

Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

На правах рукописи

Кучерявый Евгений Андреевич

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ КОМПЛЕКСА
МОДЕЛЕЙ И МЕТОДОВ РАСПРЕДЕЛЕНИЯ
РЕСУРСОВ В БЕСПРОВОДНЫХ ГЕТЕРОГЕННЫХ
СЕТЯХ СВЯЗИ**

Специальность 05.12.13 —
Системы, сети и устройства телекоммуникаций

Диссертация на соискание учёной степени
доктора технических наук

Научный консультант:
д.т.н., профессор
Самуйлов К.Е.

Москва — 2018

Оглавление

	Стр.
Введение	8
Раздел 1. Разработка методологии исследования, моделей и методов распределения ресурсов в гетерогенных сетях доступа, включая облачную архитектуру . .	24
1.1 Введение	24
1.1.1 Общие положения	24
1.2 Ожидаемая архитектура гетерогенной сети	26
1.3 Выбор критерия для справедливого распределения	31
1.3.1 Справедливое распределение ресурсов на базе критерия max-min	31
1.3.2 Пропорциональное справедливое распределение ресурсов	32
1.3.3 Обсуждение	33
1.4 Разработка нового критерия	36
1.4.1 Трафик и топологическая модель	36
1.4.2 Формулировка проблемы сетевого потока	39
1.4.3 Многослойное управление справедливым распределением ресурсов	42
1.4.4 Алгоритм решения	44
1.4.5 Эвристические подходы	45
1.5 Численные результаты	48
1.5.1 Параметры внешней среды	48
1.5.2 Проверка аналитических условий	51
1.5.3 Анализ эффективности	53
1.5.4 Балансировка производительности и справедливого распределения ресурсов	56
1.6 Гетерогенная облачная сеть радиодоступа H-CRAN	59

1.7	Распределение ресурсов в H-CRAN с использованием менеджера CRRM	65
1.8	Численные результаты исследования H-CRAN	69
1.9	Практическая реализация концепции в виде прототипа	73
1.10	Выводы	76

Раздел 2. Классификация гетерогенных сетей HetNet, разработка и исследование моделей и методов повышения их энергоэффективности		79
2.1	Введение	79
2.2	Кооперативные сети	81
2.2.1	Основные предпосылки	81
2.2.2	Введение	84
2.2.3	Основы предложенной методологии	84
2.2.4	Предлагаемая классификация гетерогенных сетей	86
2.3	Архитектура 3GPP и предлагаемые модификации	89
2.4	Повышение энергоэффективности устройств в гетерогенных сетях связи	97
2.4.1	Введение	97
2.4.2	Модель системы	101
2.4.3	Задача оптимизации энергоэффективности	105
2.4.4	Численные результаты	128
2.4.5	Заключение	138
2.5	Выводы	139

Раздел 3. Разработка и исследование моделей и методов при использовании кооперативных коммуникаций для распределения ресурсов в сетях M2M		142
3.1	Введение и предпосылки	143
3.1.1	Кооперативные коммуникации и ретрансляция	143
3.1.2	Межмашинные коммуникации	147
3.2	Трехузловая система клиентской ретрансляции с ограничением повторных передач	149

3.2.1	Описание системы и предположения	149
3.2.2	Оценка производительности	153
3.3	Схема с клиентской ретрансляцией для M2M коммуникаций	158
3.3.1	Архитектура M2M	158
3.3.2	Модель системы и предположения	160
3.3.3	Аналитические результаты	165
3.4	Клиентская ретрансляция в беспроводных сотовых сетях . .	167
3.4.1	Модель системы	167
3.4.2	Оценка производительности	171
3.4.3	Вспомогательные переменные	175
3.5	Численные результаты	176
3.5.1	Система трехузловой клиентской ретрансляции . . .	176
3.5.2	Клиентская ретрансляция для взаимодействия машина-машина	179
3.5.3	Облако клиентской ретрансляции	183
3.6	Выводы	187

**Раздел 4. Разработка и исследование моделей и методов
распределения ресурсов при использовании для
выгрузки трафика технологии D2D с сетевой
поддержкой**

		188
4.1	Введение	188
4.2	Возможности D2D в сотовых сетях	190
4.3	Способ решения системы треугольника	193
4.3.1	Описание проблематики треугольника	193
4.3.2	Специальный случай	195
4.3.3	Общий случай	196
4.3.4	Численный пример	199
4.4	Архитектурные вопросы построения динамической D2D . . .	202
4.4.1	Интеграция архитектуры	203
4.5	Оценка предложенной архитектуры	211
4.5.1	Методология оценки	212
4.5.2	Имитационное моделирование поддержки D2D	214

4.5.3	Сценарий применения гетерогенных сетей	219
4.5.4	Моделирование D2D системы	221
4.6	Результаты имитационного моделирования	223
4.6.1	Имитационное моделирование системного уровня	223
4.6.2	Результаты численного анализа	224
4.7	Практическая реализация	228
4.7.1	Программная платформа	229
4.7.2	Испытания	237
4.7.3	Перспективы внедрения	242
4.8	Дальнейшее развитие D2D	243
4.9	Выводы	245

Раздел 5. Разработка моделей и методов распределения ресурсов для доступа данных малого размера в гетерогенной сети LTE для приложений

	Интернета вещей	247
5.1	Введение и предпосылки	247
5.2	Перспективы использования стандарта IEEE 802.11ah	252
5.3	Обзор основных технических характеристик IEEE 802.11ah	254
5.3.1	Характеристики физического уровня IEEE 802.11ah	254
5.3.2	IEEE 802.11ah характеристики уровня контроля доступа к среде	255
5.4	Функционирование протокола и предложенная модель	259
5.4.1	Протокол доступа к каналу IEEE 802.11ah	260
5.4.2	Модель системы и основные предположения	261
5.5	Математический анализ	264
5.5.1	Детальный системный анализ	264
5.5.2	Основные показатели производительности	268
5.6	Методология моделирования	270
5.6.1	Описание и проверка моделирования	271
5.6.2	Текущие числовые результаты	272

5.7	Предложение по эффективной процедуре доступа для малых данных для машинно-ориентированных коммуникаций LTE	279
5.7.1	Введение в проблематику доступа данных к каналам в LTE	279
5.7.2	Традиционный доступ для данных в LTE	282
5.7.3	Предлагаемый конкурентный доступ для МТС	284
5.7.4	Системная модель и предположения	286
5.8	Анализ доступа для малых данных в LTE	289
5.8.1	Основные положения	289
5.8.2	Анализ доступа на базе PUSCH	290
5.8.3	Анализ доступа на базе PRACH	290
5.8.4	Анализ COBALT-основанного доступа для данных	291
5.9	Сравнение схем доступа для данных	293
5.10	Выводы	298

Раздел 6. Разработка моделей и методов распределения ресурсов в гетерогенных сетях дальнего радиуса действия NB-IoT для приложений Интернета вещей

6.1	Введение	299
6.2	Технологии дальнего радиуса действия для реализации решений Интернета вещей	300
6.3	Системная модель для анализа технологии NB-IoT	303
6.3.1	Модель соты LTE	305
6.3.2	Стратегии распределения ресурсов	307
6.4	Модель оценки показателей производительности	308
6.4.1	Характеристики обслуживания	310
6.4.2	Численный анализ	313
6.4.3	Марковская модель	317
6.4.4	Результаты и их анализ	319
6.5	Узкополосная технология NB-IoT с ретрансляцией	322
6.5.1	Модель системы	322

6.6	Аналитическая модель	326
6.6.1	Базовая модель системы	328
6.6.2	Усовершенствованная система с ретрансляцией	339
6.7	Численные результаты	345
6.7.1	Оценка базовой системы без ретрансляции	345
6.7.2	Оценка усовершенствованной системы с ретрансляцией	349
6.7.3	Оценка расширенной системы	351
6.8	Выводы	354
Заключение		355
Список литературы		365
Список рисунков		403
Список таблиц		408
Приложение А. Документы, подтверждающие внедрение основных результатов диссертационной работы		409

Введение

Актуальность темы исследования. Начало 21 века в области сетей и систем связи ознаменовалось появлением целого ряда технологий, которые до этого временного периода практически не использовались на сетях связи общего пользования (ССОП). Речь идет о получивших уже широкое распространение беспроводных сенсорных сетях (БСС) со всевозможными услугами и приложениями, сетях автомобильного транспорта, медицинских сетях и т.п. Разрозненное поначалу, внедрение этих сетей потребовало создания концепции, которая смогла бы не только объединить их, но и стать основой дальнейшего развития сетей и систем телекоммуникаций. Такая концепция была сформирована на базе понимания развития сети как сети не только пользователей, но и устройств, машин и т.п., получив название Интернета Вещей. Следует сразу же отметить, что концептуально Интернет Вещей смог объединить сети, ориентированные на приложения для различных отраслей, с различными требованиями по качеству обслуживания и качеству восприятия, различными характеристиками трафика, сосуществующие одновременно на одной и той же территории, в одной и той же зоне базовой станции, использующими, к тому же, и разнообразные радио технологии. Все вышесказанное привело к появлению гетерогенных сетей, модели и методы анализа которых существенно отличаются от гомогенных сетей связи. При этом в очередной раз успешное применение должна была найти и нашла теория конвергенции, предусматривающая объединение ресурсов сетей и систем с одновременным единообразным предоставлением услуг пользователям, а теперь еще и вещам. Задачи исследования гетерогенных сетей, одни из важнейших в области сетей и систем телекоммуникаций, находятся в центре внимания в диссертационной работе.

Гетерогенность сети стала в явном виде себя проявлять уже на этапе создания систем связи долгосрочной эволюции LTE (Long Term Evolution), явившихся технологической основой появления сетей связи 4G. Успешный опыт создания и внедрения систем LTE позволил научному и инженер-

ному сообществу перейти к следующему этапу развития сетей и систем телекоммуникаций, сетям пятого поколения, которые будут строиться на иных архитектурных принципах нежели все предыдущие поколения. Эти сети будут являться в полной мере гетерогенными, объединяя при этом ресурсы и как мобильных сетей телекоммуникаций, так и стационарных. Подобные беспроводные гетерогенные сети будут являться сетями, поддерживающими различные типы взаимодействий, поддерживать сверх высокую плотность пользователей, предоставлять качество обслуживания радикально разным типам трафика, что потребует для них не только разработки новых моделей и методов анализа, но даже и другой методологии распределения ресурсов (англ. Resource allocation) в сети. Кроме того, в силу ограниченной емкости беспроводных сетей на системном уровне, не очевидным является традиционное осуществление доступа через узлы сети, например, базовые станции. Появляется технология прямого взаимодействия D2D (англ. device-to-device), образующая в широком смысле и соответствующие сети D2D, для которых должны быть разработаны модели и методы анализа, в том числе и при кооперированном функционировании совместно с базовой сетью. В диссертационной работе исследования сосредоточены на разработке комплекса взаимоувязанных моделей и методов распределения ресурсов именно в гетерогенных беспроводных ССОП с учетом концепции Интернета Вещей, возможностей совместного функционирования при реализации современных сетей доступа в условиях использования множества радиотехнологий доступа, при широкомасштабном использовании межмашинного взаимодействия M2M (англ. machine-to-machine) и применении сетей D2D в условиях функционирования совместно с базовой сетью. Для проведения исследований в диссертационной работе разработана новая методология распределения ресурсов в гетерогенных сетях, отличающаяся комплексным рассмотрением различных радиотехнологий доступа.

Более того, исследование и анализ показателей качества беспроводных гетерогенных сетей практически невозможны без комплексного подхода к созданию моделей, в том числе математических, и разработке адекватных методов распределения сетевых ресурсов. Именно созданию такого

комплексного подхода посвящена тема настоящей диссертационной работы, что и обосновывает ее актуальность.

Разработанность темы исследования. Существует множество работ в области гетерогенных сетей связи различных научных школ как в Российской Федерации, так и во всем мире. Эта тематика в настоящее время является не только основной в исследовательских проектах, но и чрезвычайно востребованной вследствие своей новизны и новизны используемых моделей и методов. Вместе с тем, большинство основополагающих задач еще не решены, в том числе и на методологическом уровне.

Для решения задач создания моделей и методов распределения ресурсов в беспроводных сетях телекоммуникаций используются теория телетрафика, теория случайных процессов, теория массового обслуживания, теория графов, а также теория имитационного моделирования. Определяющий вклад в развитие теории массового обслуживания внесли российские и зарубежные ученые Г.П. Башарин, В.М. Вишневецкий, А.Н. Дудин, В.Г. Карташевский, Е.А. Крук, Б.С. Лившиц, Е.В. Морозов, А.Н. Назаров, Ю.Н. Орлов, К.Е. Самуйлов, С.Н. Степанов, Н.А. Соколов, А.П. Пшеничников, Я.В. Фидлин, О.И. Шелухин, М.А. Шнепс, А.Д. Харкевич, Г.Г. Яновский, V.-V. Iversen, M. Gerla, F. Kelly, P.J. Kuhn, L. Kleinrock, K.W. Ross, W. Whitt и др.

В диссертации в отличие от известных подходов к распределению ресурсов в гетерогенных сетях телекоммуникаций предложена методология комплексного исследования особенностей применения радиотехнологий доступа. Выбор рациональной архитектуры и оптимизация распределения ресурсов в сетях доступа являются принципиальными задачами, поскольку гетерогенность сети проявляется, в первую очередь, как раз на доступе при множестве используемых радио технологий и сетях, обеспечивающих различные приложения. Исследованию гетерогенных сетей доступа в сетях связи посвящены работы отечественных и зарубежных ученых А.С. Аджемова, В.М. Вишневецкого, Б.С. Гольдштейна, А.Е. Кучерявого, К.Е. Самуйлова, М.А. Сиверса, В.О. Тихвинского, J. Andrews, N. Himayat, F. Baccelli, S. Rangan, T. Rappoport, S. Singh, и др.

В диссертации в отличие от известных подходов предложены и исследованы гетерогенные сети доступа при кооперативном управлении посредством менеджера кооперативных радиоресурсов CRRM в H-CRAN в реальном масштабе времени на основе динамического управления с учетом справедливого распределения ресурсов для пользователей всех сетей радио доступа и общей производительности системы.

Одна из наиболее актуальных проблем в последние годы — анализ эффективности функционирования приложений концепции Интернета Вещей в условиях гетерогенных сетей радиодоступа. В этом направлении большой вклад внесли отечественные и зарубежные ученые А.Е. Кучерявый, А.В. Росляков, А.И. Парамонов, Р.В. Киричек, М. Dohler, Н. Yanikomeroglu, J. Nosek и др. В диссертационной работе для решения этих проблем предложены и исследованы модели и методы при применении клиентской ретрансляции для улучшения качества обслуживания устройств M2M, а также выявления и предотвращения перегрузки в таких сетях с учетом передачи данных малого объема.

Увеличение производительности гетерогенных систем в целом и, как следствие улучшение качества обслуживания, может быть достигнуто за счет выгрузки трафика. Проблемами выгрузки трафика в гетерогенных сетях, включая сети 4G/5G, занимались отечественные и зарубежные ученые Ю.В. Гайдамака, К.Е. Самуйлов, С.Н. Степанов, В.А. Наумов, J. Andrews, J. Araniti, M. Dohler и др. В отличие от известных работ в диссертации предложены и исследованы модели и методы выгрузки трафика в сети D2D, использующие радиотехнологию WiFi Direct. При этом предложена и исследована новая архитектура функционирования сетей D2D с поддержкой от базовой сотовой сети.

С точки зрения комплекса моделей распределения ресурсов в гетерогенных сетях созданы все теоретические предпосылки и практические наработки, позволяющие в данной диссертационной работе предложить и исследовать комплекс взаимоувязанных по структурным и нагрузочным параметрам моделей и методов распределения ресурсов для гетерогенных сетей связи, отличающийся от известных как использованием новой методологии анализа таких сетей, так и взаимным дополнением совместимых

моделей и методов при предстоящем широкомасштабном внедрении технологий M2M и D2D с использованием новых методов выгрузки трафика при взаимодействии сетей в гетерогенных условиях и кооперативных методов управления в облачных сетях доступа.

Объект и предмет исследования. Объектом исследования являются беспроводные гетерогенные радиотехнологии доступа, а предметом исследования - модели и методы распределения ресурсов в беспроводных гетерогенных сетях связи.

Цель и задачи исследования. Цель работы состоит в разработке и исследовании комплекса взаимоувязанных моделей и методов распределения ресурсов в современных и перспективных гетерогенных беспроводных сетях связи общего пользования.

Цель работы достигается путем последовательного решения следующих задач:

- анализ существующих моделей и методов распределения ресурсов в беспроводных сетях связи;
- разработка методологии распределения ресурсов в беспроводных гетерогенных сетях связи, адекватной их типовой структуре;
- поиск нового критерия оптимизации распределения ресурсов в сетях связи для обеспечения справедливого распределения ресурсов в гетерогенных сетях с учетом требуемого соотношения между сетевыми возможностями и пропускной способностью;
- разработка методологии оптимизации производительности гетерогенных сетей радиотехнологий доступа H-CRAN с использованием кооперативного управления в реальном масштабе времени;
- разработка новых архитектурных решений для гетерогенных сетей связи, не требующих обеспечения доступа ни в операторскую магистральную сеть, ни в Интернет при использовании технологий WLAN, в том числе для пользовательского оборудования не-3GPP стандарта;
- разработка модели и метода выгрузки трафика в гетерогенных сетях с использованием технологии WiFi Direct и практическая реализация, развернутая на тестовой сети LTE;

- разработка модели и метода распределения ресурсов для взаимодействия D2D в гетерогенных сетях при наличии сетевой поддержки;
- разработка модели и метода анализа механизмов доставки данных от M2M устройств в сеть LTE, в том числе при использовании технологии NB-IoT;
- разработка модели и метода доступа устройств M2M к ресурсам систем LTE с учетом специфики обслуживания данных малого объема.

Научная новизна диссертационной работы состоит в следующем:

1. Предложена новая методология исследования распределения ресурсов в гетерогенных сетях связи, отличающаяся от известных комплексным анализом и синтезом различных технологий радиодоступа.
2. Предложен новый критерий оптимизации распределения ресурсов в гетерогенных сетях связи, отличающийся от известных тем, что справедливое распределение ресурсов для гетерогенных сетей с учетом требуемого баланса между сетевыми возможностями и пропускной способностью достигается за счет использования критерия $\max\text{-min}$, основанного на полосе пропускания. Целевая функция при этом дополняется весовыми коэффициентами на основе мгновенных значений спектральной эффективности.
3. Разработан метод управления производительностью гетерогенных радиотехнологий доступа H-CRAN, отличающийся от известных тем, что используется кооперативное управление посредством менеджера кооперативных радиоресурсов CRRM в H-CRAN в реальном масштабе времени на основе динамического управления ресурсами двух альтернативных метрик: справедливого распределения ресурсов для пользователей всех доступных сетей радиодоступа и общей производительности системы.
4. Предложена новая архитектурная реализация взаимодействия беспроводных сетей 3GPP и IEEE, отличающаяся от известных тем, что используется разработанный в диссертационной работе «Шлюз

входа в сеть доступа» (AAGW), обеспечивающий такое совместное функционирование сетей 3GPP и IEEE, при котором не требуется использование ни операторской магистральной сети, ни Интернета.

5. Разработана методология эффективного распределения мощности передачи и схема управления мощностью для одновременно функционирующих двух и более радиотехнологий доступа устройства в гетерогенной сети, максимизирующая энергоэффективность мобильного устройства при удовлетворении минимальной требуемой скорости передачи пользовательских данных.
6. Разработаны модель и метод облачной клиентской ретрансляции, которые, в отличие от известных, учитывают особенности трафика для устройств M2M, расположенных на границе соты.
7. Разработаны модель и метод выгрузки трафика в гетерогенных сетях, отличающиеся от известных тем, что для выгрузки трафика используется технология WiFi Direct.
8. Разработаны модель и метод распределения ресурсов для взаимодействия D2D в гетерогенных беспроводных сетях, отличающиеся от известных использованием новой технологии сетевой поддержки D2D, что позволяет существенно сократить общие энергозатраты в гетерогенной зоне базовой станции и увеличить пропускную способность сети в целом.
9. Разработаны модель и метод доступа устройств M2M к ресурсам системы LTE, отличающиеся от известных учетом специфики обслуживания данных малого объема, что позволяет организовать эффективное обслуживание трафика M2M в LTE.
10. Разработаны модель и метод доставки данных от M2M устройств в гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяющие для ряда сценариев как повысить вероятность доставки сообщений, так и повысить энергоэффективность передающих узлов.

Таким образом, в диссертационной работе создан новый комплекс моделей и методов исследования и анализа задач распределения ресурсов в беспроводных гетерогенных сетях связи.

Теоретическая и практическая значимость работы. Теоретическая значимость работы обусловлена, прежде всего, разработкой новой методологии исследования распределения ресурсов в гетерогенных сетях связи, отличающейся комплексным рассмотрением различных радиотехнологий доступа. Новая методология исследований не только позволила в явном виде решить ряд сложнейших задач для гетерогенных сетей связи, но может быть широко использована и для различных приложений в области сетей и систем связи, например, для Интернета Вещей. Важнейшим теоретическим результатом является определение нового критерия оптимизации распределения ресурсов в гетерогенных сетях связи, отличающегося от известных тем, что справедливое распределение ресурсов для гетерогенных сетей с учетом требуемого баланса между сетевыми возможностями и пропускной способностью достигается за счет использования критерия $\max\text{-}\min$, основанного на полосе пропускания.

Метод управления производительностью гетерогенных облачных радиотехнологий доступа H-CRAN с использованием кооперативного управления посредством менеджера кооперативных радио ресурсов в реальном масштабе времени представляет собой важный вклад в исследования гетерогенных сетей связи. Разработанный комплекс моделей и методов позволяет адекватно текущему уровню развития сетей и систем телекоммуникаций решить целый ряд новых задач по выгрузке трафика, предотвращению перегрузок, а также проводить оптимизацию распределения радиоресурсов и сокращения общих энергозатрат при наличии множества радиотехнологий доступа. Все это в целом вносит существенный вклад в развитие теории массового обслуживания и теории телетрафика.

Практическая ценность работы состоит в создании научно-обоснованных рекомендаций по созданию и планированию гетерогенных сетей связи. Самостоятельную практическую ценность имеет разработанная под руководством автора система имитационного моделирования WINTERsim, используемая рядом научно-исследовательских групп.

Полученные в диссертационной работе результаты внедрены в ПАО «Ростелеком» при выполнении НИР по современным технологиям для модернизации сетей связи в направлении создания сетей 5G, в ПАО «ГИПРОСВЯЗЬ» при разработке методики по планированию гетерогенных сетей связи, в Российском университете дружбы народов при создании модельной распределенной сети для исследования и оптимизации работы механизмов в системах связи нового поколения 5G, в Национальном Исследовательском Университете «Высшая школа экономики» при чтении лекций и проведении практических занятий.

Методы исследования. Для решения поставленных в работе задач использовались методы теории вероятностей, теории массового обслуживания и теории телетрафика, имитационного и натурального моделирования.

Положения, выносимые на защиту.

1. Критерий оптимизации, для которого справедливое распределение ресурсов для гетерогенных сетей с учетом баланса между потоками трафика с разными требованиями к пропускной способности обеспечивается использованием модифицированного критерия $\max\text{-min}$, основанного на доступной полосе пропускания позволяет достичь до 60% лучшего значения 5% квантиля производительности соты по сравнению с критериями, не учитывающими динамику изменений характеристик радиоканала.
2. Метод управления производительностью гетерогенной облачной сети радиодоступа H-CRAN с использованием менеджера управления кооперативными радиоресурсами CRRM в реальном масштабе времени на основе двух альтернативных метрик справедливого распределения ресурсов для пользователей всех доступных сетей радиодоступа позволяет более чем в 2 раза увеличить общую производительность мультиуровневой системы радиодоступа.
3. Новый архитектурный элемент «Шлюз входа в сеть доступа» (AAGW), реализованный в соответствии со стандартами 3GPP, позволяет взаимодействовать технологиям 3GPP и IEEE и обеспечивает значительное сокращение сквозных задержек вследствие того, что трафик не нуждается в транзите на удаленный шлюз

3GPP PDN, и существенное сокращение капитальных затрат на развертывание дополнительной инфраструктуры 3GPP.

4. Модель выбора эффективной мощности передачи и метод управления мощностью мобильного устройства в условиях двух и более радиотехнологий доступа максимизирует энергоэффективность мобильного устройства при обеспечении минимально требуемой скорости передачи.
5. Модель и метод облачной клиентской ретрансляции, учитывающие особенности трафика для устройств M2M, расположенных на границе соты, позволяют увеличить производительность системы в 3 — 4 раза при уменьшении в несколько раз задержки пакетов.
6. Модели и методы выгрузки трафика в гетерогенных сетях, использующие для выгрузки трафика технологию D2D на базе WiFi Direct с сетевой поддержкой, обеспечивают удвоение пропускной способности соты и повышение энергоэффективности до 6 раз.
7. Модель и метод доступа устройств M2M к ресурсам сети LTE на основе модифицированной процедуры COBALT, учитывающие особенности передачи данных малого объема, позволяют уменьшить задержку пакетов по сравнению с методом PUSCH почти для 85% типов трафика и одновременно в 2 — 3 раза снизить энергопотребление по сравнению с методом PRACH.
8. Модель и метод доставки данных от M2M устройств в гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяют для ряда сценариев как повысить вероятность доставки сообщений, так и повысить энергоэффективность передающих узлов.

Степень достоверности и апробация результатов подтверждается корректным применением математического аппарата, результатами имитационного моделирования и широким спектром публикаций и выступлений как на российских, так и на Международных конференциях. Основные положения диссертационной работы были представлены и обсуждались на следующих конгрессах, конференциях и семинарах: IEEE Globecom: Anaheim CA, USA 2012, Atlanta GA, USA 2013, San Diego

CA, USA 2015, Washington DC, USA 2016, Singapore 2017; IEEE ICC (IEEE International Conference on Communications): Budapest Hungary 2013, Sydney Australia 2014, London UK 2015, Kuala Lumpur Malaysia 2016; IEEE CAMAD (IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks) Kyoto, Japan 2011; IFIP WWIC (International Conference on Wired/Wireless Internet Communication) Vilanova i la Geltrú, Spain 2011, Santorini Greece 2012, St. Petersburg Russia 2013; DCCN (International Conference on Distributed Computer and Communication Networks), ИПУ РАН, Moscow, Russia, 7 – 10 октября 2013; Конференции «XLIII Неделя науки Санкт-Петербургского политехнического университета Петра Великого», Санкт-Петербург, Россия, 4 – 5 декабря 2014; Юбилейной 70-й научно-технической конференции профессорско-преподавательского состава СПбГЭТУ «ЛЭТИ», Санкт-Петербург, Россия, 1 – 11 февраля 2017 г.; семинарах НИУ ВШЭ, РУДН, МТУСИ, ИПУ РАН, СПб ГУТ.

Публикации. Основные результаты диссертации изложены в 78 опубликованных работах, в том числе в трех монографиях, 12 работах, опубликованных в журналах из перечня ВАК Министерства образования и науки Российской Федерации; в 63 работах, опубликованных в трудах, индексируемых Scopus, из них 51 индексируемых в Web of Science.

Структура и объем работы. Диссертация состоит из введения, шести разделов, заключения, списка литературы и приложения. Общий объем диссертации 416 страниц, включая 102 рисунка, 20 таблиц, список литературы из 318 наименований. В приложении к диссертационной работе приведены документы, подтверждающие внедрение основных результатов диссертационной работы.

Личный вклад автора. Все основные результаты диссертационной работы получены автором самостоятельно. Экспериментальные исследования проведены при его непосредственном участии и под научным руководством.

Соответствие паспорту специальности. Диссертационная работа выполнена по специальности 05.12.13 - Системы, сети и устройства теле-

коммуникаций и соответствует следующим пунктам паспорта специальности: 3, 5, 11, 12, 14.

Краткое содержание работы.

В первом разделе диссертационной работы основное внимание уделяется гетерогенным сетям как таковым и их взаимодействию с облачными структурами сети доступа. Отмечается, что исследования в диссертации проводятся в условиях предстоящего в третьем десятилетии 21 века широкомасштабного внедрения сетей и систем связи пятого поколения, что требует принципиального изменения взгляда на развитие сетей в связи со многими принципиально новыми типами услуг и взаимодействий. Обеспечить требуемые характеристики сетей можно только при интеграции всех возможных ресурсов всех прогрессивных технологий, что обосновывает необходимость перехода от гомогенных структур сетей связи к гетерогенным.

В разделе рассмотрено понятие взвешенного справедливого распределения ресурсов для гетерогенных беспроводных сетей, что способно обеспечить надлежащий баланс между общей производительностью системы и справедливым распределением ресурсов. Кроме того, продемонстрировано, что это может быть сформулировано как специальный случай проблемы $\max\text{-min}$ справедливого распределения, оптимизационные задачи для которой относятся к линейному программированию.

Проведено сравнение эффективности предложенного критерия с классическими задачами $\max\text{-min}$, а также с некоторыми эвристическими методами, и доказано при этом, что предложенное в диссертационной работе решение является более выгодным для достижения баланса между общей производительностью системы и справедливым распределением ресурсов для ее пользователей.

Для оценки предложенного подхода в диссертационной работе были исследованы облачная сеть радиодоступа типовой структуры и с менеджером кооперативных радио ресурсов CRRM. Было доказано, что H-CRAN основанное управление ресурсами с менеджером CRRM эффективнее, чем запрашиваемая справедливость или требуемая производительность. Система имеет более высокие значения справедливости и производительности с увеличением доступных сетей RAN, но «пико плюс макро» более выгодно,

чем «WiFi плюс макро», даже хотя использование полосы пропускания в 20 МГц для WiFi должно было бы иметь преимущество по сравнению с 10 МГц пико сотового канала LTE.

Во втором разделе разработана новая архитектура гетерогенной сети с использованием предложенного в диссертации «Шлюза входа в сеть доступа» (AAGW), который отражает функциональность решений 3GPP во WLAN радио технологиях RAT (например, WiFi) и наоборот, что обеспечивает значительное сокращение сквозных задержек вследствие того, что трафик не нуждается в транзите на удаленный шлюз PDN, и существенное сокращение затрат на развертывание малых сот, поскольку не требуется обеспечивать ни операторскую магистральную сеть, ни доступ в Интернет.

Далее в рамках концепции гетерогенных сетей HetNet рассматривается проблема энергоэффективного управления доступными беспроводными радиотехнологиями в пользовательском устройстве. Целью ставится разработка эффективных методов, позволяющих значительно повысить производительность устройства с точки зрения результирующей энергоэффективности. Разработан алгоритм выбора мощности передачи мобильного устройства в условиях наличия двух и более радиотехнологий, способных функционировать одновременно. Далее на примерах одновременного использования радиотехнологий 3GPP LTE и IEEE WiFi исследованы области сети, разделенные по принципу выбора мощности передачи. Также предложен простой эвристический метод выбора мощности передачи устройства в гетерогенной сети, удовлетворяющий заданным ограничениям и демонстрирующий высокую энергоэффективность.

Раздел 3 посвящен вопросам клиентской ретрансляции, без анализа которой в дальнейшем не представляется возможным обосновано перейти к исследованию взаимодействия D2D и систем с клиентской ретрансляции на базе технологии NB-IoT, фактически основ построения гетерогенных сетей и систем связи пятого поколения. В разделе предложена схема клиентской ретрансляции для улучшения показателей задержки и энергетической эффективности для граничных ячеек M2M устройств в условиях плохого качества связи. Предложенный аналитический подход дает существенный

выигрыш в пропускной способности, что было подтверждено имитационным моделированием.

Были также рассмотрены беспроводные сотовые сети в условиях распределенного управления кооперативными коммуникациями посредством облака клиентской ретрансляции. Основная цель такой облачной структуры состоит в увеличении пропускной способности путем поддержки мобильных клиентов, находящихся на границе покрытия с плохим качеством связи. Определены основные характеристики производительности для предложенной структуры, такие как пропускная способность, задержки и энергетическая эффективность. Получены аналитические выражения в замкнутом виде и верифицированы путем имитационного моделирования на системном уровне.

В разделе 4 диссертационной работы предложены и исследованы модели и методы выгрузки трафика в беспроводных гетерогенных сетях, отличающиеся от известных тем, что для выгрузки трафика используется технология WiFi Direct. Кроме того, в этом же разделе разработана новая технология поддержки соединений D2D со стороны сети и соответствующие методы распределения ресурсов с целью сокращения общих энергозатрат. Объединение этих подходов позволило получить существенно лучшие характеристики сети в целом по сравнению с известными результатами. Предложенная в диссертационной работе новая архитектура сети с использованием D2D может быть сопоставлена с объектами архитектуры 3GPP с единственным дополнительным объектом — сервером D2D (называемом ProSe, Proximity Services), находящимся в EPC (Evolved Packet Core) фрагменте сети. Такое размещение позволяет серверу D2D связываться с «Центром местоположения» (Serving Mobile Location Center, SMLC), чтобы позиционировать пользовательское оборудование UE, а также эффективно взаимодействовать с серверами приложений внешних провайдеров и операторов. Для натуральных экспериментов была использована модельная сотовая сеть на базе департамента телекоммуникаций Технологического Университета Брно (BUT) в Чехии. BUT LTE зона полностью коммерчески пригодная версия для внедрения всех важнейших подсистем, входящих в современные сети 4G. Сервер D2D ProSe был реализован как виртуальное

устройство. При натуральных экспериментах наблюдался значимый синергетический эффект при совместном использовании мобильной сети и сети D2D.

Раздел 5 посвящен задачам внедрения Интернета вещей в сетях IEEE 802.11ah и 3GPP LTE. В частности, подробно рассмотрена технология IEEE 802.11ah и проведен ее детальный анализ, для чего была построена аналитическая модель и полученные результаты, такие как пропускная способность и вероятность коллизии, полученные для различных условий, проверены при помощи имитационного моделирования.

Также разработаны модель и метод доступа устройств M2M к ресурсам LTE, учитывающие специфику обслуживания данных малого объема, для организации эффективного обслуживания трафика M2M в LTE.

Далее в разделе предложен и детально исследован эффективный метод передачи малых данных, который может быть использован как альтернатива традиционной сигнализации, существенно увеличивая при этом производительность для сетей M2M. Основная идея предлагаемого конкурентного доступа для систем LTE (contention-based LTE transmission, COBALT) состоит в передаче пакетов малых данных непосредственно через канал типа PUSCH вместо стандартных каналов PUCCH/PRACH. В диссертационной работе предлагается использовать механизм COBALT для сценария M2M, когда число устройств очень велико, а трафик нерегулярен и мал. При этом ожидается, что в столь экстремальных условиях предложенная схема будет допускать малые задержки и, что даже более важно, уменьшит потребление энергии на сигнализацию для небольших M2M устройств, электропитание которых осуществляется от батареи.

В разделе 6 диссертационной работы рассматриваются задачи внедрения Интернета Вещей в условиях развертывания узкополосной технологии передачи данных 3GPP NB-IoT. В частности, рассмотрены стратегии распределения ресурсов при развертывании технологии NB-IoT на сетях LTE. Учитывая ограниченный объем радиоресурсов, выбор стратегии распределения радиоресурсов между высокоскоростными приложениями LTE и большим количеством низкоскоростных счетчиков NB-IoT является важной задачей для операторов беспроводных сетей. В рамках этой задачи

необходимо сбалансировать распределение ресурсов таким образом, чтобы характеристики доставки трафика для обоих типов были наиболее эффективными.

В разделе разработана аналитическая модель для исследования характеристик обработки трафика NB-IoT при наличии конкурирующего трафика LTE. Используя разработанный инструмент, проводится сравнение трех возможных стратегий распределения радиоресурсов: статической, динамической и динамической с резервированием. В качестве эталонной модели рассматривается внутрисполосный режим, представленный в 3GPP Release 13, где радиостанция LTE разделяет спектр частот с технологией NB-IoT. Аналитические результаты для сот размером 10 и 30 км. показали, что динамическая стратегия с резервированием рекомендуется для будущих развертываний NB-IoT в мобильных сетях LTE, а разработанная аналитическая модель может применяться для оценки эффективности будущих решений.

Далее в разделе исследуется технология NB-IoT при внедрении функции клиентской ретрансляции. Для оценки эффекта ретрансляции разработана новая аналитическая модель, которая учитывает особенности «краудсорсинга» на основе ретрансляторов, мобильность вспомогательных транспортных средств и важные особенности технологии NB-IoT. Результаты для системы с ретрансляцией показывают, что для ряда сценариев по сравнению со сценариями без ретрансляции повышается как вероятность доставки сообщений, так энергоэффективность передающих узлов. Полученные численные результаты проверяются с использованием системы имитационного моделирования WINTERsim.

Раздел 1. Разработка методологии исследования, моделей и методов распределения ресурсов в гетерогенных сетях доступа, включая облачную архитектуру

1.1 Введение

1.1.1 Общие положения

В настоящее время университеты, научно-исследовательские институты, предприятия промышленности целенаправленно продвигаются вперед в направлении разработки технологий, систем и сетей пятого поколения 5G (англ. Fifth Generation), имея конечной целью создание полнофункциональной сотовой сети связи 5G к 2020 году. Основные требования к технологиям 5G предполагают создание системы с очень большой емкостью (в 1000 раз больше по сравнению с 4G), и обеспечение равномерного бесшовного покрытия сети с возможностью высокоскоростной передачей данных [1], [2]. В длительной перспективе основные игроки на рынке систем и сетей связи, включая традиционных поставщиков телекоммуникационного оборудования, производителей устройств и операторские компании, будут заняты внедрением прогрессивных технологий 5G, тренды развития которых видны уже сегодня [3]. Отметим, что для достижения тысячекратного увеличения емкости два ключевых механизма являются необходимыми – пространственное уплотнение сети и агрегация спектра [4]. В первом случае естественным предположением по построению сетей с большой плотностью выглядит увеличение числа сот с меньшей областью покрытия в существующей сетевой архитектуре. Во втором случае предполагается совместное использование больших участков радио спектра в разнообразных спектральных диапазонах (лицензируемых и нелицензируемых), включая, в перспективе, и крайне высокие частоты.

Быстрое распространение недорогих и энергоэффективных малых сот различного размера (микро, пико, фемто и т.д.) с поддержкой разнообразных технологий радиодоступа привело к созданию так называемых гетерогенных сетей Heterogeneous Network (HetNet) [5]. Еще большее уплотнение сети способствует появлению сетей сверхвысокой плотности (Ultra-dense HetNets [6]), базирующихся на сетевой инфраструктуре с поддержкой технологий множественного радио интерфейса (например, малые соты с возможностями функционирования в различных диапазонах). В свою очередь для макро сот роль «якоря» (обеспечение покрытия крупномасштабных зон) становится даже более важной чем обеспечение необходимой емкости. Только макро сота, обладающая необходимым покрытием, может обеспечить централизованные и сбалансированное распределение пользователей по «бустерам» (малым сотам с большой емкостью), что будет способствовать эффективной выгрузке трафика данных пользователей в условиях повышенной нагрузки на сеть [7], [8].

Главным преимуществом рассмотренной выше архитектуры «якорь-бустер» является именно централизованное управление через макро соту. В первую очередь централизованное управление улучшает возможности координации между малыми сотами с множественным радио интерфейсом [9]. Таким образом, уровень сетевой кооперации смещается от «свободных» решений по взаимодействию между разными технологиями передачи данных к централизованным или распределенным схемам с более жесткой интеграцией на уровне сети радиодоступа (Radio Access Network, сокращенно RAN). Прогрессивные механизмы взаимодействия на уровне сети радиодоступа недавно были рассмотрены в стандартах 3GPP (3rd Generation Partnership Project), и эффективные решения для интеграции стандартов 3GPP и беспроводных локальных сетей (WLAN) были включены в стандарты систем длительной эволюции LTE версии 12 (Release 12) [10]. Агрегация решений WLAN/LTE с тесной интеграцией WLAN в сети операторов сотовой связи также стандартизована и для LTE версии 13 (Release 13). В целом, прирост емкости и улучшение качества обслуживания QoS (Quality of Service) [11] посредством интеграции множественных сетей связи зависят от того, как процедуры выбора и интеграции радио технологий скоорди-

нированы с уровнем управления доступа к каналу (Medium Access Control Layer) и сетевым уровнем (Network Layer) [12].

Сегодня существует множество альтернативных радио технологий (GSM/EDGE, WCDMA/HSPA, LTE-FDD/TDD, WiFi, BLE, WiGig и т.д.), поэтому появляющаяся кооперация на уровне сети радиодоступа должна быть более гибкой, чем возможные решения для «свободной» высокоуровневой координации, что позволит динамично управлять радио ресурсами для увеличения производительности системы и для пользователей [13]. В частности, интеграция на уровне сети радиодоступа предоставляет возможности для управления потоками данных в реальном времени с целью гибкого распределения трафика пользователя между доступными радио интерфейсами, оптимизируя общую производительность гетерогенной сети [14].

В этом разделе предлагается механизм динамического распределения ресурсов каналов доступа гетерогенных сетей, с целью обеспечения максимальной выгоды с точки зрения пропускной способности конечного пользователя/системы в целом, и/или максимизации честности распределения ресурсов. При этом, требуется обеспечить низкую сложность решений, что необходимо для осуществления практической реализации.

1.2 Ожидаемая архитектура гетерогенной сети

В этом разделе мы обобщим концептуальные понятия для гетерогенных сетей и продолжим детализацию для различных компонентов этих сетей. Кроме того, мы проанализируем состояние исследований и их развитие в этой предметной области, а также определим, как цели, отмеченные в предыдущих разделах, могут быть практически реализованы.

Интересно отметить, что впервые концепция гетерогенных сетей была представлена еще в спецификациях второго поколения сотовых сетей как понятие о малых сотах. Впоследствии в спецификациях 3G эта концепция не получила эволюционного развития, поскольку основные техни-

ческие усилия были сконцентрированы на оптимизации отдельных радио технологий (сотовых, беспроводных локальных сетей и т.д.).

Политика централизации управления для мульти радио сетей (например, WLAN и 3GPP систем) была введена в спецификациях четвертого поколения 4G путем использования механизма центрального сервера, известного как функция обнаружения и выбора сети доступа ANDSF (Access Network Discovery and Selection Function) [15]). Эта стандартизованная схема позволяет сотовым сетям инкорпорировать в себя политики управления обнаружением и доступом в сеть для других, т.е. не-3GPP сетей. В частности, функция ANDSF может поддерживать пользовательское оборудование UE (user equipment) для близлежащих не-3GPP сетей доступа (функция обнаружения). Также отметим, что пользовательское оборудование определяет ограничения и предпочтения для близлежащих сетей доступа посредством функции межсистемной политики мобильности. Если пользовательское оборудование позволяет использовать более, чем один простой радио интерфейс в конкурентной среде, выгрузка трафика может быть специфицирована посредством выработки межсистемной политики маршрутизации. В целом, функция ANDSF реализует политики обнаружения и доступа, которые используются для длительных периодов функционирования, но не поддерживают динамическое управление и расписание использования радио ресурсов для мульти радио технологий.

В диссертационной работе мы концентрируем внимание на концептуальных вопросах создания гетерогенных сетей [16], когда имеется больше возможностей для управления радио ресурсами, чем это возможно только с функцией ANDSF, когда мульти-радио пользовательское оборудование может одновременно взаимодействовать с любым числом радио технологий. Для представления в общих чертах свойств таких систем далее обсудим определение сетей HetNet как гетерогенных.

Каждая из радио технологий имеет свои собственные специфические архитектуру и принципы передачи данных. Эффективное одновременное использование мульти радио технологий является одной из ключевых особенностей гетерогенной сети. Существует два естественных подхода к реализации этой особенности. Первый состоит в создании централизованного

координатора, который будет обеспечивать полное управление всеми ассоциированными сетями радиодоступа. Однако, необходимость соответствующих сигнальных процедур делает этот подход осуществимым тогда, когда спецификация каждой из радио технологий разрабатывается как часть некоей взаимоувязанной структуры. Здесь трудности возникают сразу же при попытке удовлетворить разнообразные требования каждой из вовлеченных в координацию радио технологий и создать сбалансированные сигнальные процедуры, не ухудшая эффективность индивидуальных сетей. Другой подход в части координации сетей и систем гетерогенной сети состоит в обеспечении функциональной совместимости индивидуальных радио технологий, составляющих гетерогенную сеть. В этом случае, потенциально оптимальное управление гетерогенной сетью может стать невозможным из-за отсутствия полной информации о доступных сетевых ресурсах со стороны каждой из радиосетей доступа, входящих в гетерогенную сеть. Соответственно, требуемая кросс-координация для радио технологий может быть реализована посредством межсетевых транзитных линий, что может сильно увеличить нагрузку на контрольные каналы. Альтернативный подход заключается в том чтобы делегировать функцию выбора радио технологии самому пользовательскому оборудованию посредством обеспечения его всей необходимой информацией. Однако, это решение может быть неоптимальным в динамических условиях, поскольку оно предполагает достаточно большую гибкость в как в решениях производителей мобильных устройств, так и поведения конечных пользователей.

Кроме того, для сетей WPAN, WLAN, WMAN и WWAN требуется единообразный подход к планированию покрытия. Базовые станции традиционных сотовых сетей как правило классифицируются по эффективности покрытия, например, макро и микро соты. Подобным образом, соты гетерогенных сетей могут варьироваться по размеру зоны покрытия, начиная с WPANs и WLANs например, Bluetooth и WiFi технологий и заканчивая крупномасштабными зонами покрытия LTE WWANs, обычно имеющими радиус покрытия в несколько километров. Однако, существует очень большая разница в покрытии для различных технологий. Типовая макро сота WMAN имеет зону обслуживания, которая в существенной степени зави-

сит от условий эксплуатации, в то время как небольшие пико и фемто соты WLAN имеют покрытие от десятков до сотен метров.

В дополнение архитектурным различиям гетерогенных сетей, возможна также дифференциация по отношению к ожидаемому использованию отдельных радио технологий. Эта классификация относится к уровню приложений, поскольку сети 2G/3G ориентируются на передачу речи, а, например, WLAN предназначается, в основном, для передачи данных. В данном подходе обеспечение пользователей услугами становится основным в системе оптимизации. Кроме того, сценарии для межмашинного взаимодействия M2M (machine-to-machine) кардинально изменяют выбор различных вариантов построения сети. В теории, новые гетерогенные сети могут обеспечить поддержку для традиционных устройств LTE и WiFi оборудования, мульти-радио пользовательского оборудования LTE-WiFi, прямого взаимодействия устройство-устройство D2D (device-to-device) через WiFi и LTE интерфейсы, а так же масштабные машинно-ориентированные приложения.

Завершая данный раздел, мы сформулируем важнейшие характеристики модели построения гетерогенной сети, которая включает две радио технологии (3GPP LTE и IEEE 802.11 WiFi), два типа покрытия (макро и пико соты для LTE и WiFi) и простое приложение – передачу данных по модели эластичного трафика. Отдельно мы сфокусируем внимание на перспективе, когда гетерогенная система управляется с помощью простого координирующего узла. Этот узел собирает всю текущую информацию в отношении трафика пользователя, также как и доступность покрытия конкретной радио технологией в зоне обслуживания. Как только такая информация становится известна, координатор делает выбор наилучшей радио технологии для всех ассоциированных пользователей, а также рекомендует каждому пользователю, какие именно ресурсы доступных радио технологий могут быть использованы. Кроме того, мы предполагаем, что координирующий узел имеет полное знание о доступности услуг в каждой сети радиодоступа вместе с информацией о текущих характеристиках каналов для каждого пользователя.

Естественно, обсуждаемый координатор может физически располагаться на стороне макро базовой станции, куда текущая информация с пико сот и с точек доступа WiFi может поставляться по прямым проводным и беспроводным линиям. В этом случае выделенный шлюз между точками доступа WiFi и макро базовой станцией LTE должен быть установлен на этапе строительства сети. В свете текущей интеграции уровня радиодоступа 3GPP/WLAN наша ожидаемая архитектура становится структурно подобной новой концепции гетерогенной облачной сети радиодоступа, рассмотренной в [17]. Предложенная система управления функционирует на пакетном уровне в противовес потоковому уровню, что дает возможность для пакетов, принадлежащих одному и тому же потоку быть переданными и по конкурентному радио интерфейсу. Координатор при этом гарантирует, что пакеты одного и того же интерфейса будут впоследствии ассемблированы корректным образом. На практике это может быть реализовано с помощью механизма умного туннелирования между пользовательским оборудованием и узлом координатором. В доказательство этого мы уже продемонстрировали эти возможности на простом прототипе, когда пользовательское оборудование использует такое умное туннелирование (основанное на архитектуре OpenFlow) для разделения своего трафика динамически между LTE и WiFi радио интерфейсами [18].

В дальнейшем, мы усовершенствовали доступные решения упомянутой проблемы централизованного размещения радио ресурсов в мульти-радио гетерогенных сетях используя методы теории оптимизации. В этом случае задача централизации координирующего узла должна быть решена исходя из распределения общего ресурса системы, который хотелось бы оптимизировать с учетом возможных интересов всех подсистем. Следовательно, важнейшим подходом к оптимизации функционирования систем с множественными радио технологиями является выбор критерия оптимизации, который мог бы удовлетворять ожиданиям и операторов, и конечных пользователей. Мы продолжим эту дискуссию ниже.

1.3 Выбор критерия для справедливого распределения

Исходя из базовых принципов функционирования сетей связи очевидно, что следующие принципы работы представляются важными: (i) равный доступ к ресурсам, распределяемым между пользователями и (ii) максимизация общей пропускной способности системы. Типовым случаем является нахождение компромисса между этими двумя принципами. В зависимости от эффективности политики оператора сети при распределении ресурсов между конечными пользователями существует возможность баланса между этими двумя метриками. Поэтому, существуют два хорошо известных критерия оптимизация для справедливого распределения ресурсов известные как критерии max-min и пропорциональное справедливое распределение (англ. proportional fairness).

1.3.1 Справедливое распределение ресурсов на базе критерия max-min

Справедливое распределение ресурсов в соответствии с max-min является одним из наиболее распространенных критериев, изначально определенных Бертсекасом и Галлагером в [19]. Обозначая N как число пользователей, желающих быть обслуженными сетью, и P_d как число путей, доступных для обслуживания этими запросами d , $d = 1, 2, \dots, N$, задача состоит в максимизации минимальной пропускной способности $\sum_{p=1}^{P_d} x_{dp}$, $d = 1, 2, \dots, N$, при условии соблюдения ограничений на пропускную способность линии, которая в тоже время не приобретает отрицательных значений для любого распределения. Пусть $\vec{x} = (x_1, x_2, \dots, x_N)$ есть вектор ресурсов, значения которого расположены в порядке возрастания. Мы говорим, что вектор \vec{x} обеспечивает max-min распределение, если он является лексикографически максимальным среди всех возможных векторов распределений, значения которых также расположены в порядке возраста-

ния [20]. Вектор \vec{x} лексически больше, чем \vec{z} , если существует такой индекс k , что $x_i = y_i$, $i = 1, 2, \dots, k$, и $x_k > z_k$. Следовательно, распределение называется max-min оптимальным, если нет другого пути увеличить его для конкретного запроса i за счет запросов j с большим распределением. Лексикографическая максимизация относится к задачам линейного программирования, что позволяет эффективно использовать множество вариантов решений, например, симплекс-метод [21]. Важно отметить, что в классическом понимании max-min критерий всегда изменит общую пропускную способность системы в пользу справедливого распределения ресурсов. В то же самое время, простота оптимизационной задачи в данном случае является очень привлекательной, что потенциально дает возможность использовать ее в реальном времени для управления радио ресурсами.

1.3.2 Пропорциональное справедливое распределение ресурсов

В течение двух последних десятилетий ряд авторов поставили под сомнение целесообразность использования критерия max-min. Например, Массоули и Робертс [22] писали: *«На самом деле нет четких экономических причин, почему критерий max-min предпочтительнее, чем другие методы распределения полосы пропускания. Более рациональным было бы максимизировать общую выгоду при учете затрат и осознанной ценности или минимизировать ожидаемое время ответа на запрос при любом способе передачи информации»*. Кроме того, в [23] обращает внимание на то, что в одной из наиболее успешных процедур управления скоростью на основе окон для протокола TCP используется близкое к пропорциональному распределение ресурсов для путей передачи потоков данных.

Продолжая сказанное, отметим, что традиционно используемый критерий для справедливого распределения ресурсов

$$\sum_{d=1}^N \log x_d \rightarrow \max, \quad (1.1)$$

имеет ряд полезных свойств, которые вытекают из математических особенностей функции $f(x) = \log(x)$. Во-первых, когда требуемая скорость для определенного запроса мала, выражение (1.1) имеет очень большую отрицательную составляющую. Во-вторых, хотя $f(x) = \log(x)$ и является монотонно возрастающей функцией, анализ поведения ее производной $f'(x) = (\log x)' = 1/x \ln(x)$ показывает, что скорость роста уменьшается, если x увеличивается. Такое поведение функции исключает экстремально большие распределения ресурсов. Исходя из сказанного, можно также отметить, что основание логарифмической функции в (1.1) не имеет значения. По сравнению с max-min справедливым распределением пропорциональное справедливое распределение обеспечивает лучшие результаты для общей пропускной способности системы за счет предотвращения экстремально больших или малых распределений. Это, однако, приводит к некоторой потере справедливости распределения ресурсов между пользователями. Очень важно, что в классической формулировке подобно критерию max-min пропорциональное распределение ресурсов не обеспечивает гибкости в выборе точки баланса справедливого распределения пропускной способности. Это критическое рассмотрение является введением для поиска альтернативных целевых функций со свойствами, подобными функции $f(x) = \log(x)$. В первую очередь, это класс оптимизационных задач, в которых целевая функция задается посредством (1.1), известных как выпуклое программирование. Такие задачи существенно сложнее для решения по сравнению с линейным программированием, что представляет проблему для эффективного внедрения метрики справедливого пропорционального распределения для алгоритмов распределения ресурсов в реальном времени, особенно в крупномасштабных и сверхплотных гетерогенных сетях.

1.3.3 Обсуждение

Как следует из предыдущего анализа, сложная топология гетерогенной сети совместно с уникальными свойствами отдельных компонентов ра-

дио технологий требуют пересмотра выбора критерия справедливого распределения ресурсов. В частности, в проводных сетях рассмотренные критерии штрафуют более длинные потоки строже, чем короткие (в терминах числа шагов) при условии, что они конкурируют за ресурсы в пределах определенной области, представляющей взаимный интерес. Поскольку пользователи с более высоким качеством канала могут более эффективно использовать беспроводные ресурсы, эквивалент критерия пропорционального справедливого распределения ресурсов для гетерогенных сетей должен быть основан на некоторой метрике, учитывающей текущее состояние канала пользователей. Далее отметим, что критерий *max-min* пытается справедливо распределить ресурсы настолько, насколько это возможно. Это свойство, полезное в существующих сетях, может быть неподходящим для гетерогенных сетей. В самом деле, текущее расстояние между пользователем и базовой станцией BS (Base Station) в гетерогенной сети может существенным образом влиять на выбор схемы модуляции и кодирования MCS (Modulation and Coding Scheme) и, таким образом, на эффективную скорость обслуживания, обеспечиваемую радио технологией. При использовании критерия *max-min* для пользовательского оборудования с худшими условиями канала компенсация должна осуществляться за счет большего количества ресурсов, выделенного для этого оборудования. Последнее, в свою очередь, предполагает, что попытка обеспечить точное справедливое распределение ресурсов приведет к резкому уменьшению пропускной способности всей сети, что может быть неприемлемо для операторов сетей связи (см., например, [22] для более детальной дискуссии).

Как адекватную альтернативу для учета баланса между сетевыми возможностями и распределением пропускной способности, мы предлагаем использовать основанный на полосе пропускания *max-min* критерий справедливости. Соответственно, вместо того, чтобы справедливо распределять ресурсы между пользовательским оборудованием, мы изначально предлагаем разделить совокупную полосу пропускания доступных технологий доступа как можно более справедливо. Очевидно что пользователи, которые находятся ближе к обслуживающей их базовой станции, получают возможность использовать более высокую скорость передачи, чем те,

которые расположены дальше вследствие использования различных MCS. Однако предложенный критерий не допускает бесконечно малых скоростей для пользователей с плохими условиями канала, и, в тоже время, высокие скорости пользователей с хорошими условиями ограничены возможностями MCS, которые определяются конкретной радио технологией. Таким образом, предложенная метрика выражается в компромиссе между пропускной способностью системы (выгода от естественного динамического выбора MCS) и справедливым распределением скорости передачи данных. Далее, для обеспечения обмена пропускной способности на справедливость и наоборот, мы дополним базовую целевую функцию весовыми коэффициентами. Эти коэффициенты определяются посредством тщательно подобранных функций использующих мгновенные значения спектральной эффективности пользовательского оборудования и, таким образом, позволяют гибко корректировать текущий оперативный режим функционирования гетерогенной сети. Итак, предложенный критерий может рассматриваться как разновидность пропорциональной справедливой метрики, специально разработанной для гетерогенных сетей, когда пользователи, расположенные на больших расстояниях от базовой станции получают больше или меньше ресурсов, в зависимости от выбранных весовых коэффициентов. Важнейшим фактом является то, что как мы продемонстрируем далее, задача оптимизации все еще относится к задачам линейного программирования. Далее предположим, что для эффективного решения используем алгоритмы, чья вычислительная сложность хорошо масштабируется с размерностью задачи, т.е. размером гетерогенной сети и числом пользователей.

1.4 Разработка нового критерия

1.4.1 Трафик и топологическая модель

Спецификация видов пользовательского трафика зависит от его важности в гетерогенной сети, что в свою очередь влияет на выбор целевой функции и, соответственно, алгоритма решения. Для нашего исследования мы предполагаем, что генерируемый пользователями трафик является «жадным» и эластичным (также известен как полностью буферизированный или насыщенный в соответствии со спецификациями 3GPP). Напомним, что «жадный» трафик занимает все доступные радио ресурсы, в то время как эластичность подразумевает адаптивность к распределяемым ресурсам. На практике, эластичный трафик является типовым не только для сессий передачи данных протокола TCP, но и для современных видео и голосовых приложений, реализованных посредством алгоритмов с динамической адаптацией скорости.

Далее рассмотрим характеристики сценария для простой соты в 3-х уровневой конвергентной сети (см. рисунок 1.1.A). Следующие составляющие входят в интересующую нас систему: N пользователей через 3 отдельных яруса сетей радиодоступа (в дальнейшем будем называть их уровни); $M_1 = 1$ базовые станции первого уровня (например, макро соты LTE); M_2 базовые станции второго уровня (например, пико соты LTE); и M_3 базовые станции третьего уровня (например, точки доступа WiFi). В целом, в рассматриваемой структуре есть $M = M_1 + M_2 + M_3$ базовых станций, а также один координирующий узел (например, физический агрегатор, ассемблирующий трафик после его разделения в различных радио технологиях).

Резюмируя, мы делаем следующие предположения:

- пользователь имеет доступ к более, чем одной базовой станции на каждом уровне HetNet;
- «жадный» и эластичный трафик;
- местоположение пользователя известно на координаторе;

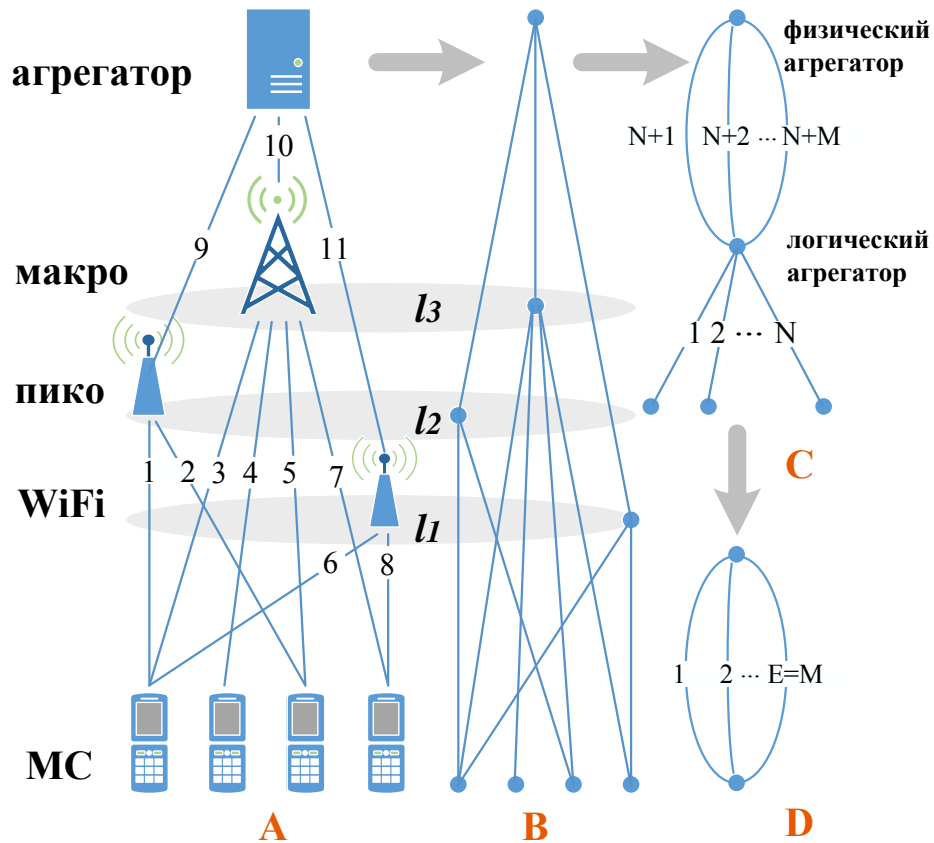


Рисунок 1.1 — Формализация системной топологии HetNet

– достижимая скорость передачи данных в конкретной сети RAN зависит от текущего MCS.

Координатор, который является модулем централизованного распределения ресурсов, физически интегрируется в оборудование базовой станции первого уровня. Следовательно, решения о соответствующем распределении ресурсов принимаются в то время, когда новые пользователи (MC) входят в систему. Пример топологии гетерогенной сети с четырьмя пользователями приведен на рисунке 1.1.В. Далее, мы сконцентрируем внимание на так называемом разветвленном распределении ресурсов, принципы которого предполагают, что любой пользователь одновременно может использовать более, чем один радио интерфейс и разделять свой трафик произвольно между ними.

Анализируя топологию гетерогенной сети, представленную на рисунке 1.1.В, мы видим, что невозможно явно определить скорость передачи данных между пользователями и базовыми станциями. Это связано с тем, что у многих пользователей имеются общие линии для связи с различными

уровнями базовых станций. Кроме того, отметим, что топология является избыточной, поскольку соединение базовых станций с агрегатором в правильно определенной системе должно иметь равную или большую пропускную способность, чем это обеспечивается базовыми станциями по радио интерфейсу. Например, пусть $e = 1, 2, \dots, E$ будут набором линий и используется c_e для обозначения их пропускной способности на физическом уровне радио, и мы видим из рисунка 1.1.B, что $c_9 \geq c_1 + c_2$, $c_{10} \geq c_3 + c_4 + c_5 + c_7$, и $c_{11} \geq c_6 + c_8$. Следовательно, линии c_e , $e = 9, 10, 11$ не дают никаких дополнительных ограничений.

Упрощенная сетевая топология приведена на рисунке 1.1.C, где вышеупомянутые «лишние» линии удалены. Здесь показаны два ключевых дополнительных элемента системы: логический и физический агрегаторы (что согласуется с агрегатором на рисунке 1.1.A). С этой простейшей топологией мы не можем немедленно идентифицировать возможные пути между источником (S) и получателем (D). Однако, дополнение их возможными путями для пары S-D, позволяет нам идентифицировать такие маршруты однозначно. Заметим далее, что мы все еще не можем явно определить скорости передачи данных для линий $e = 1, 2, \dots, N$ для логического агрегатора, хотя они на самом деле представляют собой потоки, реализующие требования пользователей.

Наконец, наша окончательная модель топологии сети HetNet показана на рисунке 1.1.D, где все требования сконцентрированы между двумя узлами: логическим и физическим агрегаторами [24]. Число линий в такой системе равно числу базовых станций $M = M_1 + M_2 + M_3$, в тоже время их пропускные способности равны эффективным пропускным способностям соответствующих базовых станций. Некоторые из этих линий (но не все) являются общими для требуемых пар. В конечном счете, пути для каждого требования являются известными явно, и они однозначно идентифицируют потоки данных, которые должны быть реализованы в системе, для удовлетворения текущих потребностей пользователей.

1.4.2 Формулировка проблемы сетевого потока

Здесь мы впервые рассмотрим простое распределение ресурсов на основе скоростей передачи данных. В этом случае, текущая спектральная эффективность (Spectral efficiency, SE) пользователя не принимается в расчет, а все пользовательские узлы равны в терминах распределения полосы пропускания. Далее, мы расширим эту формулировку, учитывая SE, как часть конструкции пропускной способности. В последнем случае ожидается более справедливое распределение ресурсов, и в тоже время должен быть достигнут компромисс между справедливым распределением и общей производительностью системы. Итак, мы расширяем последний случай за счет управляемого справедливого распределения посредством введения весовых коэффициентов как функций от спектральной эффективности. Стоит еще раз заметить, что нас интересуют разветвленные решения, т.е. такие схемы, когда все радио технологии для пользовательского оборудования могут быть использованы одновременно. Ниже мы специфицируем задачу max-min распределения, основанного на полосе пропускания, и расширенного за счет использования SE пользовательских узлов.

Обозначим как N число требований, которые реализуются в системе. Вспомним, что каждый пользовательский узел ассоциируется с одним требованием. Пометим требования индексом d , $d = 1, 2, \dots, N$. Объем требований измеряется в битах в секунду (bps) и обозначается как

$$h_d, \quad d = 1, 2, \dots, N. \quad (1.2)$$

Заметим, что значения h_d неизвестны заранее вследствие предположения о «жадном» эластичном трафике. Эти величины должны быть определены таким образом, чтобы удовлетворять вполне определенному критерию справедливого распределения ресурсов. Давайте обозначим также как Π_d набор путей для запроса d , т.е.,

$$\Pi_d = \{P_{d1}, P_{d2}, \dots, P_{dP_d}\}, \quad d = 1, 2, \dots, N, \quad (1.3)$$

где P_d есть общее число доступных путей для запроса d . На практике, P_d это число базовых станций на всех уровнях гетерогенной сети, с которыми ассоциирован конкретный пользователь. В нашей заданной топологии эти пути легко доступны как все подмножества P_{dp} , $p = 1, 2, \dots, P_d$, $d = 1, 2, \dots, N$, содержащие, по крайней мере, один элемент, связанный с агрегаторами. Поскольку мы можем избегать использования путей с дополнительными переменными, формулируя задачу исключительно в терминах линий (существует взаимно однозначное соответствие между ними), мы все еще предпочитаем использовать их, чтобы избежать путаницы с обозначениями, характерными для классических задач о сетевых потоках.

Далее, пусть

$$x_{dp}, \quad d = 1, 2, \dots, N, p = 1, 2, \dots, P_d, \quad (1.4)$$

Представляют собой потоки с частью запроса d через путь p . Эти термины известны как *переменные потока*. Для распределения ресурсов, основанного на полосе пропускания, эти переменные потоков измеряются в бит/с. В ином случае, когда SE конечных узлов принимаются во внимание, переменные потоков x_{dp} измеряются в Гц.

Первый набор уравнений, который нам необходимо определить – это так называемые *ограничения на запросы*. Эти ограничения дают уверенность в том, что все запросы h_d полностью учтены при использовании переменных потока x_{dp} , $p = 1, 2, \dots, P_d$. *Ограничения на запросы* условно определены как

$$\sum_{p=1}^{P_d} x_{dp} = h_d, \quad d = 1, 2, \dots, N. \quad (1.5)$$

Далее, будем использовать переменную δ_{edp} , как

$$\delta_{edp} = \begin{cases} 1, & e \cap P_{dp} = e, \\ 0, & \text{иначе,} \end{cases} \quad (1.6)$$

тогда *ограничения пропускной способности* можно записать как

$$\sum_{d=1}^N \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = R_e, \quad e = 1, 2, \dots, E, \quad (1.7)$$

где R_e есть скорости, относящиеся к базовым станциям и измеренные в битах. Пропускная способность при этом ограничена, что гарантирует отсутствие перегрузок линий связи в нашей сети. Заметим, что $\delta_{edp} = 1$, если путь p запроса d использует линию e .

Вводя вектор распределения ресурсов как

$$\vec{h} = (h_1, h_2, \dots, h_N), \quad h_d = \sum_{p=1}^{P_d} x_{dp}, \quad (1.8)$$

мы видим, что проблема оптимизации является лексикографической максимизацией \vec{h} с заданными ограничениями (1.5) и (1.7). Соответствующее решение сразу же дает max-min распределение ресурсов.

Теперь рассмотрим второй случай, когда принимается во внимание роль SE пользовательских узлов. Это в свою очередь требует модификации ограничений и для запросов, и для пропускной способности следующим образом. Пусть s_{dp} есть текущий SE для потока с переменной x_{dp} реализующий часть запроса d через путь p и измеряемый в бит/с на один Гц (бит/с/Гц). Вспомним, что для SE использование справедливого распределения ресурсов для потока x_{dp} зависит от MCS используемой в текущий момент времени. В соответствии с обеспечением max-min распределения ресурсов различных MCS в пользовательских узлах, нам необходимо распределить большую полосу пропускания для UE, имеющих MCS с более низкой скоростью передачи данных. Заметим, что теперь распределение для потоков x_{dp} измеряется в Гц, а не в битах в секунду.

Ограничения по потокам задаются тогда следующим образом:

$$h_d = \sum_{p=1}^{P_d} s_{dp} x_{dp}, \quad d = 1, 2, \dots, N, \quad (1.9)$$

где $x_{dp}s_{dp}$ обеспечивают распределение ресурсов, измеряемое в бит/с.

Ограничения по пропускной способности остаются такими же, как в (1.7) со скоростями передачи данных R_e в полосе пропускания B_e измеренной в Гц. Задача состоит в лексикографической максимизации вектора \vec{h} . Выше сформулированная задача оптимизации (с принятием во внимание SE) обеспечивает max-min распределение скоростей передачи данных для пользовательских узлов. В этом случае, доля полосы пропускания, выделяемая для пользовательского оборудования, будет изменяться пропорционально отношению весов для его MCS. Поскольку целевая задача определяется как поиск справедливого распределения ресурсов при изменчивой пропускной способности, можно заменить неравенства на равенства с ограничениями по пропускной способности. Заметим также, что для ресурсов, выделяемых по расписанию в базовых станциях LTE, и основанных на дискретном распределении ресурсов по блокам, фактическое распределение ресурсов будет слегка отличаться от теоретического прогноза.

1.4.3 Многослойное управление справедливым распределением ресурсов

Рассмотрим теперь модифицированные требования к вышеупомянутой задаче оптимизации, которые основываются на возможности изменения общей производительности системы для справедливого распределения ресурсов [25]. Эти модификации базируются на нескольких специализированных функциях, применяемых к SE пользовательского оборудования. Отметим, что по умолчанию рассматриваемая схема будет выделять ресурсы с максимально возможным справедливым распределением. Это оставит пользователей с высоким SE и лучшей связностью (например, имеющих доступ ко всем трем уровням сети) с недостаточно большой долей полосы пропускания в целях повышения производительности для узлов с плохой связностью и низким MCS. Хотя эта стратегия действительно может оказать положительное влияние на нижние процентиля интегральной функ-

ции распределения скорости передачи данных на одного пользователя, общая производительность системы останется на низком уровне (см. результаты раздела 1.5). Однако мы предлагаем модифицировать SE, s_{dp} посредством введения функции управления $f(x)$ таким образом, что абсолютная разница между пользовательским оборудованием с высокими и низкими SE будет уменьшена. В результате предложенная схема будет назначать большие ресурсы для пользовательского оборудования с лучшими SE с целью добиться компромисса для справедливого распределения ресурсов в пользу общей производительности системы.

Ключевым исследовательским вопросом для предложенной модификации является выбор подходящей функции управления $f(x)$. В диссертации мы рассматриваем две альтернативы: $f(x) = x^\alpha$ и $f(x) = \alpha^x$, где x SE пользовательского оборудования, рисунок 1.2. Первая опция проста для понимания. В самом деле, исходное справедливое распределение ресурсов достигается посредством установления $\alpha = 1$, и в тоже время максимальная производительность наблюдается при $\alpha \rightarrow 0$. Для случая распределения ресурсов на основе скорости передачи данных, когда спектральная эффективность не принимается во внимание ($s_{dp} = 1, \forall d, p$). Однако, выбор второй функции, которая широко используется, например, при сжатии диапазона в радиочастотных усилителях, может потенциально обеспечить лучшую альтернативу вследствие наличия специальной точки балансировки $\alpha = 1$, рисунок 1.2. Значение α меньше, чем 1 позволяет лучше распределить ресурсы для пользователей с высокими SE, таким образом увеличивая производительность системы. Далее, выбор $\alpha = 1$ обеспечивает распределение ресурсов, основанное на скорости передачи данных, что способствует увеличению производительности системы. Для $\alpha > 1$, схема имеет тенденцию к улучшению справедливого распределения ресурсов, в конечном счете достигая случая для справедливого распределения ресурсов на основе полосы пропускания. Дальнейшее увеличение α , однако, не дает роста в справедливом распределении ресурсов, как это подчеркивается дальше. В любом случае, использование $f(x) = \alpha^x$ может быть более гибким по сравнению с $\alpha \rightarrow 0$, позволяя ориентировать результаты на весь диапазон управления для гетерогенных сетей. Важно отметить, что пред-

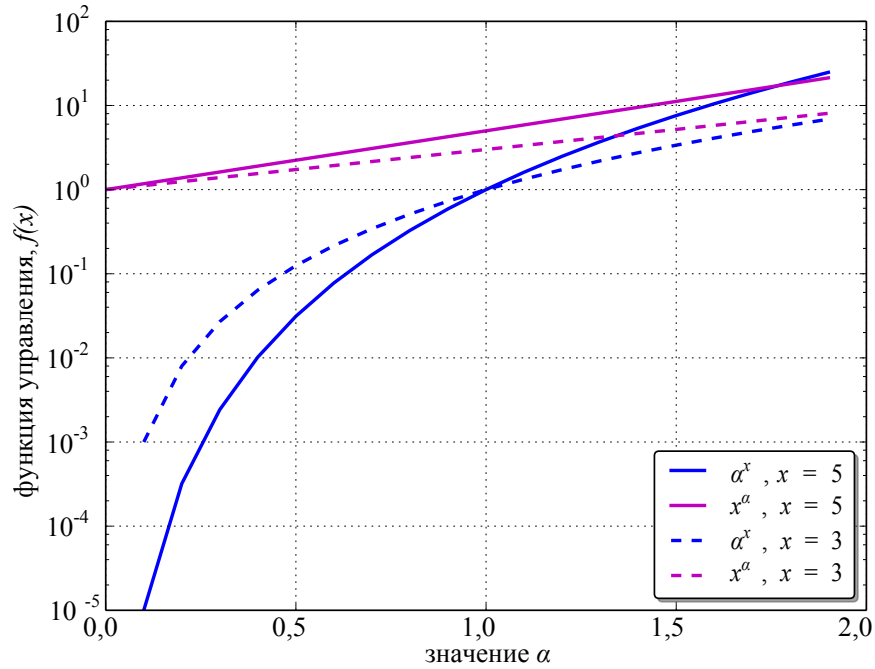


Рисунок 1.2 — Сравнение различных функций управления

ложенные модификации не являются альтернативными для используемого класса оптимизационных задач, т.к. они применяются к x_{dp} только в части весовых коэффициентов.

1.4.4 Алгоритм решения

Решение max-min оптимизационной задачи является расширением алгоритма для задачи нахождения единственного пути, известной по, например [26]. В нашей работе мы используем Matlab для разработки соответствующего алгоритма распределения ресурсов. Эта задача классифицируется как задача линейного программирования, содержащая как первый этап исходное «заполнение водой» для выявления максимального распределения, которое может быть предложено для всех потоков одновременно, и последующее уточнение распределений для выбранных потоков. Первая стадия может иметь множественные решения как противовес единственному решению распределения фиксированных путей. Вследствие этого нам необходимо выполнять дополнительно так называемые неблокирующие те-

сты для достижения результатов max-min распределения. Для более детального взгляда на эту проблему можно проанализировать структуру алгоритма 1.

Предложенный алгоритм используется на макро сотовой базовой станции LTE в то время, когда новый пользователь прибывает в систему или существующий пользователь изменяет состояние своей связности. Причина этого в том, что max-min справедливое распределение ресурсов для всех пользователей зависит от доступности беспроводных интерфейсов для каждого конкретного пользовательского оборудования. Также важно отметить, что исходящие сессии не прерываются вследствие прихода новых пользователей, поскольку только доступные распределения ресурсов пересчитываются и сообщаются пользовательскому оборудованию. По этой же причине нет необходимости решать основные задачи распределения ресурсов, когда не наблюдается изменений в связности пользователей. В заключение подчеркнем, что поскольку решаемые задачи относятся к классу задач линейного программирования, вычислительная сложность для макро сотовой базовой станции LTE вполне приемлема даже для сценариев сетей со сверхвысокой плотностью.

1.4.5 Эвристические подходы

В этом подразделе мы предлагаем две простые эвристические стратегии распределения ресурсов, которые будут служить в качестве рейтинговых тестов для подхода, предложенного выше. Первая схема – механизм централизации для пользовательского оборудования, когда пользовательское оборудование независимо решает, как много ресурсов должно быть выделено на каждом сетевом уровне. Важно отметить, что в этой процедуре пользователь не знает, сколько ресурсов в общем доступно на каждом уровне сети и как пользовательское оборудование может локально конкурировать за каждый доступный радио интерфейс. Кроме того, предположим, что «жадный» эластичный тип трафика не предоставляет лучших решений

- 1 Модификация изначального вектора SE (s_{dp0}) для получения обменного соотношения пропускной способности к справедливости.
 $s_{dp} = f(s_{dp0})$, где $f(s_{dp0})$ – выбранная функция (например,
 $f(x) = \alpha^x$)
- 2 Оценка Δ как решения задачи линейного программирования
- 3 Задание $n = 0$, $\Delta^{(0)} = \Delta$
- 4 Определение $Z_0 = \{1, 2, \dots, N\}$.
- 5 А: Задание $n = n + 1$ и оценка пропускной способности для каждого d
- 6 Пользователи, распределение ресурсов для которых может быть увеличено, обозначаются как $Z_n \subseteq Z_0$
- 7 **if** $Z_n = \emptyset$ **then**
- 8 | go to B
- 9 **else**
- 10 | $Z_{n-1} = Z_{n-1} \setminus Z_n$.
- 11 **end**
- 12 Решение следующей задачи линейного программирования:

$$\max \Delta,$$

При условии что:

$$\sum_{p=1}^{P_d} s_{dp} x_{dp} = h_d, \quad d = 1, 2, \dots, N,$$

$$\Delta - h_d \leq 0, \quad d \in Z_n,$$

$$\Delta^{(k)} - h_d \leq 0, \quad k = 0, 1, \dots, n - 1,$$

$$\sum_{d=1}^N \sum_{p=1}^{P_d} \delta_{edp} x_{dp} = B_e, \quad e = 1, 2, \dots, E,$$

- 13 Задание $\Delta^{(n)} = \Delta$
- 14 go to A
- 15 В: Применение функции $f_1(x)$ из шага 1.
- 16 Финальный вектор пропускных способностей h_{d1} будет иметь форму
 $h_{d1} = h_d s_{dp0} / s_{dp}$.

Алгоритм 1: Алгоритм взвешенного max-min для решения задачи распределения ресурсов в сети доступа мульти радио.

для распределения ресурсов. Существует два ограничения, препятствующие внедрению простой «жадной» схемы: пользовательское устройство возьмет так много ресурсов базовой станции, как позволяет планировщик базовой станции. В оставшейся части текста мы будем называть этот метод эвристикой максимального использования.

Второй подход является примером распределения ресурсов с сетевой поддержкой: беспроводная система помогает пользовательскому оборудованию сделать выбор более интеллектуальным путем обеспечения информацией об обслуживании близлежащих объектов. Такие алгоритмы в настоящее время находятся в исследовательских проектах 3GPP для версии 12 и последующих для LTE (например, [27]). В нашем исследовании мы реализуем простой пример алгоритма, в котором пользователям обеспечивается информация только о предпочтительном пороге ассоциации с каждым уровнем HetNet. С помощью соответствующей регулировки этого порога сеть может уменьшать/увеличивать эффективную дальность покрытия пико сот LTE и точек доступа WiFi для лучшей выгрузки трафика пользовательского оборудования от/к макро соте LTE.

Кроме того, мы предполагаем, что пользовательское оборудование может одновременно подключиться только к одному (например, WiFi или пико сота LTE), который рассматривается в продолжающейся стандартизации 3GPP. Зная об этом ограничении, сначала пользователь пытается подключиться к точке доступа WiFi, что является типовым решением для высокой скорости передачи данных. Если обеспечиваемый сетью порог ассоциации не достигается на WiFi, пико сотовая базовая станция LTE является следующей в очереди попыток ассоциации с сетью. В конечном итоге, если ни одна из попыток не заканчивается успешно, пользовательскому оборудованию остается только соединиться с макро базовой станцией LTE. Как правило, выделенные пороги могут быть использованы индивидуально для каждой базовой станции. В диссертации мы вначале выбираем порог, основанный на средних значениях SE (s_{dp}) пользовательского оборудования на каждом уровне. Далее, мы рассматриваем различные пороги в соответствии с компромиссом между справедливым распределением ресурсов и

общей производительностью системы. В дальнейшем мы будем именовать этот эвристический алгоритм как «WiFi-предпочтительный».

1.5 Численные результаты

1.5.1 Параметры внешней среды

Типовой сценарий HetNet, который мы рассматриваем в диссертации, включает следующие системные составляющие: простая макро сотовая базовая станция BS (называемая eNodeB), несколько пико сотовых базовых станций eNodeB и точки доступа WiFi (WiFi APs), а также пользовательские узлы с различными мульти радио технологиями (WiFi и LTE). Следуя обсуждениям проблемы в 3GPP [28;29], пико сотовые базовые станции, точки доступа WiFi, и UE размещаются в зоне покрытия макро сотовой базовой станции eNodeB. Каждый пользователь предположительно имеет несколько активных радио соединений, что зависит от доступных радио технологий и воспринимаемого уровня сигнала (он должен превышать уровень ассоциации для AP/eNodeB). Последнее оценивается для различных моделей потерь в тракте радиопередачи.

В соответствии со стандартами 3GPP [28;29], для пользователей макро сот дистанционно зависимые потери определяются моделью МСЭ для макро каналов в городской среде. Для WiFi APs и пользователей пико сот LTE потери аппроксимируются с использованием модели МСЭ для микро городских каналов, специфицированной в [30] со следующими параметрами: средняя высота зданий равна $h = 20$ и средняя ширина улицы также равна $W = 20$. Кроме того, для более точного моделирования потерь в среде распространения требуются подходящие модели медленного и быстрого замирания сигналов. Модель медленного замирания принимается во внимание в том случае, когда мы рассматриваем нашу сеть в дискретные моменты времени $n\Delta t$, $n = 0, 1, \dots$. Эффект быстрого замирания проявляется

Основные параметры и их значения

Параметр	Значение
Частотная конфигурация LTE/WiFi	10 МГц FDD / 20 МГц
Конфигурация сценария	1 Макро сота, несколько LPN
Модель канала Макро/LPN-UE	ITU UMa/UMi [30]
Усиление антенны Макро/LPN	17/6 dB
Максимальная мощность Макро/пико/WiFi	43/23/20 dBm
Максимальная мощность UE	23/20 (LTE/WiFi) dBm
Схема контроля мощности LTE/WiFi	макс. мощность
Высота антенны UE/макро/LPN	1.5/25/10 m
UE noise figure/потери кабеля антенны	5 dB / 0 dB
Модель трафика	полный буфер
Тип распределения LPN/UE	равномерный [28]
Минимальная дистанция LPN/UE-макро	> 75/35 m [29]
Минимальная дистанция LPN/UE-UE	> 40/10 m [29]
Количество экспериментов	1000

на протяжении длительных периодов времени, когда нас интересуют средние величины метрик функционирования сети. Следовательно, параметры, относящиеся к медленному и быстрому замираниям, могут быть включены в модель с использованием постоянных граничных значений (равных стандартным отклонениям, [30]), которые добавляются к порогам ассоциации с основными радио технологиями. В диссертации мы также аппроксимируем текущие уровни интерференции посредством подобных методик. Обсудим эти важные аппроксимации более детально ниже. Все базовые параметры сведены в таблицу 2.3, в которой термин низко потребляющий узел LPN (low-power node) относится и к WiFi AP, и к пико сотовой базовой станции LTE.

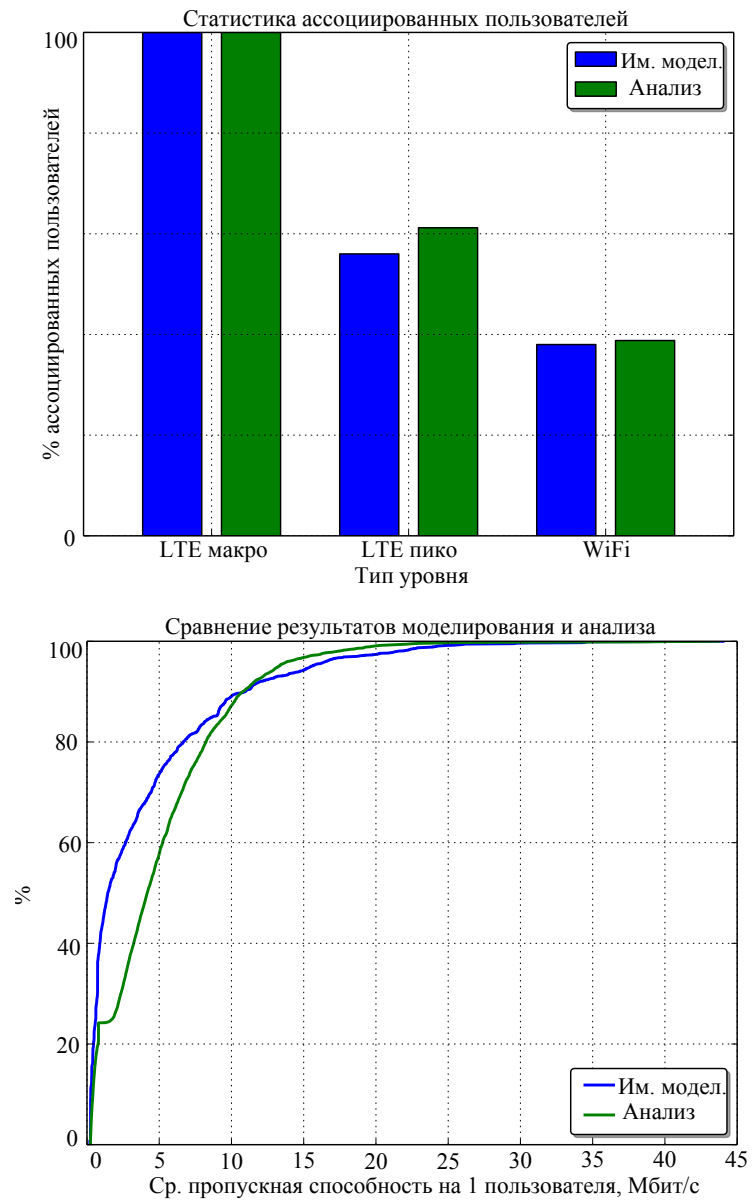


Рисунок 1.3 — Калибровка аналитической модели на базе анализа подключений (сверху) и функции CDF (снизу)

1.5.2 Проверка аналитических условий

Предложенные аналитические модели включают некоторое число простых предположений. Для верификации этих предположений в отношении интересующих нас метрик (включая производительность и справедливое распределение ресурсов) нужно протестировать модель в более реалистичных условиях гетерогенной сети. Для этих исследований была использована система имитационного моделирования WINTERsim, которая учитывает все существенные детали функционирования HetNet и тщательно откалибрована в публикациях автора [31; 32]. Текущая открытая версия WINTERsim доступна на сайте [33]. Ниже мы сравним результаты функционирования простых эвристических стратегий, рассмотренных выше, с результатами, полученными как с использованием аналитического оптимизационного инструментария, так и с результатами, полученными с помощью симулятора. Сопоставление наблюдаемого функционирования эвристических схем должно подтвердить то, что поведение предложенной аналитической структуры повторяет по существу результаты, полученные при помощи симулятора.

Во время калибровочной фазы эвристическая схема с максимальным использованием была применена как для аналитической модели, так и для условий симулятора с ограничениями сценария тестирования в виде фиксированного покрытия и пропускной способности. Исходными для процесса тестирования параметрами являются число пользователей (60 единиц пользовательского оборудования равномерно распределены в зоне макро соты), число пико сотовых базовых станций LTE и точек доступа WiFi (5 узлов каждого типа с равномерным распределением размещения в зоне макро соты), а также условия передачи и каналов. Одним из ключевых индикаторов функционирования, обсуждаемых ниже, является справедливое распределение ресурсов. Для оценки результата справедливого распределения ресурсов для отдельного набора распределения будем использовать

индекс Джейна, определяемый как

$$J = \left(\sum_{i=1}^N x_i \right)^2 \frac{1}{N \sum_{i=1}^N x_i^2}, \quad (1.10)$$

где x_i – производительность пользовательского оборудования и N – число пользователей.

Результаты моделирования, усредненные для 100 реализаций (реальное время составляло 4 секунды для каждой реализации в симуляторе) продемонстрированы на рисунке 1.3, где левый график показывает статистику ассоциированных пользователей (число пользователей, соединенных с каждой радио технологией), в то время как на правом графике приведена средняя пропускная способность пользовательского оборудования. Итак, мы подтвердили, что наблюдается разница в значениях метрик для двух подходов в 5%, которая является достаточно точной, учитывая детали расчета для симулятора. Наблюдаемые меньшие отклонения в основном связаны с отсутствием учета межсотовой интерференции в аналитической модели.

В то время как построение пико сотовой инфраструктуры LTE осуществляется в соответствии со стандартами 3GPP [28; 29], которые должны обеспечивать определенное минимальное расстояние между базовыми станциями пико сот LTE, интерференция между близлежащими узлами все еще не является ничтожно малой [34]. Наблюдая CDF функцию пропускной способности, мы отмечаем небольшое отклонение в точке 25%. Такое поведение объясняется разницей в производительности пользователей, имеющих только макро соту LTE для связи, и теми, которые имеют дополнительные радио технологии для ассоциации. Заметим, что в симуляторе этот эффект сглажен вследствие случайной картины интерференции. Подобный эффект сглаживания может наблюдаться в случаях, когда пользовательское оборудование с лучшей связностью (например, ассоциированное со всеми тремя уровнями в одно и то же время) не имеет преимуществ в производительности по сравнению с пользовательским оборудованием, имеющим связь только с макро сотой, вследствие большой интерферен-

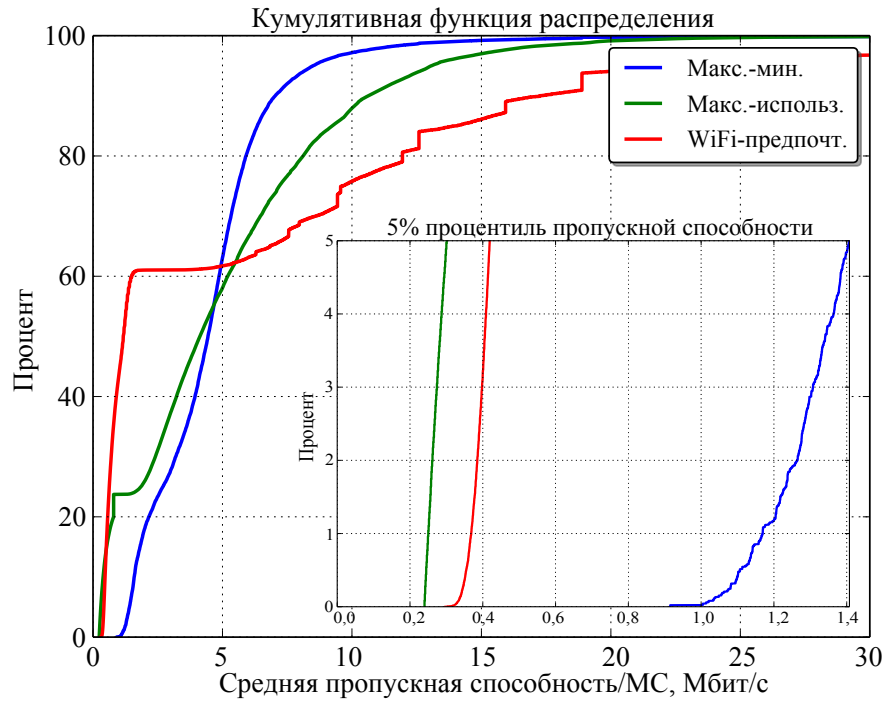


Рисунок 1.4 — Функция CDF (аналитика)

ции на уровне WiFi и/или пико уровнях. В заключении, продемонстрируем результаты, которые позволяют сделать вывод о том, что разработанный инструмент аналитического моделирования является достаточно точным, чтобы оценить влияние предлагаемых стратегий на распределение ресурсов.

1.5.3 Анализ эффективности

Сравним теперь эффективность предложенной аналитической схемы распределения ресурсов, основанной на полосе пропускания, ($f(x) = s_{dp}^\alpha$ with $\alpha = 1$) с простыми эвристическими процедурами. Результаты такого сравнения проиллюстрированы на рисунке 1.4 и рисунке 1.5. В частности, рисунок 1.4 представляет функцию распределения производительности и 5%-процентиль для всех рассмотренных стратегий. Хотя на первый взгляд, средняя производительность почти равна для всех трех альтернатив, можно видеть, что разница в справедливом распределении ресурсов для 5%-процентиля существенна. Для WiFi-предпочтительной схемы наибольшая

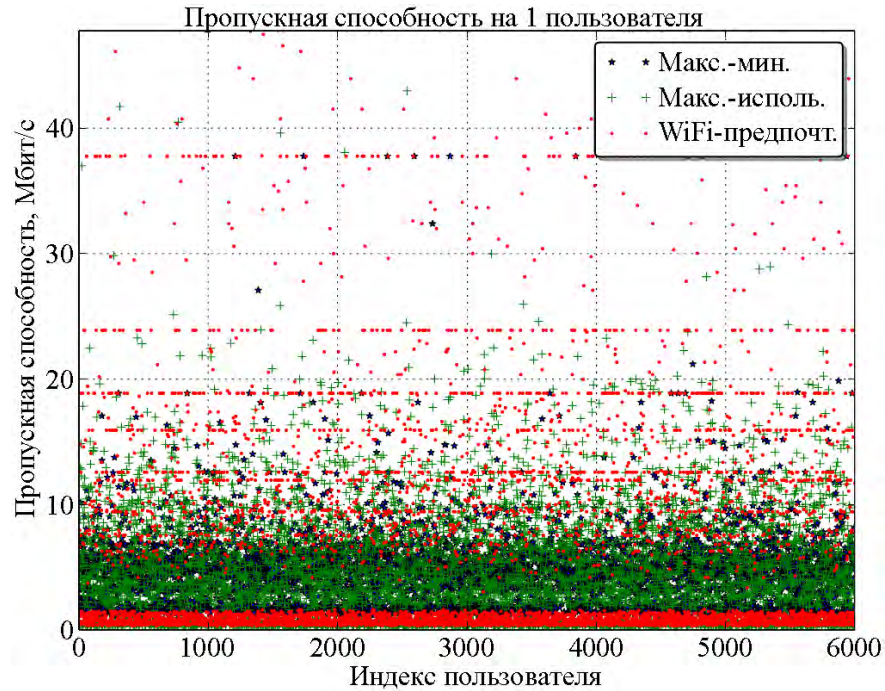


Рисунок 1.5 — Диаграмма пропускной способности (аналитика)

пропускная способность достигается устройствами, имеющими связность с WiFi, в то время как пользователи пико сот и макро сот LTE имеют существенно более низкую производительность (разница достигает 60%). Кроме того, для такого механизма распределения неравномерные изменения в кривой CDF могут быть отмечены только для WiFi пользователей. Этот эффект вызван отсутствием разветвления потока (бифуркацией) для каждого пользовательского оборудования, использующего простую радио технологию в конкретный момент времени. Предложенная схема обеспечивает лучший 5% перцентиль производительности, что однако достигается некоторой деградацией средней производительности на одно пользовательское оборудование.

Более детально, рисунок 1.5 представляет значение производительности на одно пользовательское оборудование для каждого пользователя в виде диаграммы распространения. Эта иллюстрация показывает, что красные точки (WiFi-предпочтительная схема) в наибольшей степени сконцентрированы в верхней и нижней частях графика, вследствие неразветвленной природы распределения ресурсов, как это отмечено выше. Наоборот, зеленые плюсы (схема максимального использования) расположены почти везде, в то время как синие плюсы (max-min схема) имеют тенденцию груп-

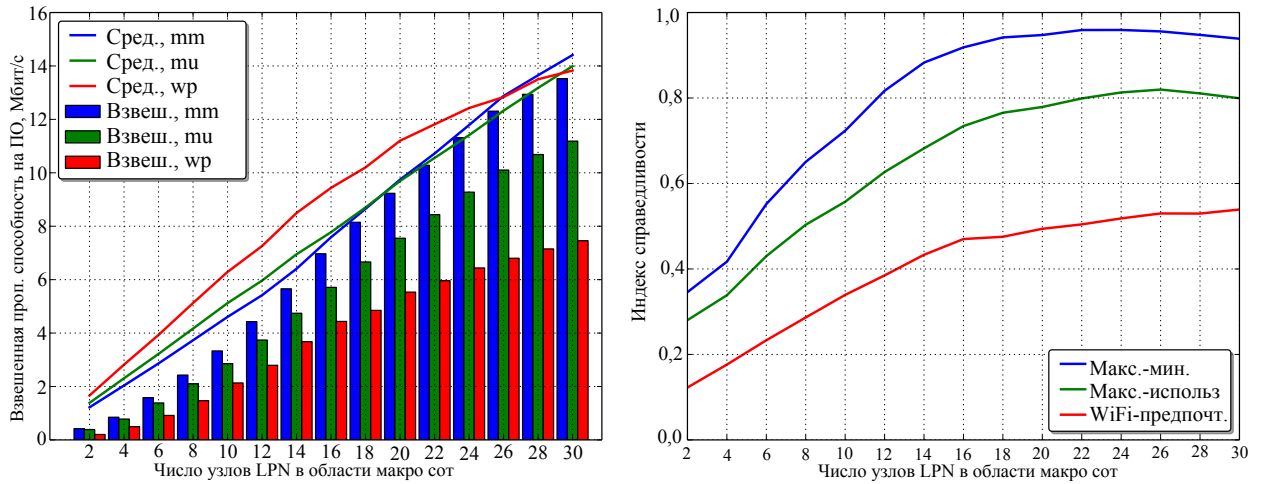


Рисунок 1.6 — Взвешенная справедливая и средняя пропускная способность (слева), индекс Джайна (справа) для различных сценариев

пироваться около центра графика, получая таким образом наилучшее справедливое распределение ресурсов и 5%-квантиль max-min справедливого распределения ресурсов, основанного на полосе пропускания.

Для лучшего понимания влияния пропускной способности и связности на эффективность предложенных алгоритмов мы также оценим поведение рассматриваемых схем распределения ресурсов для различных значений плотности инфраструктурных узлов на различных уровнях гетерогенной сети. Результирующее справедливое распределение ресурсов и производительность в расчете на одно пользовательское устройство для всех трех схем приведены в сравнении на рисунком 1.6, где количество узлов LPN LTE и WiFi варьируется. Анализируя полученные результаты, можно отметить, что когда связность пользователей и пропускная способность всей сети малы (например, почти все пользовательское оборудование использует только макро соты LTE) степень несправедливого распределения ресурсов велика. Также можно отметить, что среди рассмотренных схем нет стратегии, которая может улучшить такую ситуацию.

С увеличением связности пользователей, средняя производительность на единицу пользовательского оборудования при WiFi-предпочтительной схеме становится значимо выше по сравнению с другими стратегиями. Однако этот эффект существенен только для части пользователей, связанных с WiFi, которые получают наибольшую производительность от системы. В то же время, доля этой производительности

сти (множественная пропускная способность по индексу Джейна) WiFi-предпочтительной схемы все еще мала как результат очень плохого распределения ресурсов в такой системе. Однако, если связность пользователей увеличивается, max-min механизм формировать лучшие условия с точки зрения и средней производительности, и справедливого распределения ресурсов, в конечном счете достигая и даже опережая и схемы максимального использования, и WiFi-предпочтительные решения. Другой интересный вывод может быть сделан при анализе поведения индекса Джейна (правый график на рисунке 1.6). Как уже было отмечено, однажды уже используя степень связности, получить дальнейшее улучшение справедливого распределения ресурсов не представляется возможным. Мы называем этот эффект «насыщением справедливостью» и отмечаем, что он пригоден для всех рассмотренных схем. Это явление является следствием близкой к идеальной связности (когда почти каждое пользовательское оборудование может получить доступ к базовым станциям всех трех уровней) и равномерного распределения пользовательского оборудования и базовых станций в зоне покрытия макро соты. Абсолютная величина насыщения справедливостью также зависит от деталей конкретной стратегии распределения ресурсов.

1.5.4 Балансировка производительности и справедливого распределения ресурсов

В этом подразделе мы сконцентрируем внимание отношении эффективности схем max-min и максимального использования для лучшего управления достижением компромисса между производительностью системы и справедливым распределением ресурсов. Напомним, что в случае max-min справедливого распределения ресурсов мы можем использовать две альтернативные весовые функции для текущих значений SE пользователей. На рисунке 1.7 подчеркивается баланс между справедливым распределением ресурсов и производительностью на одно пользовательское обо-

рудование для max-min стратегии распределения ресурсов и двух различных функций управления α^x и x^α . Как и ожидалось, в определенном диапазоне α эффективность max-min схемы для обеих функций почти идентична. Это наблюдается для $f(x) = s_{dp}^\alpha$ с $\alpha \in (0,1)$.

Однако, мы также напоминаем, что функция управления $f(x) = \alpha^x$ имеет больший диапазон величин α , потенциально предоставляя таким образом достижение более высокой производительности. В дополнение к сказанному, рисунок 1.7 показывает предельные значения для реального управления вследствие ограничений по пропускной способности и покрытию. Например, может быть неосуществимо достичь какой-либо средней производительности на одно пользовательское устройство, делая α меньше, чем 0.1 для $f(x) = \alpha^x$, или в качестве рассматриваемой схемы попытаться распределить больше ресурсов на пользовательское оборудование с высоким SE, что может быть невозможно вследствие их неэффективного использования. Кроме того, любое последующее уменьшение α приведет только к деградации общей эффективности системы. Подобное поведение наблюдается, когда рассматриваемые функции применяются к увеличению справедливости, например, делая α в $f(x) = \alpha^x$ больше, чем 1.5. В этом случае, препятствием являются пределы покрытия, и дальнейшее увеличение в α приведет к чрезмерному распределению ресурсов на пользовательское оборудование с низкой связностью. Это, в свою очередь, является расточительным и несправедливо по отношению к пользовательскому оборудованию, расположенному в «хороших» местах.

В заключении, сравним эффективность управления для нашей схемы взвешенного max-min с «жадной» полосой пропускания и эвристического подхода с максимальным использованием для двух различных условий покрытия и пропускной способности. В этих экспериментах мы будем использовать мульти радио технологическую гетерогенную сеть с целью идентификации доступных режимов эксплуатации таких сетей. Соответствующие результаты показаны на рисунке 1.8, где функция управления для взвешенного max-min управления выбирается в виде $f(x) = \alpha^x$. В первом сценарии пять точек доступа WiFi (40% от покрытия макро сотой LTE) и пять пико сотовых базовых станций LTE (60% от покрытия макро сотовой станцией

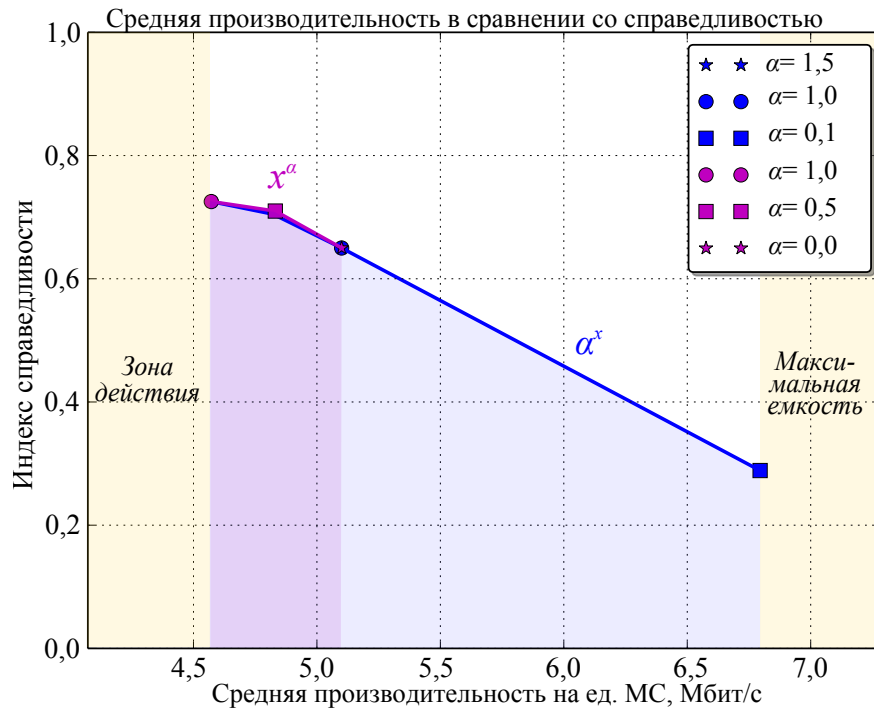


Рисунок 1.7 — Средняя пропускная способность в зависимости от справедливости

LTE) расположены в зоне действия макро соты. Вторым сценарий предусматривает развертывание сети сверхвысокой плотности, когда 15 узлов LPN каждого типа расположены в той же зоне, покрывая таким образом почти 100% уровней и пико сот LTE, и WiFi.

В рассматриваемых сценариях в дополнение к централизованному max-min решению возможно также приспособить справедливое распределение ресурсов за счет WiFi-предпочтительной схемы путем изменения установки порога ассоциации очень низким (более высокая справедливость) или очень высоким (большая пропускная способность). Однако, как видно из рисунка 1.8 такое приспособление возможностей эвристического механизма дает значительно более скромные результаты, чем наш централизованный взвешенный max-min алгоритм, при этом справедливость распределения ресурсов эвристического подхода лежит ниже. В случае сверхплотной сети max-min управление теряет возможный диапазон значений – близкое к максимальному отношение справедливость/пропускная способность уже достигнуто с помощью параметров по умолчанию (распределение ресурсов, основанное на полосе пропускания), и нет необходимости в дополни-

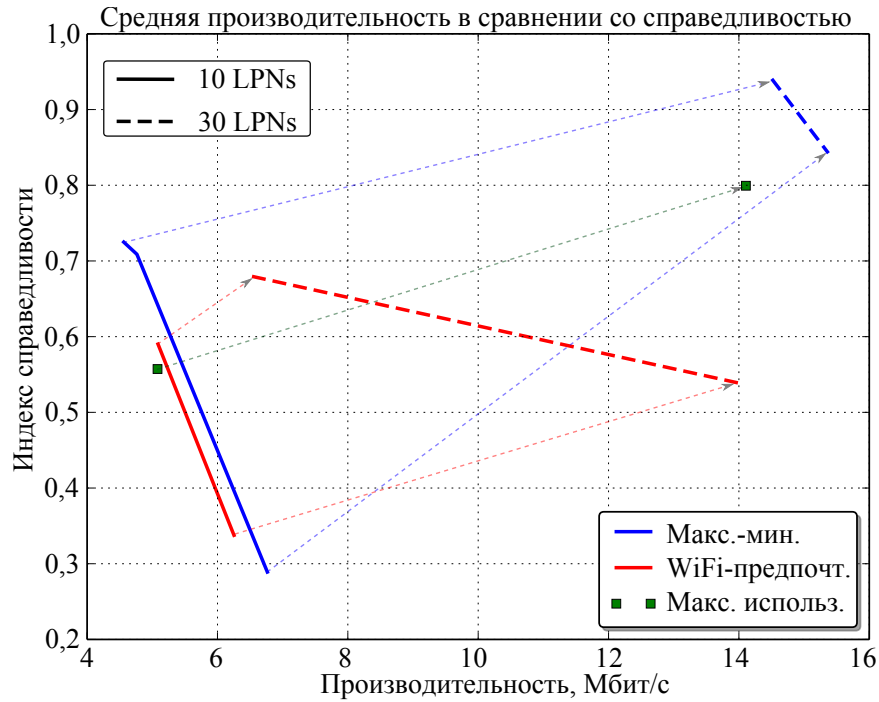


Рисунок 1.8 — Диапазоны работы различных схем

тельных регулировках. Другой интересный эффект наблюдается, когда для пользовательского оборудования централизованно применяется схема максимального использования. В условиях сверхвысокой плотности размещения узлов сети этот подход становится лучше, чем WiFi-предпочтительная стратегия с сетевой поддержкой. Причина этого состоит в том, что такой сценарий даже в простейшей разветвленной схеме превосходит большинство неразветвленных решений.

1.6 Гетерогенная облачная сеть радиодоступа H-CRAN

В зависимости от эффективности ограничения транзитных соединений, различные уровни координации в архитектуре 5G гетерогенных сетей HetNet могут быть возможными. Например, если доступна магистральная сеть только мобильных операторов, координация посредством архитектуры «якорь-бустер» может быть использована, при этом «якорем» будет являться базовая станция макро соты, обеспечивая управление сетью и маленькими мульти радио сотами «бустерами» для передачи данных поль-

зователей посредством возможностей выгрузки трафика. Как альтернатива, в случае магистральной сети с высокой пропускной способностью и малыми задержками (например, оптоволоконной), сигналы с низко потребляющих небольших сот могут быть приняты и обработаны на удаленной централизованной платформе. Эта архитектура, получившая название облачной, сети радиодоступа Cloud RAN становится предпочтительной у сетевых операторов, которые все чаще отдают предпочтение сетям с преобладанием оптоволоконных систем и недорогим беспроводным сетям прямого доступа, в первую очередь для сверхплотных гетерогенных сетей, покрывающих зоны с высокой интенсивностью запросов трафика. Сегодня, когда до 80% капитальных затрат операторов сотовых сетей используется на сети радиодоступа, концепция облачной сети радиодоступа Cloud RAN позволяет существенно снизить капитальные/эксплуатационные расходы, а также значительно сократить потребление энергии беспроводной сети [35].

В Cloud RAN, удаленный головной радиоузел RRH (Remote Radio Head), который является простым низко потребляющим устройством, использует высокоскоростные линии прямого доступа, компрессирует и направляет модулированные сигналы от мобильного пользовательского оборудования на централизованный узел BBU (Base Band Unit), функционирующий как программируемый ретранслятор. Ограничения на пропускную способность прямого доступа накладывают фундаментальное ограничение на функционирование такой системы и передовых решений обработки сигналов [36], совместно с динамическим управлением радио ресурсами [2] требуется поддерживать приемлемый уровень производительности Cloud RAN. Доступная пропускная способность Cloud RAN также зависит от (i) практических ограничений для магистральной сети [37] и соответствующей оптимизации [38], (ii) используемых стратегий ассоциации исходящих линий RRH [39] и соответствующих ограничений по внедрению и сложности использования радио ресурсов [40], (iii), а также от используемых децентрализованных диаграммообразующих алгоритмов [41] и крупномасштабного распределения мощности и выбранных антенных схем [42].

Совсем недавно, концепция гетерогенного облака для радиосетей доступа гетерогенной облачной сети радиодоступа Heterogeneous Cloud RAN

(H-CRAN) была предложена [17] как экономически эффективное решение для дальнейшего улучшения доступной выгоды кооперации в гетерогенных сетях посредством их комбинирования с «облачной обработкой сигналов». Удобно располагаясь между гетерогенной сетью и облачными платформами, H-CRAN наследует привлекательные преимущества обеих сфер, способствуя уменьшению интерференции, увеличению масштабируемости и возможностями по управлению ресурсами радиосвязи за счет кооперации методов обработки сигналов и сетевых технологий. Соответственно, управление ресурсами низко потребляющих узлов LPN продвигается в направлении создания виртуальных базовых станций, которые являются частью процессорной мощности, расположенной в пуле физических BBU облачного сервера. Резюмируя, отметим, что в то время как технологические возможности и принципы построения ядра H-CRAN формируются в направлении создания коммерческих H-CRAN для систем 5G [17], основные исследовательские задачи состоят в анализе теоретической эффективности и оптимального распределения ресурсов для понимания потенциала использования этого инновационного предложения.

Базируясь на знании этой технологии и последних разработок 3GPP в области стандартизации, мы сфокусируем внимание на проблеме кооперативного управления радио ресурсами в 5G H-CRANs посредством создания комплексной методологии оптимизации производительности H-CRAN в реальном масштабе времени. Наше предлагаемое решение позволяет динамически управлять количеством ресурсов, размещенных у конечных пользователей с помощью двух альтернативных метрик интереса, называемых соответственно справедливым распределением ресурсов всех доступных сетей радиодоступа и общей производительностью системы. Мы специально сконцентрируем внимание на динамическом управлении системами H-CRAN для достижения гибкого компромисса между этими метриками. Далее, мы тщательно сравним эффективность механизмов распределения ресурсов для сетевого централизованного, с поддержкой сети и централизованного для пользовательского оборудования механизмов в характерных для H-CRAN условиях с различными уровнями доступа к LTE/WiFi, а также сделаем обзор нашей концепции для доказательства ее обоснованности.

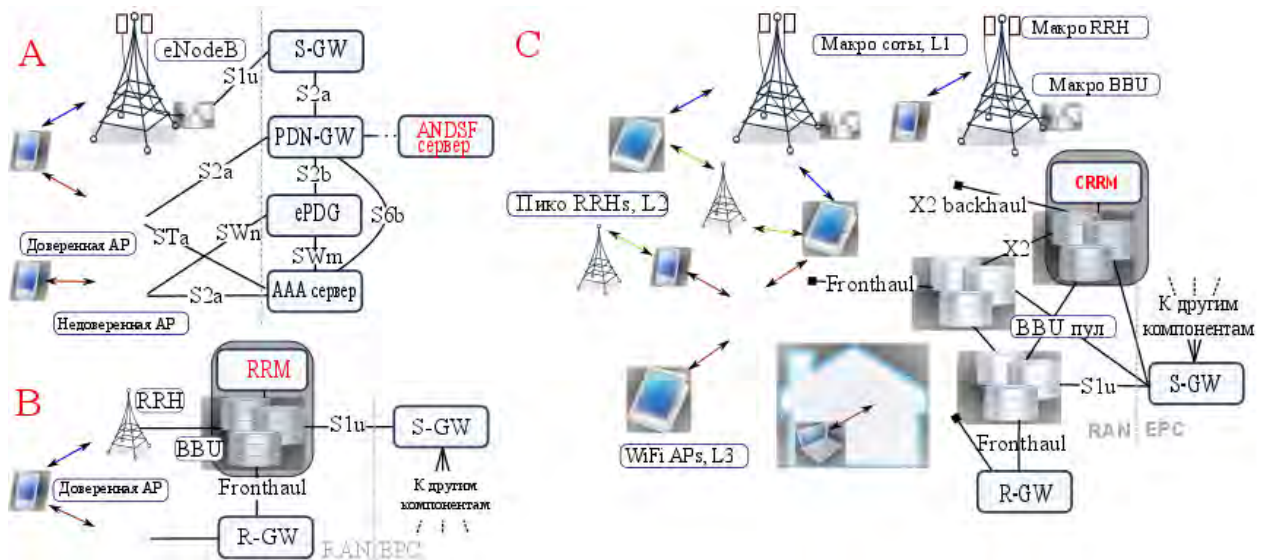


Рисунок 1.9 — Примеры развертывания сети H-CRAN

H-CRAN технология все еще далека от реального создания, но текущие работы по стандартизации в 3GPP положили начало использованию мульти радио технологий как части операторского управления мульти радиосетями. Продолжим далее с кратким обзором относящихся к этому направлению работ.

В широкой координационной модели версии Release 8 вводится концепция открытой сети доступа и функции выбора ANDSF (Access Network Discovery and Selection Function), которая используется для управления взаимодействием между локальными беспроводными сетями WLAN и сетями 3GPP с помощью сервера политик ANDSF с ядром сети. При этом, оператор может специфицировать относящиеся к статическим политикам открытие и использование ресурсов WLAN для сети, оставляя задачу выбора сети пользовательскому оборудованию, которое в своих решениях определяет факторы локальной среды и динамически изменяющиеся условия радиоканалов. Основанная на функции ANDSF архитектура показана на рисунке 1.9.A. Однако, централизованные решения для пользовательского оборудования являются субоптимальными, поскольку пользовательское оборудование не знает об условиях, в которых находятся линии, и о требованиях к радио ресурсам других пользователей, также их использующих. Кроме того, узлы мульти радио сети с созданием HetNet обладают очень ограниченными знаниями о радио ресурсах друг друга, что еще больше

снижает эффективность использования ресурсов радиосвязи. Более того, управление мобильностью между WLAN и 3GPP линиями осуществляется в ядре сети (как правило, в P-GW), делая обмен трафика между WLAN и 3GPP линиями очень дорогим при адаптации к динамическим изменениям условий для линий.

В Release 12 были предприняты работы по решению проблем взаимодействия 3GPP/WLAN (TR 37.834), изложенные выше. Для частной координации с радио сетью доступа сотовой базовой станции eNB стал доступен набор порогов, позволяющий соотнести качество линии и загрузку WLAN, что дает возможность лучшего обмена трафиком между WLAN и сетями доступа 3GPP. Однако, координация, требуемая в сети для набора подходящих порогов, до сих пор не специфицирована. Однако, поскольку мобильный якорь для линий WLAN/3GPP все еще находится в ядре сети, это не дает возможности быстрого эффективного обмена трафиком. Ранее, мы рассмотрели механизмы на основе этого варианта интеграции.

Текущие усилия в 3GPP ориентируются на тесную интеграцию WLAN в 3GPP RAN. Существующие предложения стремятся к использованию WLAN в качестве вторичного транспорта, с якорем в базовой станции eNB в пределах 3GPP RAN. Предложенная архитектура расширяет возможности Release 12, предусматривая двойное использование системы «якорь-бустер», введенное для малых сот 3GPP, также, как и существующая структура агрегации в 3GPP включает в себя не-3GPP радио технологии, например, доступ WLAN. Ожидается, что 3GPP может рассмотреть такие архитектурные решения при стандартизации следующих версий.

Также ожидается, что 3GPP будет рассматривать стандартизацию интерфейса между базовой станцией eNB и WLAN APs для не совместимых WLAN/3GPP. Заметим, что если эта адаптация будет сделана 3GPP, последнее интегрирует архитектуру контроля сети с расширением LTE-основанной схемы «якорь-бустер» и сделает управление координацией радио ресурсов также доступным в не-3GPP WLAN сетях. Важно, что использование eNB как якорного узла для соединений WLAN, позволяет пользователям использовать LTE сеть для реализации функций управления ресурсами WLAN исключительно для выгрузки трафика. Упрощенная

архитектура для такого подхода показана на рисунке 1.9.В. Функция администрирования радио ресурсов отмечена как RRM модуль, в то время как интерфейс от WiFi к BBU еще не стандартизован, предполагается, что соединение будет проходить через специальный шлюз (так называемый шлюз RAN или R-GW), который выполняет согласование интерфейсов.

В то время как архитектура «якорь-бустер» позволяет координировать использование радио ресурсов в зоне покрытия якорных сот (как правило, макро сот), эффективность системы в целом может быть увеличена, если координация радио ресурсов через якорные соты также будет включена. На практике, различные подходы могут быть использованы для достижения такой координации. В распределенной модели, координация базовой станции eNB с другой базовой станцией eNB может быть осуществлена через интерфейс X2. Как альтернатива, контроллер централизованного распределения радио ресурсов может быть использован для системного администрирования радио ресурсов. 3GPP изучает в настоящее время такую архитектуру для сетей с мульти радио технологиями (в соответствии с документом 3GPP TR 37.870).

Для создания сетей, которые могут использовать высокоскоростные соединения по оптоволокну, архитектура облачной сети доступа Cloud RAN становится осуществимой. При этом линии RRH с простой функциональностью централизуются на уровне пула BBU в облаке. Для таких архитектур, в сети можно легко разместить не-3GPP RRH узлы, что позволяет централизовать координацию для мульти радиосетей.

Концепция H-CRAN введена в [17] и рассмотрена в разделе 1.6, подобные принципы для мульти радио облачной сети доступа обсуждены здесь, но она также позволяет осуществлять централизованную обработку в EPC, когда узлы используют интерфейс S1 для связи с централизованным сервером. Отметим, что централизация сети для координации/кооперации администрирования радио ресурсов позволяет принимать во внимание не только проблемы интерференции, но и изменения нагрузки (например, час наибольшей нагрузки) и мобильность пользовательского оборудования (например, пользовательское оборудование с высокой мобильностью может быть выгружено в макро соту).

В наших дальнейших исследованиях будем предполагать, что H-CRAN позволяет централизовать администрирование радио ресурсов с помощью выделенного объекта, называемого менеджером кооперативных радио ресурсов CRRM (Cooperative Radio Resource Manager), при этом подключение к CRRM серверу осуществляется с помощью такого же интерфейса X2, как и для магистральной сети. Структура и системная модель для такого подхода показаны на рисунке 1.9.C.

1.7 Распределение ресурсов в H-CRAN с использованием менеджера CRRM

В диссертации мы рассматриваем H-CRAN с некоторым числом мульти радио сетей доступа RANs, называемых далее уровнями. Следуя концепции, изложенной в разделе 1.6, все узлы радио доступа на всех уровнях должны иметь возможность взаимодействия с CRRM, который отвечает за централизованное распределение ресурсов радио технологий. Отдельная радио технология, как правило, имеет индивидуальный набор возможностей для адаптации каналов более низкого уровня и механизмы обработки сигналов для управления радио линиями ассоциированных пользователей. Эти важнейшие параметры могли бы быть доступны для CRRM. Вследствие этого, CRRM становится ответственным за общую эффективность оптимизации H-CRAN. Однако разработка приемлемых процедур оптимизации и относительно несложных алгоритмов является очень сложной комплексной исследовательской проблемой. С другой стороны, роль CRRM может быть ограничена исключительно оптимизацией трафика в H-CRAN. В диссертации мы сконцентрируем внимание именно на такой методологии кооперативного администрирования радио ресурсов.

Системное состояние H-CRAN в любой заданный момент времени может быть описано состоянием запросов трафика в отдельных узлах, их возможностями связи с радио сетью доступа и географическим расположением зоны обслуживания. Информация об изменении таких метрик сообща-

ется от пользовательского оборудования или узлов радио доступа в CRRM. В моменты изменений состояния CRRM использует величины этих метрик как исходные для оптимизации соответствующего распределения ресурсов через доступные сети радио доступа. Оптимальное распределение затем доводится до пользовательского оборудования посредством интерфейсов управления и используется до тех пор, пока не наступит следующее изменение состояния. Частота проведения такой оптимизации зависит от динамики исходных параметров.

Далее рассмотрим некоторый момент времени t , когда изменение состояния системы запускает функцию оптимизации ресурсов в CRRM. Соответствующая модель распределения ресурсов включает в себя три составляющие: (i) модель топологии H-CRAN, (ii) спецификацию подходящей модели оптимизации и (iii) эффективный алгоритм решения. Для исходящих линий H-CRAN нас интересует так называемое *разветвленное* распределение ресурсов, позволяющее гибко разделить отдельные запросы трафика между доступными радио интерфейсами. Это соответствует случаю, когда мульти радио пользовательское оборудование может вести передачу через один или более радио интерфейсов одновременно и результаты проявляются в относительной простоте алгоритма распределения ресурсов, также как и при рутинной оптимизации ядра сети, делая их внедрение прагматичным для CRRM.

Традиционный подход к вышеупомянутой проблеме распределения состоит в максимизации минимального отношения сигнал/интерференция плюс шум (SINR) для всех пользователей, подобно методу max-min рассмотренному ранее. Однако это может быть не очень гибким решением вследствие ряда присущих системе ограничений. Первая проблема состоит в том, что потребуется выпуклое программирование со значительно более интенсивными вычислениями, чем модели линейного программирования. Второе ограничение связано с тем, что существующие мульти радио сети доступа могут работать в различных полосах пропускания, и тогда целевая функция примет комплексную форму. С целью уменьшения сложности в диссертации предлагается методология оптимизации, которая базируется на нескольких аппроксимациях точных уравнений интерференции.

Другое важное предположение состоит в том, что все запросы трафика представляются как «жадные» и эластичные. Напомним, что «жадный» трафик занимает все распределенные ресурсы, в то время как эластичность подразумевает адаптацию к реальному доступному распределению ресурсов. С этими двумя предположениями рассматриваемая H-CRAN система полностью характеризуется возможностями пользовательской связности во время t . Как увидим далее, сформулированная модель обеспечивает точные результаты функционирования и в то же время высокую вычислительную эффективность. Как следствие, предложенный подход является вполне подходящим для оптимизации функционирования в реальном времени крупномасштабных и высоко гетерогенных H-CRANs.

Узким местом для H-CRAN являются радио интерфейсы. Тогда здесь, как и ранее, можно абстрагироваться от линий, связывающих узлы сети радиодоступа с пулом BBU's, а результирующий график показан в части C рисунка. Далее, убирая дополнительный набор линий, связывающих пользовательское оборудование с узлами сети радио доступа, приходим к представлению общей топологии с двумя вершинами графа, одна из которых является логическим концентратором, а другая – физическим. Число линий, соединяющих концентраторы равно числу узлов сетей радио доступа на всех уровнях. В такой интерпретации связность пользователей поддерживается набором доступных путей между двумя узлами (часть D рисунка).

Соответственно, пропускная способность и ограничения по запросам представляются следующим образом:

$$\begin{aligned} \sum_{p=1}^{P_d} \alpha_{dp} x_{dp} &= h_d, \\ \sum_{d=1}^M \sum_{p=1}^{P_d} \delta_{edp} x_{dp} &= B_e, \end{aligned} \quad (1.11)$$

где δ_{edp} есть заданная переменная в пределах 0 – 1, при этом линия e используется для обслуживания части запроса d через путь p , B_e есть полоса пропускания радио доступа узла e . Основная задача в случае max-min со-

стоит в лексикографической максимизации вектора распределения полосы пропускания \vec{h}_d , что является задачей линейного программирования.

Естественно, чтобы установить коэффициент α_{dp} в рассмотренных выше спецификациях используется текущая спектральная эффективность мобильного пользователя d беспроводного интерфейса p . Однако, компромисс с выбором этих коэффициентов должен быть в согласии со справедливым распределением ресурсов. Для обеспечения возможности контролируемого баланса между справедливостью распределения ресурсов и производительностью предлагается описанная ранее форма функции спектральной эффективности. Определив $\alpha_{dp} = \beta^{s_{dp}}$, можно использовать параметр β в предложенном выше оптимизационном алгоритме. Принимая $\beta = 1$, приходим к классическому max-min распределению ресурсов в терминах справедливого распределения скоростей передачи данных. При $\beta > 0$, пользователи имеют более высокую спектральную эффективность, что приводит к лучшей производительности системы. При ограничениях $\beta \rightarrow \infty$, занимает вся полоса пропускания и, следовательно, максимизируется производительность системы.

Предложенная структура оптимизации является основой, которая может быть далее расширена путем учета дополнительных метрик. Одно из потенциальных расширений базируется на введении дополнительных коэффициентов (1.11). Например, добавляя коэффициенты γ_d , $d = 1, 2, \dots, M$ к первому уравнению можно учесть приоритетное распределение ресурсов для некоторого заданного числа классов обслуживания. Далее, вводя дополнительные весовые коэффициенты, связанные с потреблением энергии различными радио интерфейсами и/или спектральной эффективностью, можно оптимизировать потребление энергии пользовательским оборудованием при совместной работе с H-CRAN. Кроме того, ограничение запросов трафика может быть введено в модель без каких-либо значимых модификаций посредством специфицирования их в явном виде. Принципиально то, что задача оптимизации не может иметь решения тогда, когда запросы превышают общую пропускную способность системы.

Модель, предложенная в диссертации, может быть модифицирована также для ориентации на различные цели и/или условия внешней сре-

ды. Не разветвленное распределение ресурсов (когда пользовательское оборудование не может делить свой трафик между несколькими радио технологиями) делает задачу оптимизации такой, что требуется использование смешанно-целочисленного программирования (MIP), существенно увеличивая сложность алгоритма решения. Кроме того, эта задача известна как NP-сложная даже в случае, если имеется простой источник и множественные направления приема. Однако в случае H-CRAN, когда все пути передачи данных имеют фиксированную длину равную двум, наличие эффективного алгоритма оптимизации представляется возможным.

Модифицируя предложенную целевую функцию, можно также рассмотреть случаи с различными стоимостями маршрутизации через конкретные сети доступа. Например, более высокие затраты на маршрутизацию через макро сеть LTE могут заставить пользовательское оборудование выгружать больше трафика через WiFi или интерфейсы маленьких сот. Заметим, что эти модификации не оказывают влияния на задачу оптимизации, которая по-прежнему входит в структуру линейного программирования. Если целью оптимизации для H-CRAN являются задержки, то тогда требуется изменить целевую функцию и использовать выпуклое программирование. В заключении отметим, что с приемлемыми модификациями предложенная модель пригодна и для оптимизации нисходящих потоков трафика.

1.8 Численные результаты исследования H-CRAN

Для того, чтобы тщательно оценить предложенные в диссертации оптимизационные решения, проанализируем ряд схем администрирования радио ресурсов в реалистичных условиях внедрения H-CRAN путем построения ряда представительных сценариев создания сети. Для этих сценариев мы используем допустимые аналитические условия, которые оцениваются с помощью системы имитационного моделирования WINTERsim. Обсуждаемая среда комплексно характеризует функционирование ядра сети

при альтернативных стратегиях распределения ресурсов посредством абстрагирования от малоинформативных практических деталей реализации сети с целью уменьшения сложности использования решений.

Поскольку ряд особенностей системного уровня уже подвергся упрощению в самом представляемом инструменте (например, межсотовая интерференция и изменения замираний фиксируются с помощью подходящих предельных значений), используемые в диссертации аналитические условия представляются достаточно точными, чтобы обеспечить первичное понимание обсуждаемых концепций. С упрощением межсотовой и межуровневой интерференции исследования концентрируются на типовых характеристиках сот 3GPP LTE при покрытии макро базовой станцией. Другие инфраструктурные узлы, такие как пико соты и точки доступа WiFi размещаются равномерно в зоне покрытия в соответствии со спецификациями 3GPP.

В дальнейшем, рассмотрим три предлагаемых в диссертации сценария создания H-CRAN, связанные с типовыми стратегиями операторов и относящиеся к соответствующему выбору интеграции LTE/WiFi: (i) оператор создает макро сеть LTE в лицензионном спектре и собственную WiFi сеть в не лицензируемом спектре; (ii) только LTE технология используется оператором с применением как макро сот, так и пико сот на отдельных лицензированных частотах и (iii) все возможные опции доступны оператору, а именно: макро и пико соты LTE и WiFi сети радио доступа. Естественно, что последний сценарий обладает большей потенциальной пропускной способностью и лучшими характеристиками гибкости.

Плотность пико сот и точек доступа WiFi в сценариях создания сети может быть различна, но целевое значение в 4 доступных радио узла на одну макро соту предполагается имитирующим современные городские условия. Общее число единиц пользовательского оборудования зафиксировано на уровне 60 в соответствии с настоящими документами 3GPP и его распределение в зоне предполагается равномерным. Как отмечалось ранее, предполагается, что линии магистральной сети и линии прямого доступа имеют большую пропускную способность, чем рассматриваемые радиолнии, и внимание концентрируется на эффектах, относящихся исключитель-

но к пропускной способности радиоканалов. Предполагается также, что все пользовательское оборудование статично в течение эксперимента, в то время как их географическое положение может изменяться в различных реализациях сценария. Другие важные параметры системы приведены в рассмотренной ранее таблице 2.3. И для точек доступа WiFi, и для пико RRHs LTE используется единое наименование – низко потребляющие узлы LPN.

Далее проведем сравнение эффективности двух путей создания H-CRAN (позволяющих обеспечивать полное управление сетью): централизованного для пользовательского оборудования и интегрированной архитектуры мульти радио с поддержкой сети. Как было проиллюстрировано, в примере централизованного решения для пользовательского оборудования используется простая «жадная» схема (максимального использования), в которой запросы пользовательского оборудования используют все доступные радио ресурсы всех возможных радиосетей доступа для соединения с ними. Такая интуитивная стратегия не накладывает жестких сигнальных требований ни на сеть, ни на пользовательское оборудование, и является достаточно хорошей с точки зрения производительности для конкретного пользователя. Более важно, что такая схема очень проста в реализации.

Другой характерный пример, а именно схема распределения ресурсов с сетевой поддержкой является примером концепции увеличения соты зоны покрытия. Они управляют эффективностью порогов ассоциации для маленьких сот WiFi и LTE пользователей зоны покрытия макро соты. Этот подход позволяет учесть эффект колебаний загрузки конкретных сетей радиодоступа путем увеличения/уменьшения числа пользователей, ассоциированных с маленькими сотами. Естественно, что такой метод требует отдельных решений, внедряемых в eNB вместе с соответствующей сигнализацией для поддержки пользовательского оборудования.

При внедрении решений с сетевой поддержкой рассматриваются не разветвленные опции, которые позволяют пользовательскому оборудованию выбрать предпочтительную сеть радиодоступа в конкретный момент времени. Краткое описание этого алгоритма следующее: если первая попытка со стороны пользовательского оборудования занять ресурсы уровня

WiFi с плохим сигналом неудачна, и вторая попытка для пико соты также безуспешна, то пользовательское оборудование будет обслужено на уровне макро соты. В заключении будет рассмотрен также централизованный механизм, основанный на H-CRAN/CRRM, который централизованно оптимизирует распределение ресурсов между радио технологиями на основе качества радиолиний для пользовательского оборудования.

На рисунке 1.10 приводится детальное сравнение эффективности. Изучается динамическая эффективность, исследуется распределение ресурсов для трех различных сценариев: WiFi и макро сота LTE (левая часть), пико и макро соты LTE (центральная часть) и, наконец, полностью гетерогенный сценарий с интеграцией макро соты LTE, WiFi и пико сот LTE. Различные исходные параметры используются (посредством изменения β коэффициентов, раздел 1.7) при выборе сетевых решений для того, чтобы исследовать соответствующие пределы эффективности.

Естественно, увеличение производительности вызывает деградацию уровней справедливости, которые отмечены индексом Джейна по вертикальной оси. Кроме того, ограничения доступного компромисса возможностей показано на рисунке различными цветами. Например, если попытка запроса для дальнейшего увеличения справедливости встречает ограничения по покрытию, то выделение дополнительных ресурсов для пользовательского оборудования с плохой связностью будет необоснованно, в то время как пользовательское оборудование с хорошей связностью не будет больше иметь дополнительных привилегий в справедливости. С другой стороны, если будут предприняты действия по увеличению производительности сверх установленного предела, то улучшение производительности хорошо связанных пользователей сталкивается с ограничениями по пропускной способности: не будет больше распределено ресурсов вследствие ограничений по полосе пропускания.

Подобный компромисс наблюдается для схемы с сетевой поддержкой, когда варьируется порог ассоциации, основанный на SNR, для малых сот LTE и WiFi. Однако также наблюдается и существенно более широкий диапазон балансировки для централизованной сетевой стратегии, так же, как и высокая результирующая производительность и эффективность

справедливости. Практически это означает, что основанное управление ресурсами на базе H-CRAN показывает большую эффективность с точки зрения справедливости распределения ресурсов и общей производительности сети. Итак, система имеет более высокие значения справедливости и производительности с увеличением доступных сетей RAN, но сочетание уровней «пико плюс макро» более выгодно, чем «WiFi плюс макро», несмотря на то, что использование полосы пропускания в 20 МГц для WiFi должно было бы иметь преимущество по сравнению с 10 МГц пико сотового канала LTE. Это происходит вследствие изначально более лучшего покрытия для уровня пико сетей (за счет более высокой мощности передачи и усиления передатчик / приемник) вместе с большей спектральной эффективностью технологии LTE.

1.9 Практическая реализация концепции в виде прототипа

Для подтверждения практической ценности рассмотренных выше концепций H-CRAN и расширения возможных опций интеграции 5G H-CRAN функциональности в сети мобильных операторов под руководством автора были созданы несколько прототипов для экспериментальной проверки полученных результатов. Это привело к реализации тестовой среды, демонстрирующей полную интеграцию WiFi технологий в существующую сеть 3GPP LTE. В этом разделе рассматриваются условия тестовой среды в настоящее время, основные проблемы, с которыми приходится сталкиваться при создании таких сетей, и предлагаются важные решения по расширению предложенной технологической концепции для операторов при развертывании сетей 5G.

Важно отметить, что большинство существующих терминалов пользователей уже сегодня поддерживают и сотовые интерфейсы, и интерфейсы беспроводных локальных сетей. Однако, борьба за энергетическую эффективность производителей оборудования требует ограничений по использованию пользовательского оборудования в части применения в одно и

то же время только одного радио интерфейса. В настоящее время эта ситуация изменяется вследствие предстоящего широкого внедрения систем 5G, которые будут высоко синергетичной интеграцией различных технологий и решений радиодоступа, а не одной новой технологией. Вследствие этого, современные требования по увеличению беспроводной связности заставляют основных производителей пользовательского оборудования поддерживать разработку перспективных платформ для доступа и эффективного использования всех доступных радио интерфейсов. Как адекватный пример этого положения можно назвать телефоны платформы пользовательского оборудования Jolla, поддерживающие операционную систему Sailfish OS. В экспериментах в созданной тестовой среде тестирование таких устройств доказало большую гибкость и открытую архитектуру, дополненную современным аппаратным обеспечением с возможностями использовать среду разработки, что позволяет модифицировать системный уровень и, таким образом, достигать необходимого уровня связности.

Далее, расширяя видение H-CRAN, можно рассмотреть возможные опции по интеграции сотовых технологий и технологий локальных беспроводных сетей. В идеале, интеграция точек доступа WiFi на стороне сетевого провайдера потребует доступа к открытой сотовой базовой станции, которая тяжело доступна сегодня в исследовательских условиях. Поэтому, необходимая эффективная агрегация и WiFi, и сотовых средств в одной и той же операторской сети может быть выполнена посредством инкапсуляции соответствующих радиолиний в выделенных OpenVPN туннелях. Они будут заканчиваться в виртуальной машине, функционирующей в операторском узле агрегации, что может расширить функциональность пакетного шлюза или выделенного объекта в пуле BBU.

Разработанная тестовая среда H-CRAN использует современную архитектуру программно-конфигурируемой сети SDN (Software-Defined Network) и, в частности, протокол OpenFlow для динамического и эффективного управления связностью пользовательского оборудования [43]. Все радиолинии, доступные отдельному пользовательскому оборудованию, просматриваются в простом плане направлений на телефоне, что эмулируется посредством программного обеспечения Open vSwitch. Реальный вы-

бор направления выполняется программным контроллером реализуются на основе правил OpenFlow, основываясь на активных измерениях условий радиолиний. В тестовой конфигурации контроллер работает на самом пользовательском оборудовании, но предложенное решение является более общим и конфигурация пользовательского оборудования может легко быть передана сети провайдера под управление одного контроллера для всех обслуживаемых пользователей. Однако последний вариант, естественно, накладывает дополнительные требования к доступности канала управления.

Более конкретно, якорные точки для туннелей VPN на операторской стороне функционируют в изоляции внутри соответствующих контейнеров Linux, которые поддерживаются в Docker Engine. Виртуальные магистральные линии от контейнеров подключены Open vSwitch daemon эмулятору операторского плана маршрутизации. Разработка Open vSwitch предполагает, что все линии виртуального интерфейса доступны для обработки кадров Ethernet [43], сотовые линии открываются в системе посредством RmNet или RmNet USB драйвера. Поэтому, добавление радио линий в виртуальный план маршрутизации требует соответствующей модификации vSwitch daemon. Кроме того, для протоколов OpenFlow необходимо рассчитывать желательные заголовки протокола или пропускать заголовки более низкого уровня с предварительно определенной компенсацией. В архитектуре предложенной тестовой среды введен уровень абстракции для линий, определенных посредством GRE туннелей, эмулирующий общий уровень фрагмента сети между пользовательским оборудованием и агрегатором на операторской стороне.

В заключении отметим, что предложенная комбинация механизмов аппаратного обеспечения и виртуальной контейнерно-основанной операционной системы дает возможность быстро и гибко создавать различные сценарии в контексте использования современной экосистемы H-CRAN. Все необходимые компоненты собраны в интегрированную среду, что позволяет легко конструировать необходимые практические структуры. Гибкость предложенной тестовой среды всесторонне поддерживает современное видение операторской сети.

1.10 Выводы

1. Анализ развития новых технологий, концепций развития сетей и систем телекоммуникаций, таких как Интернет Вещей, показывает, что в обозримой перспективе решающую роль в реализации сетей и систем пятого поколения и последующих будут играть гетерогенные структуры, основанные на интеграции множества используемых радио технологий в условиях сверхплотных сетей связи. При этом для создания перспективных сетей и систем связи требуется как разработка новой методологии исследования гетерогенных сетей, так и комплекса взаимоувязанных адекватных новому этапу развития сетей моделей и методов распределения ресурсов.
2. Предложен новый критерий оптимизации распределения ресурсов в гетерогенных сетях связи, отличающийся от известных тем, что справедливое распределение ресурсов для гетерогенных сетей с учетом требуемого баланса между сетевыми возможностями и пропускной способностью достигается за счет использования критерия $\max\text{-min}$, основанного на полосе пропускания. Целевая функция при этом дополняется весовыми коэффициентами на основе мгновенных значений спектральной эффективности.
3. Использование предложенного критерия позволяет сформулировать задачи оптимизации как задачи линейного программирования, предложенные решения хорошо масштабируемы с ростом числа пользователей и инфраструктурных узлов, даже в условиях создания сетей сверх высокой плотности.
4. Разработан метод управления производительностью гетерогенных радиотехнологий доступа H-CRAN, отличающийся от известных тем, что используется кооперативное управление посредством менеджера кооперативных радиоресурсов CRRM в H-CRAN в реальном времени на основе динамического управления ресурсами двух альтернативных метрик: справедливого распределения ресур-

сов для пользователей всех доступных сетей радиодоступа и общей производительности системы.

5. Разработана тестовая среда для H-CRAN, которая использует современную архитектуру программно-конфигурируемой сети SDN и протокол OpenFlow для динамического и эффективного управления связностью пользовательского оборудования. Экспериментальные исследования на тестовой среде подтвердили основные преимущества, которые получает облачная сеть доступа при использовании менеджера кооперативных радио ресурсов.

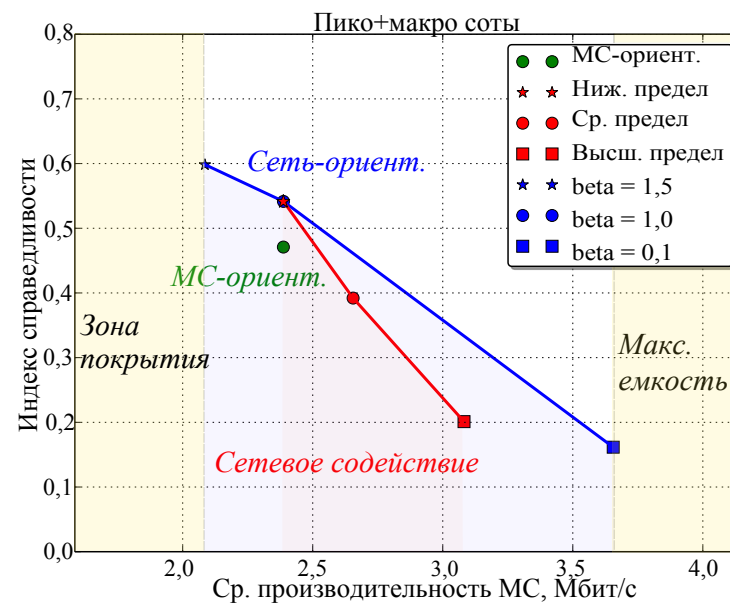
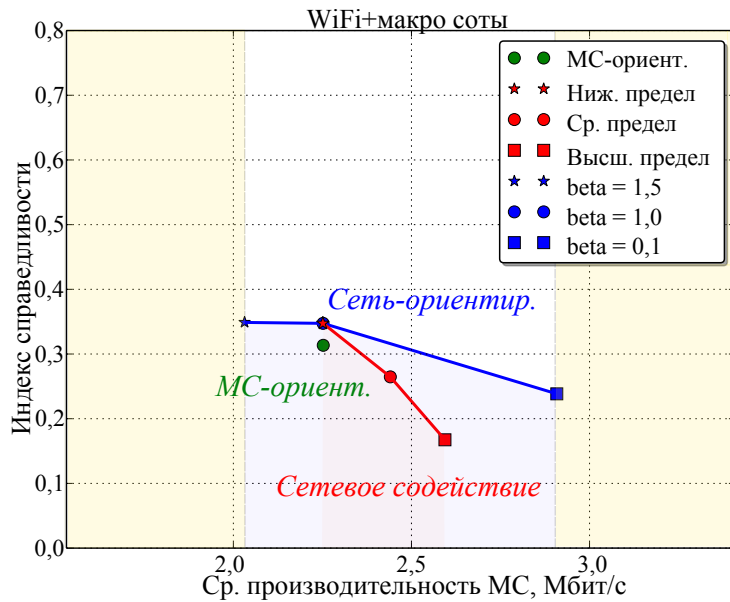
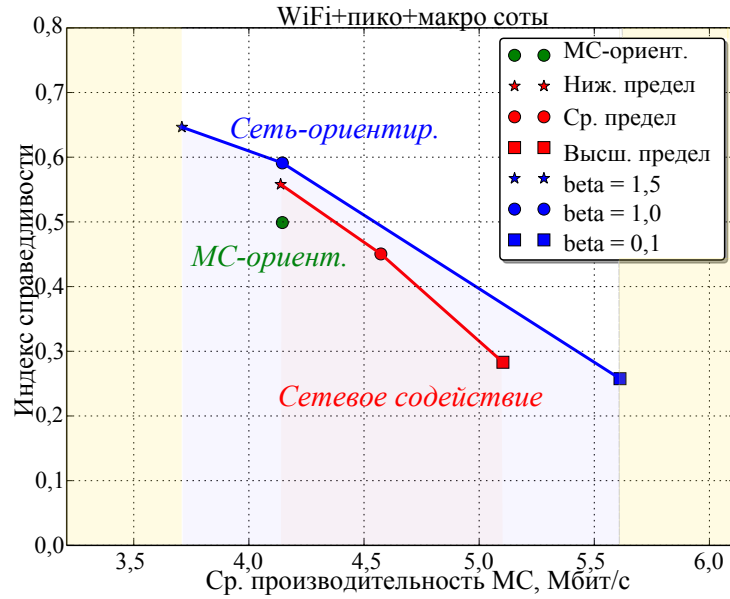


Рисунок 1.10 — Сравнение эффективности для трех различных сценариев

Раздел 2. Классификация гетерогенных сетей HetNet, разработка и исследование моделей и методов повышения их энергоэффективности

2.1 Введение

Рассматривая более чем сорокалетнюю историю мобильной связи (первый мобильный вызов был совершен 3 апреля 1973 года) от аналоговых систем до систем LTE и последующих, можно заметить, что поколения мобильных сетей сменяются примерно каждые двадцать лет. Новая технология достигала своего пика, в среднем, в течение половины времени своего развития, а в оставшийся промежуток проходила путь до последнего пользователя. Началу внедрения каждой из новых технологий предшествовали несколько лет интенсивных исследований и разработок, и, исходя из опыта, можно ожидать начало широкого внедрения беспроводных сетей связи пятого поколения (5G) в 2020 году [44].

Исторически каждое следующее поколение беспроводной связи порождалось возрастающими требованиями пользователей и технологическими недостатками предыдущих версий мобильных систем. Рассмотрим эволюцию мобильных сетей, основные характеристики которой приведены в таблице 2.1. В настоящее время можно предполагать, что критерии сетей 5G сформировались, а основными из них являются следующие [45]

- доступность услуг «в любое время и в любом месте» (англ. anytime, anywhere)
- высокая скорость передачи данных (1-10 Гбит/с)
- экстремально низкая задержка (1мс и ниже)
- высокая доступность и надежность связи для ряда сценариев (99.999%)
- снижение энергопотребления (на 90%)
- массовое увеличение числа соединений (в 10-100 раз)

Таблица 2.1

Эволюция поколений сотовых сетей в терминах услуг и производительности. Источник: GSMA Intelligence

	Базовые услуги	Ключевые особенности	Недостатки
1G	Аналоговые телефонные вызовы	Мобильность	Низкая спектральная эффективность, проблемы информационной безопасности
2G	Цифровые телефонные вызовы и сообщения	Безопасность, массовое использование	Ограниченная скорость передачи данных — сложности в поддержке растущего спроса на Интернет-соединения
3G	Телефонные вызовы, сообщения, передача данных	Улучшенное качество Интернета	Реальная производительность не справляется со спросом, трудности WAP-доступа в Интернет
3.5G	Телефонные вызовы, сообщения и широкополосные данные	Широкополосный Интернет, приложения	Привязка к архитектуре и протоколам предыдущего поколения
4G	Все услуги на базе IP (включая передачу голоса и сообщений)	Высокоскоростной Интернет, более низкая задержка	...не до конца определены

Методы обеспечения некоторых из этих технических требований (например, такие как высокая скорость и небольшая задержка), уже активно разрабатываются операторами связи. Остальные требования слишком разнообразны, и одной технологии радиодоступа (Radio Access Technology, RAT) в данном случае недостаточно. В связи с этим, ожидается, что технологии пятого поколения будут существенно отличаться от всех предшествующих (см. [13;46], и [16]). Сегодня многие исследователи пришли к согласию [4;47], что перспективные сети 5G будут основаны на использовании сочетания различных технологий радиодоступа (таких как 3G, 4G, WiFi, протоколы, поддерживающие передачу в миллиметровом спектре, и другие) и предоставят пользователям более высокие скорости, лучшую связность и покрытие. Благодаря подобной интеграции различных технологий в пользовательских устройствах, 5G открывает возможности для построения так называемого «гиперсвязанного общества» с высоким социальным

и экономическим уровнем и жизненно важной ролью беспроводной связи в развитии человека.

Таким образом, естественным ключевым вопросом и фундаментальной задачей для создания сетей 5G является выбор подходящих стратегий управления множеством радиотехнологий [48], что поможет обеспечить наилучшее функционирование сетей связи. Это означает, что сети необходимо проектировать таким образом, чтобы создать определенную гибкость или *интеллектуальность* системы в целом, с возможностью адаптации к различным требованиям, условиям окружающей среды и вариантам развертывания сети.

При проведении научно-исследовательских работ необходима эффективная методология, которая позволит сфокусироваться на наиболее важных для 5G целях, таких как, например, ультранизкая задержка и высокая скорость передачи данных, а также предоставит возможность достаточно быстро предсказывать производительность конкретной 5G системы в зависимости от изменения условий функционирования. Хотя инструменты имитационного моделирования могут значительно помочь и в этом случае, прогнозируемое уплотнение сети и системные требования диктуют необходимость создания простого метода *первичной оценки производительности*, способного надежно предсказывать поведение системы.

2.2 Кооперативные сети

2.2.1 Основные предпосылки

Несмотря на достаточно большую вариативность с точки зрения скорости передачи данных, существующие радиотехнологии все еще подвержены большим задержкам, а иногда даже и отказам в обслуживании в силу недостаточного покрытия или интерференции. Сотовое покрытие в помещениях также зачастую неидеально, несмотря на переиспользование спек-

тра (spectrum reuse) и сложные методы погашения интерференции [49; 50]. Ситуацию усугубляет и то, что в сети появляется большое число машин (machine-type devices) [51], формируя так называемый Интернет Вещей (Internet of Things, IoT) и фундаментально изменяя глобальную сеть. Подобные изменения выдвигают новые требования к сетям и системам 5G и формируют спрос на новые технологические решения [1].

Одним из таких решений, направленных на уменьшение диспропорции между ожидаемым качеством обслуживания и доступными беспроводными ресурсами, является поддержка высокой плотности устройств *малых сот* в существующей архитектуре сотовых сетей ([52], [53], [2] и [6]). Это повышает пропускную способность сети за счет большего переиспользования частот (frequency reuse) и увеличения средней скорости передачи данных (т.к. для сот меньшего размера характерны более короткие расстояния). Кроме того, поскольку размер соты уменьшается, увеличивается перекрытие сотовых, локальных и персональных сетей, что создает предпосылки для одновременного использования нескольких радиотехнологий с целью увеличения емкости и связности сети [54], [55].

Далее, в соответствии с тенденциями роста мобильного трафика [56], еще одно решение может быть найдено в выгрузке трафика пользовательских сессий в соединение непосредственно между устройствами (device-to-device, D2D). В рамках концепции D2D близлежащие беспроводные устройства могут взаимодействовать между собой без использования инфраструктуры сети, увеличивая переиспользование спектра [57]. Кроме того, близость пользовательских устройств обеспечивает более высокую скорость передачи, низкие задержки и уменьшение энергопотребления [58; 59]. Потенциальные приложения D2D в сотовых сетях весьма разнообразны [60] и включают как локальное голосовое обслуживание (т.е. выгрузку голосовых вызовов), так и обмен мультимедиа контентом, игры, мультикастинг, контекстно-зависимые приложения и приложения, связанные с общественной безопасностью.

Таким образом, растут потребности в эффективной координации между альтернативными радиотехнологиями [61]. В рамках этого, распределенные сети с нелицензируемым спектром (например, WLAN) мо-

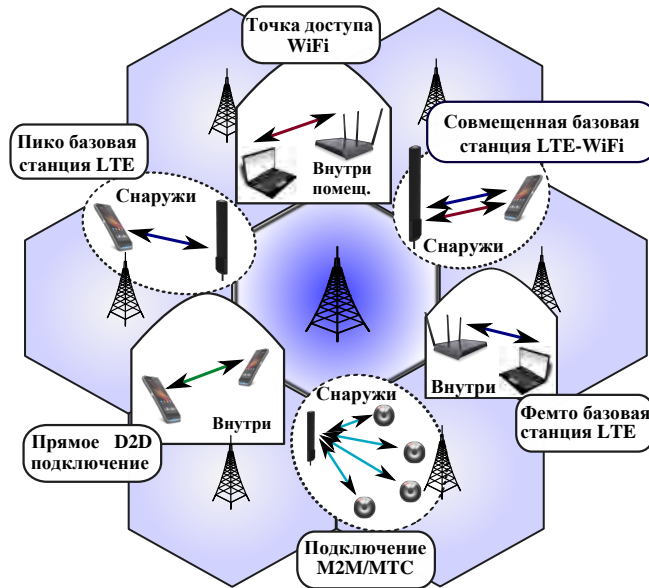


Рисунок 2.1 — Архитектура гетерогенной сети 5G

гут, например, использовать для эффективной динамической ассоциации *централизованную* функцию управления, предоставленную сотовой сетью. Однако в то время как основные усилия были направлены на оптимизацию производительности индивидуальных радиотехнологий, исследованиям механизмов *поддержки со стороны сети* при совместном использовании нескольких радиотехнологий, было уделено ограниченное внимание.

В заключении отметим, что в настоящее время ожидается, что емкость сетей, в основном, будет увеличиваться за счет усовершенствованных архитектур и протоколов, использующих нелицензионный спектр и характер взаимодействия между устройствами и сетью (а также непосредственно самих устройств между собой) посредством *гетерогенных* (heterogeneous) сетей. Подводя итог, можно сказать, что эффективное управление набором различных радиотехнологий позволяет поддерживать многообразие требований, выдвигаемых новым поколением связи 5G; в частности, технологии как большего, так и меньшего радиуса действия, могут функционировать в тесной взаимосвязи с целью увеличения пропускной способности и качества обслуживания сети.

2.2.2 Введение

2.2.3 Основы предложенной методологии

Основная цель этого раздела состоит в разработке единой методологии классификации и исследования моделей различных гетерогенных сетей для оценки их производительности в будущих сетях 5G. Представим предложенную в диссертации методологию в форме простого *набора правил*, т.е. в виде элементарных блоков, которые могут быть скомбинированы для построения требуемой системной модели и расчета необходимых характеристик производительности сети. Ниже приведем краткое описание того, как работает предложенная конструкция.

Основа структуры предложенной в диссертации методологии, а также промежуточные и связанные с ними результаты, представлены на рисунке 2.2. Как видим, существует дифференциация между *пространственной компонентой*, которая взаимосвязана с пользователем и инфраструктурой (на рисунке слева), и *временной компонентой*, относящейся к появлению пользователей, обслуживанию и их уходу из системы (на рисунке справа). Обе единицы содержат несколько последовательных блоков (например, блок, описывающий распределенную инфраструктуру), для каждого из которых имеется несколько альтернативных вариантов (например, пуассоновский точечный поток, Poisson Point Process, PPP). После выбора опции для каждого блока в соответствии с принятой классификацией результаты для пространственной компоненты (такие как распределение расстояний, SINR (Signal Interference plus Noise Ratio) и требуемые ресурсы для обслуживания) могут быть переданы во временную компоненту в форме переходных вероятностей для основного марковского процесса, описывающего динамику системы и трафика.

Структура основного марковского процесса базируется на соответствующем выборе во временной компоненте. Комбинируя выходные данные от пространственной компоненты со структурой процессов во времен-

ной компоненте, можно получить формулировку задачи теории массового обслуживания с известными вероятностями перехода. В зависимости от сложности обеих компонент: (i) выражений для переходных вероятностей и (ii) структуры марковского процесса, можно получить результаты в численной или в замкнутой форме. Подчеркнем, что в зависимости от выбора блоков обе компоненты могут быть главными для нетривиальных формулировок.

Характеризация пространственных компонентов

Пространственное распределение инфраструктуры
 - Точечный пуассоновский процесс (PPP)
 - Кластерный пуассоновский процесс
 - Жесткий точечный процесс
 - Детерминантный процесс (на пр. по Гинибре)
 - ...

Пространственное распределение пользователей
 - Точечный пуассоновский процесс (PPP)
 - Кластерный пуассоновский процесс
 - Жесткий точечный процесс
 - Детерминантный процесс (на пр. по Гинибре)
 - ...

Рассматриваемая область
 - Круг
 - Ячейка Вороного
 - Шестиугольник
 - Треугольник
 - ...

Характеризация временных компонентов

Процесс появления пользователей
 - Пуассоновский процесс
 - ПП модулированный по Маркову
 - D-BMAP процесс
 - ...

Тип (QoS) класса

Трафик на сессию (время)
 - Экспоненциальный
 - По Эрлангу (Erlang-k)
 - Гиперэкспоненциальный
 - ...

Трафик на файл (размер)
 - Экспоненциальный
 - По Эрлангу (Erlang-k)
 - Распределение с медленно убывающим "хвостом"
 - ...

Распределение ресурсов/контроль мощности/доступ
 - Round robin (RR) - Контроль ресурсов - Фиксированная мощность
 - Взвешенный RR - Контроль скорости - Контроль мощности
 - ... - ... - ...

Модель интерференции

Модель канала

Вероятности перехода для соответствующей цепи Маркова

Выходные пространственные характеристики (распределения)

- Расстояния между приемником и передатчиком
- Мгновенных пропускных способностей
- Средних пропускных способностей
- Занимаемых частей ресурсов
- Всех занимаемых ресурсов

Выходные временные характеристики

- Система очередей A/S/c/K/N/D
- Стационарное распределение для текущих пользователей
- Характеристики системы, усредненные по пространству и времени

Характеристики усредненные по пространству и времени

- Среднее количество пользователей
- Средняя задержка
- Средний расход энергии
- Вероятности успеха доступа
 - ...

Рисунок 2.2 — Структура набора данных и результатов в предложенной методологии

2.2.4 Предлагаемая классификация гетерогенных сетей

В данном разделе рассмотрим характеристики гетерогенной сети как пример демонстрации предложенной выше методологии в приложении к оценке производительности сетей связи. Начнем рассмотрение с описания предложенной в диссертации *классификации* для различных уровней гетерогенной сети HetNet, охватывающих все ее выдающиеся особенности с точки зрения аналитических решений.

Продолжая в этом же ключе, рассмотрим одну меченую макросоту с базовой станцией (base station, BS), расположенной в центре соты и собирающей всю необходимую для управления информацию для обеспечения *сетевой поддержки* посредством механизмов *контроля доступа в сеть*. Для каждого уровня в такой гетерогенной сети HetNet, будем проводить дифференциацию между *тремя* основными компонентами (оси на рисунке 2.3), которые изначально определяются соответствующими математическими конструкциями, а именно:

- интерференция (незначительная, поэтому можно ей пренебречь, или существенная, которая должна учитываться в явном виде),
- управление мощностью/распределение ресурсов (варьируется от фиксированного распределения мощности, например, при отсутствии контроля мощности, до алгоритма распределения ресурсов с помощью циклического опроса (round-robin), в то же время оптимальное распределение мощности может быть достигнуто между этими двумя граничными случаями [62]),
- использование ресурсов (выделенные ресурсы для линии связи или общий канал доступа для нескольких линий).

На рисунке 2.3 проиллюстрировано трехмерное пространство, сформированное на основе представленных выше трех критериев и демонстрирующее различные виды гетерогенных сетей HetNet в качестве примеров такого пространства. Разделим предложенные примеры на три группы, в дальнейшем будем рассматривать эти группы независимо как индивидуальные модели:

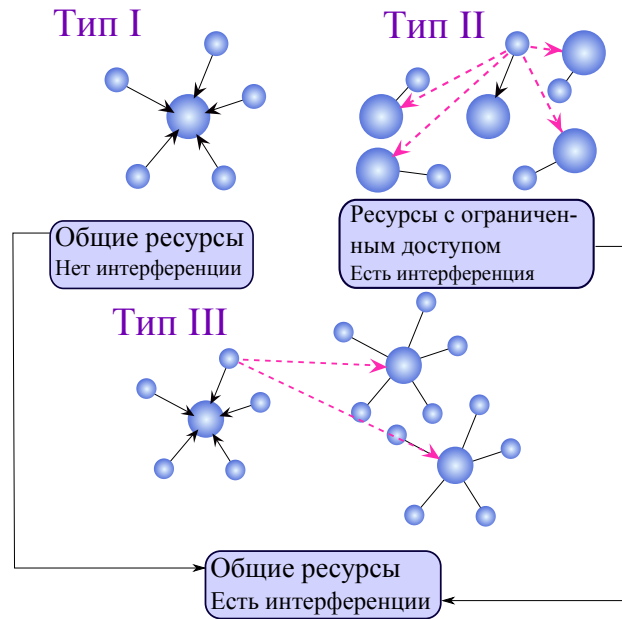


Рисунок 2.3 — Иллюстрация предложенной классификации для различных видов гетерогенных сетей HetNet

- Тип I (условно назовем «макро»): ресурсы распределяются между несколькими линиями, влияние интерференции может быть нивелировано благодаря технологическим особенностям (таким как координация, частотное планирование, направленное излучение, а также другие современные и перспективные технические решения).
- Тип II (условно назовем «D2D»): ресурсы эксклюзивно принадлежат одной линии, но интерференцию следует принимать во внимание.
- Тип III (условно назовем «малая сота»): ресурсы распределяются между несколькими линиями, и интерференция должна приниматься во внимание для близлежащих сот.

Итак, типы гетерогенных сетей, предложенные выше, пронумерованы в соответствии с их возрастающей сложностью (рисунок 2.4). Простейший Тип I («макро») эквивалентен единственной соте в зоне покрытия макро базовой станции (один объект), когда интерференция от других объектов сети может рассматриваться как фоновый шум вследствие сложных процедур управления интерференцией, распределяющих ресурсы базовых станций.

Более сложный сценарий рассматривается для Типа II («D2D»), когда ресурсы эксклюзивны для приемопередающей пары и, таким образом, не могут быть использованы для других пользователей, а интерференция от других соединений (объектов) существует и должна приниматься во внимание. И, наконец, наиболее сложная опция — Тип III («малая сота»), подразумевает значимую интерференцию между объектами сети, а также использование общих ресурсов несколькими передатчиками.

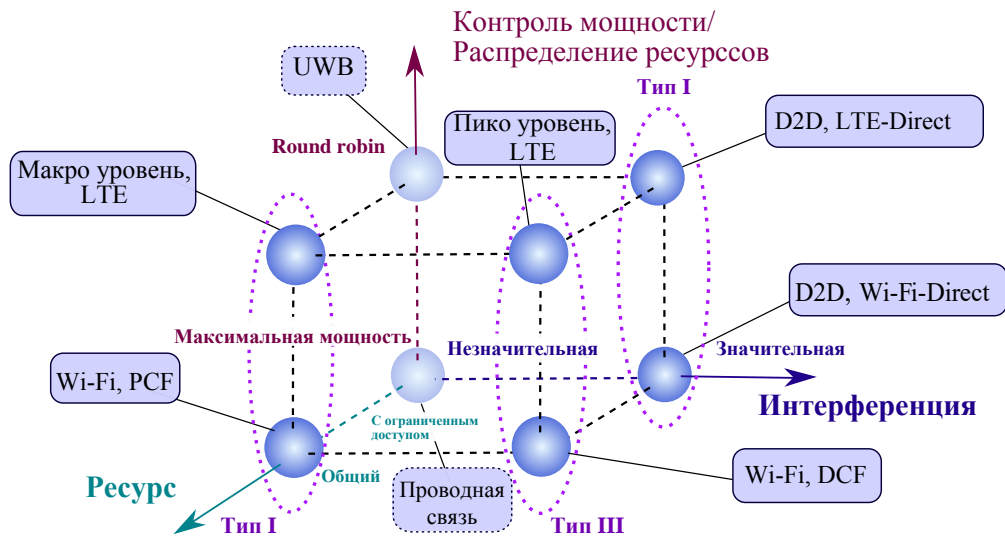


Рисунок 2.4 — Предложенная классификация для различных типов гетерогенных сетей HetNet

Отдавая дань широкому разнообразию механизмов управления мощностью, скомбинируем все возможные схемы распределения ресурсов и управления мощностью в одну вертикальную группу (рисунок 2.4) и в дальнейшем соотнесем их с *типами системы*, как указано выше. В качестве примера рассмотрим точку на рисунке, обозначенную как «WiFi, DCF», использующую механизмы на основе стандарта IEEE 802.11; эта система относится к случаю значимой интерференции между «сота́ми» WiFi, в то время как ресурсы одной точки доступа (access point, AP) распределяются между несколькими пользователями с фиксированной мощностью передачи. Другим примером является точка, обозначенная как «WiFi, LTE-Direct», где интерференция между ближайшими линиями может быть достаточно высокой, но ресурсы исходящей связи отдаются эксклюзивно одному соединению.

2.3 Архитектура 3GPP и предлагаемые модификации

Технической причиной важности интеграции является необходимость наличия жизнеспособной сетевой архитектуры. В этом разделе, прежде чем предложить некоторые улучшения в архитектуре, а также несколько новых дополнительных модулей в ней, кратко наметим задачи в соответствии с изложенными подходами к интеграции.

Текущее видение операторов и задачи

На сегодняшний день типичный портфель современного сетевого оператора включает (i) многоуровневую сотовую сеть, имеющую узлы как традиционной, так и LTE инфраструктуры; (ii) сеть доступа к IP-сетям, включая масштабно развернутые широкополосные беспроводные сети; (iii) существенный набор услуг IMS (например, мобильного ТВ и радио), облегчающий доставку объемного контента пользователям с достаточной пропускной способностью; и, наконец, (iv) множество пользователей с безлимитными объемами данных, что продиктовано популярностью смартфонов и планшетов.

Резкое увеличение числа пользователей с безлимитным тарифом представляет собой серьезную проблему для возможностей беспроводных сетей доступа RAN, при этом всевозможные формы выгрузки мобильного трафика становятся ключевыми для операторов мобильной сети [63]. Таким образом, интегральная гетерогенная сеть с несколькими типами «сторонних» радиотехнологий RAT (таких как, например, WiFi, которые могут бесшовно взаимодействовать с мобильной архитектурой 3GPP) является важным инструментом оптимизации затрат оператора. Однако, стоит заметить, что подобная интеграция нескольких радиотехнологий RAT в единую систему требует особого подхода. В частности, пользовательское

оборудование UE должно работать эффективно без нарушения требований по качеству обслуживания.

Тем не менее, текущие варианты интеграции, предусмотренные 3GPP, далеко не настолько гибки, чтобы обеспечить эффективную связь для систем с несколькими радиотехнологиями [64] на желаемом уровне. Например, любая технология доступа, отличная от указанных в 3GPP, должна быть преобразована в шлюзе пакетной передачи данных сети (Packet Data Network Gateway, PDN GW). Таким образом, любое пользовательское оборудование UE, которое планирует использовать не-3GPP (а какую-то иную) радиотехнологию RAT вместе с сотовым доступом 3GPP, должно направлять свой трафик на шлюз PDN оператора, откуда пакеты будут перенаправлены адресату, следуя логике мобильной IP-адресации (Internet Protocol). Поскольку мобильный IP-трафик блокируется большинством устройств преобразования сетевых адресов (Network Address Translation, NAT), которыми изобилует инфраструктура IPv4, сеанс мобильного IP-трафика необходимо туннелировать в шлюзе PDN через VPN. Такое решение чрезвычайно хрупкое и громоздкое, но в настоящее время *только* оно является доступной опцией. В случае если IPv6 доступен, туннель VPN больше не нужен, что делает архитектуру существенно более эффективной. Тем не менее, это не разрешает *ключевую проблему* — весь трафик должен доходить до шлюза PDN, прежде чем он будет маршрутизирован далее.

Это, к сожалению, может привести к ненужному перенаправлению трафика, увеличению задержек и возникновению перегрузок. Важно отметить, что в то время как использование мобильного IP делает мобильный WiFi несколько похожим на сотовую мобильность, между этими механизмами существует значительная разница: не-3GPP доступ не может использоваться без перерывов в связи вследствие того, что есть хэндовер, а также задержки в переключении и соответствующая реакция со стороны мобильного IP. В результате существующая архитектура 3GPP для доступа не-3GPP не решает проблему построения эффективной радиосети на основе нескольких технологий, поскольку каждый раз, когда передача данных переключается с LTE на WiFi и обратно, ряд пакетов может быть потерян, и

одновременной использование технологий становится невозможным. Ниже детально проанализируем существующую архитектуру 3GPP и предложим решение для улучшения данной ситуации.

Существующий выбор интеграции LTE/WiFi

Как известно, все пользовательские данные в LTE представлены в виде IP-пакетов, и все IP-пакеты перенаправляются с фиксированным уровнем качества обслуживания QoS через соответствующие системы пакетной передачи (EPS) (которые являются своего рода VPN-туннелями, использующими протокол GTP-U для инкапсуляции данных и протокол GTP-C для установления соединения). Как видим, сеть LTE внутренне работает как система с коммутацией каналов, а для внешнего мира является системой с коммутацией пакетов. Это позволяет потокам данных маршрутизироваться и распределяться по приоритетам в сети оператора любым желаемым способом, так что QoS может поддерживаться и при перемещении пользователей. В тоже время несмотря на то, что действительно обеспечивается гибкость, которую требуют сотовые системы, функционирование системы длительной эволюции существенно отличается от того, как работает IP. В результате, не прямой IP-трафик от пользователей фактически разрешен внутри сети LTE.

Вышеизложенное имеет важное значение для эффективной интеграции LTE/WiFi, особенно в случае мобильности. Если мобильный пользователь занимает ресурсы внутри сети LTE, хэндовер может быть надежно скрыт от восприятия пользователем с помощью модуля управления мобильностью (MME), который обеспечивает переключение туннеля. Если, однако, хэндовер должен осуществляться в не-3GPP радиотехнологии RAT, например WiFi или WiGig, то даже если точка доступа AP совмещена с одной из базовых станций LTE eNodeB, мобильный IP нуждается в переключении потоков, что дает потери пакетов. Кроме того, шлюз PDN должен использоваться как домашний агент. В конечном итоге, если

IPv6 не поддерживается в какой-либо точке маршрута между WiFi точкой доступа AP и шлюзом PDN, туннель VPN должен использоваться для подключения к шлюзу PDN до взаимодействия с мобильным IP.

Подобные технические трудности возникают при разработке решений для совмещенных устройств LTE/WiFi в гетерогенных сетях HetNet. При этом, по крайней мере, требуется, чтобы LTE или WiFi были доступны для доступа в Интернет в любое время; предпочтительно также, чтобы переключение между двумя радиотехнологиями RAT было прозрачным. Чтобы достичь этого, нужно было бы подключить WiFi точку доступа AP к Интернету, так чтобы пользовательское оборудование UE могло иметь доступ к шлюзу PDN посредством своего туннеля, а затем полагаться на то, что мобильный IP работает достаточно быстро при хэндовере. Таким образом, каждая совмещенная малая сота LTE/WiFi должна быть подключена не только к внутренней интернет сети оператора (в которой работают интерфейсы S1, S3, S4 и X2), но также и к Интернету, усложняя, таким образом, сеть. Единственная причина, по которой последнее необходимо, заключается в том, чтобы позволить пользовательскому оборудованию UE иметь доступ к шлюзу PDN с внешней стороны по отношению к пакетному ядру сети (EPC).

Однако было бы проще и эффективнее симулировать точку доступа AP для WiFi в ядре EPC в виде eNodeB, при этом ее технология на физическом уровне фактически совпадает с LTE. Это позволит использовать все протоколы, адаптируемые к IP, поверх протокола доступа к среде MAC для WiFi. Основываясь на этом понимании, предложим альтернативный подход к тому, как технологии, такие как WiFi, должны быть интегрированы в жизнеспособную архитектуру 3GPP. При этом оставим вне рассмотрения потенциальные проблемы совместимости радиотехнологий, которые могут иметь место в совмещенных технических средствах, поскольку они обычно разрешаются с минимальными затратами.

Предлагаемые изменения в архитектуру 3GPP

Целью предлагаемых усовершенствований системы для гетерогенных сетей HetNet является улучшение управления не-3GPP потоками данных в рамках пакетного ядра сети EPS. Поскольку это выходит за рамки домашнего eNodeB, называемого HeNB (Home eNodeB), местного доступа к IP (LIPA), а также и за рамки функции доверенного WiFi, для интерфейсов WiFi и EPS потребуются новый модуль, который маскирует различия между WiFi точкой доступа AP и фактическим eNodeB в направлении пользовательского оборудования UE, а также ядра сети, и который обеспечивает всю необходимую сигнализацию на всех соответствующих интерфейсах. Значительно расширяя существующие функциональные возможности HeNB-GW, этот модуль, назовем его *шлюз входа в сеть доступа* (Alien Access Gateway, AAGW), может рассматриваться как часть малых сот или сегментов WiFi на пути к действительно интегрированным системам LTE/WiFi.

Модуль AAGW действует как прозрачный прокси-сервер на канальном уровне, где сотовая сторона выполняет все характерные протоколы, такие как SCTP, для связи с другими базовыми станциями и MME, в то время как не-3GPP сторона доступа использует WiFi непосредственно для передачи всех пакетов, которые обычно передаются через U-интерфейс между eNodeB и UE (наиболее важно, что RRC-пакеты будут необходимы для передачи). На стороне пользовательского оборудования UE оба интерфейса WiFi и туннельный связаны с верхними уровнями стека протоколов сотовой сети, который теперь может открыть полностью функциональность ядра сети EPS для шлюза PDN через физический PHY и MAC уровни WiFi. Как это ни странно, помимо модуля AAGW, никаких других изменений в архитектуре 3GPP не требуется, так как остальная часть сети может с уверенностью предполагать, что новый объект WiFi является еще одним eNodeB. Здесь также важно отметить, что AAGW не является базовым модулем в любом из предполагаемых развертываний сети. В качестве чисто программного решения он обладает максимальной гибкостью разверты-

вания и минимальными капитальными затратами. Поскольку единственной реальной альтернативой предложенному решению является отсутствие прозрачного переключения потока, нет необходимости в дальнейшей аргументации необходимости такого модуля.

В результате создания AAGW в малых сотах и на точках доступа WiFi AP, весь не-3GPP трафик поступает до обслуживающего шлюза, маскируя точку входа как другой вид eNodeB. Фактически, совместно расположенная базовая станция LTE-WiFi теперь будет функционировать так, как если бы это были две отдельные соты; все, что ей необходимо для функционирования — внутренняя сеть оператора (такая, что интерфейсы S1 и X2 с обслуживающими шлюзами и другими eNodeB могут быть активированы). Таким образом, развертывание совмещенных базовых станций будет существенно упрощено, поскольку эти базовые станции больше не нуждаются в отдельном доступе к сети Интернет, как показано на рисунке 2.5. Наконец, с модулем AAGW возможны и все возможные обычные сценарии сотовых сетей между LTE и WiFi так же естественно, как если бы они были той же самой технологией.

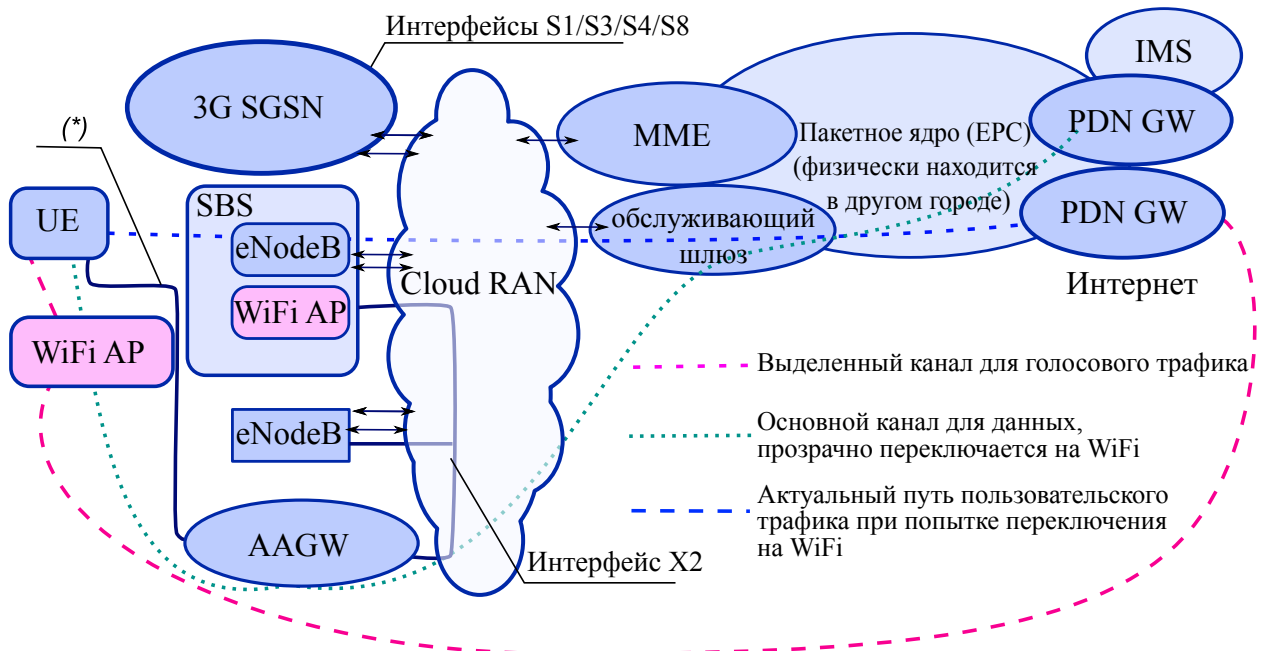


Рисунок 2.5 — Предложенная высокоуровневая архитектура и результирующие потоки данных для сетей пятого поколения 5G с несколькими радиотехнологиями RAT для LTE/WiFi сетей

Продолжим анализ случая, когда точка доступа WiFi не может напрямую подключаться к интерфейсам сети S1/X2. Оператор может избежать такого подключения по соображениям безопасности или точка доступа AP может быть развернута третьей стороной (торговый центр, владелец здания, железнодорожная станция и т.д.) [52]. Эта ситуация, вероятно, была основной причиной, побудившей 3GPP использовать доступ через виртуальную частную сеть VPN и шлюз PDN в существующей в стандартах архитектуре. Однако, если оператор не устанавливает шлюзы PDN в каждом районе, где существуют точки доступа WiFi, направление, использующее такую топологию, всегда останется субоптимальным. Напротив, чтобы обеспечить экономичное функционирование без расширения присутствия ядра сети EPS по всей зоне покрытия, можно было бы еще больше расширить возможности предложенного в диссертации модуля AAGW, что будет рассмотрено ниже.

Используя подходящий протокол туннелирования (например, такие как L2TP или IPSec), можно безопасно и эффективно туннелировать сообщения L2 от третьей стороны с помощью модуля AAGW до пользовательского оборудования UE. Все, что требуется, это заменить интеграцию WiFi с VPN-сервером — и пакеты из пользовательского оборудования UE могут быть переданы в AAGW с любой точки доступа. Чтобы идентифицировать пользователей, можно использовать сервер аутентификации на основе SIM-карты или логина и пароля. Как только пакеты UE поступят на модуль AAGW, он может использовать все инструменты управления потоками, предоставляемые сетью LTE, включая прозрачный хэндовер от одной точки доступа к другой. Очевидно, чтобы использовать предлагаемый модуль AAGW (расположенный в совмещенной базовой станции или через VPN), для пользовательского оборудования UE, возможно, потребуются установить необходимые драйверы для взаимодействия с объектом, который ведет себя как базовая станция LTE, но доступен по совершенно другой технологии L2. Это является основным препятствием в процессе развертывания, однако подобные проблемы возникают и при использовании существующей архитектуры 3GPP (поскольку мобильный IP должен быть установлен на пользовательском оборудовании UE). Наконец, поль-

зовательскому оборудованию UE необходимо будет узнать IP-адрес своего ближайшего AAGW с помощью соответствующего механизма, такого как, например, ANDSF. Когда все вышеперечисленное комплексно реализовано, использование WiFi оператором будет неразличимо для верхних уровней от использования соты LTE, что автоматически обеспечивает все преимущества сотовой архитектуры, такие как прозрачная мобильность.

В конечном счете, если нет необходимости в технологиях доступа для обеспечения того же уровня обслуживания, что и на доступе в сотовой сети, то нет смысла даже смешивать их вместе: пользовательское оборудование UE может быть идеально совместимо с несколькими возможными маршрутами передачи данных одновременно с верхним уровнем подобным многомаршрутному TCP [65]. Наконец, поскольку все точки доступа WiFi теперь, по существу, имитируются как eNodeB в предлагаемой архитектуре гетерогенной сети HetNet, возможно динамически разделять потоки пакетов между несколькими сотами LTE и точками доступа AP WiFi, используя существующие процедуры сигнализации протокола CoMP. Этот протокол позволяет отправлять произвольные части одного и того же процесса переноса информации на разные соты, таким образом, разрешая настоящее многомаршрутное функционирование без необходимости дополнительных решений на транспортном уровне, например, многомаршрутного TCP. В дальнейшем эта новая возможность гибкого разделения потока данных станет основой математического анализа для сверхплотного развертывания малых сот с совмещенными интерфейсами LTE/WiFi.

Теперь предположим, что есть сеть оператора, на которой установлены модули AAGW в каждом городе внутри своей зоны покрытия, и все связанные с этой сетью сети WiFi в городе позволяют мобильному пользователю UE получить к этим модулям доступ через процедуры ANDSF, и в то же время все базовые станции этого оператора eNB оснащены WiFi интерфейсами, которые настроены для функционирования в качестве совмещенных точек доступа AP. В разделе 4 данной диссертационной работы ставятся и решаются задачи, связанные с использованием технологии D2D для выгрузки трафика.

2.4 Повышение энергоэффективности устройств в гетерогенных сетях связи

2.4.1 Введение

В этом разделе исследуется работа мобильного устройства с несколькими радиоинтерфейсами в гетерогенных беспроводных сетях. Предполагается, что такое устройство может гибко управлять своими радиоинтерфейсами при использовании имеющихся технологий радиодоступа. В частности, исследуется потенциал эффективного распределения мощности передачи и разрабатывается оптимальная схема управления мощностью, которая строго максимизирует энергоэффективность мобильного устройства и в то же время удовлетворяет минимальному требуемому уровню скорости передачи данных. При сравнении с более простыми (эвристическими) стратегиями управления мощностью предлагаемое решение на основе использования нескольких радиотехнологий всегда демонстрирует наибольшую энергоэффективность, что делает представленное решение полезным ориентиром для беспроводных сетей 5G.

Гетерогенные сети HetNet как составляющая беспроводной связи нового поколения приобретают все большую популярность, и, как ожидается, позволят улучшить доступ абонентов к службам и приложениям, которые ранее поддерживались только через фиксированные широкополосные системы. Существующие беспроводные сети не всегда способны обеспечивать своим пользователям желаемое качество связи в силу недостатка емкости и отсутствия равномерности обслуживания [66].

На протяжении многих лет спектр беспроводной связи оставался одним из самых ценных ресурсов, что подчеркивает важность его эффективного использования и необходимости повышения спектральной эффективности. Тем не менее, энергоэффективность также становится все более важным фактором, в первую очередь, для небольших мобильных устройств, где потребление энергии за счет беспроводной связи доминирует в общем

бюджете мощности. Это связано с увеличением диспропорции между доступной и требуемой емкостью аккумулятора, продиктованной мультимедийными приложениями [67]. Чтобы компенсировать этот растущий разрыв, необходимы глобальные изменения во всех аспектах проектирования беспроводной системы [68].

В настоящее время беспроводные сотовые, локальные и персональные сетевые технологии, а также поддерживающие их сетевые архитектуры развиваются в сторону более совершенных и сложных гетерогенных сетей. С другой стороны, потребительская электроника порождает огромный всплеск в количестве и разнообразии радиоустройств [69], обусловленных потребностью пользователя в беспроводной связи «в любое время и в любом месте».

Ниже рассматривается проблема энергоэффективного управления доступными беспроводными радиотехнологиями в пользовательском устройстве. Целью ставится разработка доказуемо эффективных методов, позволяющих значительно повысить производительность устройства с точки зрения результирующей энергоэффективности.

Стандартные беспроводные устройства передают свои данные, выбирая один из фиксированных наборов схем модуляции и кодирования, жертвуя гибкой адаптацией мощности ради упрощения [70]. Это часто приводит к чрезмерному потреблению энергии или заниженным скоростям передачи данных, выбранных для текущего качества канала [71]. Следовательно, параметры физического уровня должны гибко настраиваться для того, чтобы непосредственно учитывать требования качества обслуживания (quality of service, QoS) клиента, а также состояние беспроводного канала в целях достижения баланса между затраченной энергией и спектральной эффективностью [72]. В связи с этим, оптимизация пропускной способности уже давно является привлекательным направлением исследований [73], [74]. Однако, поскольку беспроводные клиенты становятся все более мобильными, фокус недавних исследований имеет тенденцию смещаться к исследованию потребления энергии на всех уровнях систем связи [75].

Для максимизации энергоэффективности часто учитываются так называемые метрики «бит-на-джоуль» [76] или «пропускная способность на

один джоуль» [77]. Известны различные подходы к учету энергоэффективности. Некоторые включают в себя, например, управление мощностью с помощью алгоритмов, которые оптимизируют пропускную способность с учетом ограничения на мощность передачи [78], [79], а также адаптацию как общей мощности передачи, так и ее распределения по каналам согласно информации об их состоянии [71], [80].

Однако подавляющее большинство существующих информационно-теоретических подходов (см., например, [81] и [82]) при исследовании потребления энергии учитывают только мощность передачи данных. Как правило, клиентское устройство также потребляет дополнительную мощность схемы (circuit power), которая возникает независимо от скорости передачи данных [83], [84] и может оказаться сопоставимой с максимальной мощностью передачи. Таким образом, мощность схемы должна рассматриваться при оптимизации энергоэффективности в явном виде [85]. В недавних работах для установления оптимальных по энергии параметров связи предлагается сбалансировать потребляемую мощность передачи и схемы, используя теорию оптимизации [71], [86]. Эти результаты показывают, что обычный water-filling подход, увеличивающий время передачи, может быть оказаться не всегда оптимальным, поскольку потребление энергии схемы растет с продолжительностью передачи.

С уменьшением размера сот, происходящим с целью повышения пропускной способности беспроводных систем, диапазон охвата сотовых, локальных и персональных сетей все чаще перекрывается. Также стоит отметить, что современные городские беспроводные сети часто включают области, где по определению доступны разные технологии связи [87], [88]. Пока эти технологии занимают непересекающиеся полосы частот [88], они могут сосуществовать одновременно без какого-либо значительного ухудшения производительности сетей. Это создает привлекательную возможность для совместного использования нескольких сетей радиодоступа для улучшения качества беспроводной связи в целом [89].

За последние несколько лет в литературе появились [90] различные исследования межсетевых решений в пределах базовой (core network) и других участков сети, включающие бесшовную мобильность (seamless

mobility) между технологиями 3GPP и WLAN, надежный доступ к сервисам 3GPP для WLAN-устройств и поддержку функции обнаружения и выбора сети (Access Network Discovery Selection function, ANDSF) [91]. В частности, недавно была изучена проблема выбора сети в гетерогенной беспроводной среде с использованием протоколов IEEE 802.21 и IEEE 1900.4 [49]. Мы подчеркиваем, что наш основной акцент в этом разделе заключается, однако, в совместном использовании нескольких сетей, сосуществующих на уровне сети радиодоступа (Radio Access Network, RAN), что обеспечивает более гибкое управление параметрами передачи [92]. Мы также ожидаем, что эта работа будет полезна в текущих обсуждениях 3GPP по радиосвязи WLAN/3GPP [27].

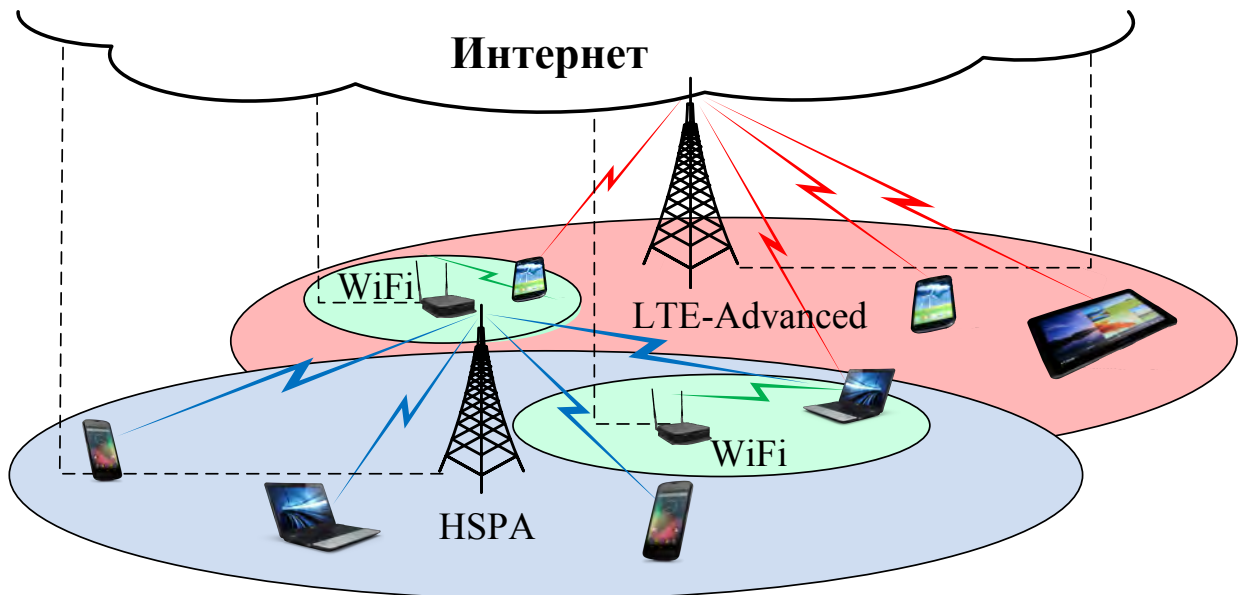


Рисунок 2.6 — Пример топологии гетерогенной сети

Мы ожидаем, что интеллектуальное взаимодействие между несколькими технологиями радиодоступа (такими как LTE-Advanced, HSPA, WiMAX, WiFi, Bluetooth, ZigBee и т. д.) позволит эффективно работать с несколькими радиоприемниками одновременно и, таким образом, формировать единый беспроводной пользовательский интерфейс. Для достижения этой цели, а также для увеличения пропускной способности системы и улучшения непрерывности обслуживания [93] необходимо обеспечить совместную работу технологий как ближнего, так и дальнего радиуса действия — например, WiFi и LTE-Advanced (см. рисунок 2.6). В связи с вышеизло-

женным, мы стремимся изучить потенциал алгоритма адаптивного управления мощностью в целях повышения энергоэффективности многодиапазонного устройства с использованием гетерогенной связи при различных параметрах.

2.4.2 Модель системы

В этом разделе вводятся основные предположения модели системы. В частности, рассматривается восходящее соединение одного пользовательского устройства, обладающего и управляющего несколькими радиоинтерфейсами (см. рисунок 2.6). В гетерогенной среде такое многодиапазонное устройство может эффективно использовать различные технологии радиодоступа. Этот раздел концентрируется на изучении достижимой скорости передачи данных, мощности и энергоэффективности, связанных с оптимальным управлением различными радиоинтерфейсами.

Наша аналитическая модель основана на предположениях, введенных ниже, и ее основные параметры суммируются в таблице 2.2. Здесь и далее изучается работа одного пользовательского устройства (см. рисунок 2.7), которому необходимо достичь фиксированной целевой скорости передачи r_0 (заданной, например, активным мобильным приложением). Другими словами, для обеспечения соответствия требований QoS этого устройства должна быть гарантирована скорость передачи данных не ниже r_0 .

Чтобы абстрагироваться от конкретной модели трафика, мы предполагаем, что трафик верхнего уровня является насыщенным, что соответствует случаю максимальных доступных итоговых выигрышей в системе. Также предполагается, что устройство, функционирующее в некотором гетерогенном деплоimente, может использовать K технологий радиодоступа RAT. Для выполнения требования по целевой скорости устройство может отправлять данные, используя одну или все имеющиеся RAT.

Потребляемая мощность устройства при использовании конкретной RAT (то есть мощность в данном канале связи) включает в себя две ком-

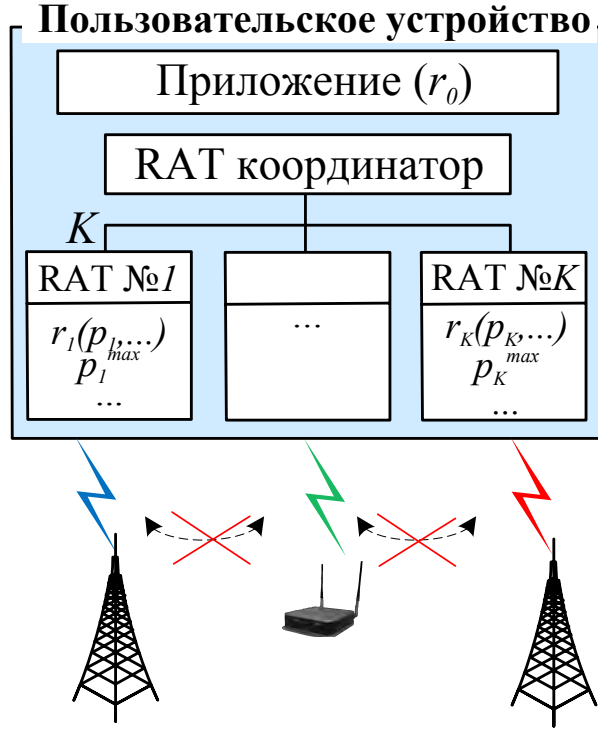


Рисунок 2.7 — Структура поддержки нескольких радиотехнологий доступа в мобильном устройстве

поненты: p_i^{tx} и p_i^c , где $p_i^{tx} \leq p_i^{max}$ — мгновенная мощность передачи, не превышающая максимальный предел мощности, а p_i^c — мощность схемы (circuit power), потребляемая независимо от того, какая мощность передачи используется в соответствующем канале. Руководствуясь теми же принципами, что и, например, [71], можно предположить, что значение p_i^c является постоянным (дополнительное энергопотребление самого устройства не учитывается). Однако метод, приводимый ниже, позволяет ослабить это предположение без потери общности.

В заданном канале i связь между мгновенной мощностью p_i^{tx} и достижимой скоростью передачи данных r_i может быть определена некоторой функцией $p_i^{tx}(r_i)$. Рассмотрим важный пример, когда это соотношение задается теоремой Шеннона-Хартли для канала «точка-точка» и $r_i(p_i^{tx}) = w_i \ln(1 + \gamma_i p_i^{tx})$ или, альтернативно

$$p_i^{tx}(r_i) = \frac{1}{\gamma_i} \left(e^{\frac{r_i}{w_i}} - 1 \right), \quad (2.1)$$

где p_i^{tx} — мощность передачи (Ватт), γ_i — отношение сигнал-шум, w_i — выделенная полоса пропускания канала (Гц), и r_i — мгновенная скорость передачи данных по каналу i (бит/с). Важно отметить, что основные результаты этой работы также применимы в случае произвольной функции скорости вида $r_i = A_i \ln(1 + B_i p_i)$ при любых положительных значениях параметров A_i и B_i , что может отражать более реалистичные свойства каналов.

Предполагается, что все каналы K в рассматриваемой модели не интерферируют, что может быть, например, связано с использованием неперекрывающихся частот. Наконец, можно предположить, что устройство может адаптивно управлять своей мощностью передачи в заданном канале $\{p_i^{tx}\}_{i=1}^K$, используя информацию о доступных РАТ, например, $\{w_i, \gamma_i, p_i^c, p_i^{\max}\}_{i=1}^K$.

Таким образом, общая скорость передачи данных устройства представляет собой сумму *индивидуальных* скоростей передачи данных по всем каналам $r = \sum_{i=1}^K r_i$. Соответственно, полная мощность передачи есть сумма мощностей, используемых для каждого канала в отдельности $p^{tx}(\mathbf{r}) = p^{tx}(r_1, \dots, r_K) = \sum_{i=1}^K p_i^{tx}(r_i)$, где $\mathbf{r} = (r_1, \dots, r_K) \in R^K$ — вектор передаваемых степеней.

Таким образом, общее энергопотребление устройства составляет $p(\mathbf{r}) = p^{tx}(\mathbf{r}) + p_c$, где $p_c = \sum_{i=1}^K p_i^c$. Для простоты предположим, что $p_i^c > 0$ для любого интерфейса i вне зависимости от того, активен он или нет.

Возможны и альтернативные формулировки (см., например, [94]); случай, когда $p_i^c = 0$, если $p_i^{tx} = 0$, будет кратко рассмотрен в разделе 2.4.3. Далее устанавливается оптимальная дисциплина управления мощностью, учитывающая как скорость передачи данных устройства, так и его энергопотребление.

В общем случае, пропускная способность канала представляет максимальную скорость передачи данных, при которой в системе возможна надежная связь. Пропускная способность определяется отношением сигнал-шум, отражающее характеристики распространения сигнала. На практике это зависит от множества факторов, таких как свойства беспроводной среды, высота антенны, расстояние до приемника и т.д.

Параметры аналитической модели

Обозначение	Параметр
r_0	Целевая скорость передачи данных
K	Количество доступных каналов (радиоинтерфейсов)
r	Общая скорость передачи данных устройства
p^{tx}	Общая мощность передачи пользовательского устройства
p_i^{tx}	Передача мощности в канале i
p_i^{max}	Максимальная мощность передачи в канале i
p_i^c	Мощность схемы для канала i
r_i	Скорость передачи данных в канале i

Теорема Шеннона-Хартли (2.1) представляет собой подходящую модель, которая позволяет оценить эффективность практической беспроводной системы с помощью аналитических выкладок. Теорема моделирует тот факт, что пользовательское устройство может снизить потребляемую мощность передачи, в некоторой степени жертвуя скоростью передачи данных, что часто оказывается предпочтительнее для небольших мобильных устройств с питанием от батареи [7]. Однако при этом пользовательское устройство также должно обеспечить требуемую скорость $r_0 > 0$, что приводит к задаче оптимального управления мощностью.

В то время как использование результата теоремы Шеннона-Хартли в качестве преобразования мощность-пропускная способность может обеспечить интуицию о возможных методах уменьшения мощности передачи, подход данным подразделе этим не ограничивается и может быть воспроизведен для более широкого класса функций, которые удовлетворяют следующим критериям:

1. Связь между мощностью передачи и скоростью передачи данных в канале i представлена биективной функцией $p_i^{tx}(r_i)$, такой, что $p_i^{tx}(0) = 0$
2. Производная $\frac{dp_i^{tx}}{dr_i} > 0$ на интервале $[0, \infty)$ (т.е. функция $p_i^{tx}(r_i)$) непрерывно дифференцируема и монотонно возрастает. Далее, необходимо выполнение требования $\frac{d^2 p_i^{tx}}{dr_i^2} > 0$, т.е. функция $\frac{dp_i^{tx}}{dr_i}$ также должна монотонно возрастать относительно r_i .

2.4.3 Задача оптимизации энергоэффективности

В этом разделе рассматривается задача максимизации энергоэффективности многодиапазонного мобильного устройства. Формулируется общая постановка задачи, не ограничивающаяся конкретной функцией отображения скорости. В качестве переменной рассматривается достижимая скорость передачи данных по каждому доступному каналу связи (которому соответствует своя мощность передачи). Затем перечисляются некоторые реалистичные ограничения и формулируется задача оптимизации энергоэффективности с ограничениями. Далее, эта задача оптимизации решается напрямую, т.е. при помощи подхода Каруша-Куна-Таккера [95], путем составления системы соответствующих уравнений и неравенств. Для получения оптимального решения, находятся стационарные точки целевой функции для задачи без ограничений, и далее решается поставленная задача при заданных ограничениях.

Общие утверждения

Ограничения задачи, следующие из практических соображений

В рамках рассматриваемой задачи управления мощностью мобильного устройства с несколькими радиointерфейсами необходимо ввести следующие реалистичные ограничения [96; 97]. Рассматривается минимальная общая скорость передачи данных по всем каналам (в свою очередь, это означает, что используется хотя бы один из доступных каналов K)

$$r = \sum_{i=1}^K r_i \geq r_0 > 0.$$

Также устанавливается естественное ограничение на достижимую скорость передачи данных r_i (и, следовательно, на p_i^{tx}), таким образом, что она не

может быть отрицательной

$$r_i \geq 0, \quad i = \overline{1, K},$$

где выражение $\overline{1, K}$ обозначает множество индексов $1, 2, \dots, K-1, K$. Наконец, учитывается максимально допустимая мощность передачи

$$p_i^{tx}(r_i) \leq p_i^{\max}, \quad i = \overline{1, K}.$$

В дальнейшем для краткости будем обозначать p_i^{tx} как p_i . Кроме того, поскольку функция $p_i(r_i)$ является биективной, можно использовать эквивалентную формулировку

$$r_i \leq r_i^{\max}, \quad i = \overline{1, K},$$

где $r_i(p_i^{\max})$ может быть задана функцией, обратной к (2.1).

Целевая функция В этом подразделе рассматривается задача максимизации энергоэффективности пользовательского устройства, представленной отношением общей скорости передачи данных r к общей мощности p , затраченной пользовательским устройством

$$\eta(\mathbf{r}) = \eta(r_1, \dots, r_K) = \frac{r}{p} = \frac{\sum_{i=1}^K r_i}{\sum_{i=1}^K p_i(r_i) + p_c},$$

где r_i и p_i — скорость передачи данных и мощность в канале i и $p_c = \sum_{i=1}^K p_i^c$. Таким образом, ниже можно сформулировать исходную задачу оптимизации с точки зрения энергоэффективности пользовательского устройства.

Исходная задача с ограничениями (Original Constrained Problem, OCP)

$$\max_{\{r_i\}_{i=1}^K} \eta(\mathbf{r}) = \max_{\{r_i\}_{i=1}^K} \frac{\sum_{i=1}^K r_i}{\sum_{i=1}^K p_i(r_i) + p_c},$$

которая подчиняется ограничениям, описанным в разделе 2.4.3. Следовательно, общая оптимальная скорость передачи данных складывается из

компонент вектора отдельных оптимальных скоростей передачи данных в каждом отдельном канале

$$\mathbf{r}^* = \arg \max_{\{r_i\}_{i=1}^K} \eta(\mathbf{r}).$$

В связи с тем, что решение ОСР в текущей форме может оказаться достаточно сложным, можно заметить, что на самом деле было бы проще рассмотреть минимизацию функции $U(\mathbf{r})$, обратной к $\eta(\mathbf{r})$. Такое преобразование возможно в том случае, если $\sum_{i=1}^K r_i \neq 0$. Такая (эквивалентная) проблема оптимизации и рассматривается далее. Здесь можно заметить, что $\eta(\mathbf{r}) > 0$ для любого вектора \mathbf{r} , и, как следствие, компоненты этого вектора неотрицательны, т.е. $R_i \geq 0$, $i = \overline{1, K}$.

Эквивалентная задача с ограничениями (ЕСР)

$$\min_{\{r_i\}_{i=1}^K} \frac{1}{\eta(\mathbf{r})} = \min_{\{r_i\}_{i=1}^K} U(\mathbf{r}) = \min_{\{r_i\}_{i=1}^K} \frac{\sum_{i=1}^K p_i(r_i) + p_c}{\sum_{i=1}^K r_i}, \quad (2.2)$$

которая также подчиняется ограничениям, описанным в разделе 2.4.3.

Наконец, можно переформулировать ЕСР как

$$\min_{\{r_i\}_{i=1}^K} U(\mathbf{r}) = \min_{\{r_i\}_{i=1}^K} \frac{\sum_{i=1}^K p_i(r_i) + p_c}{\sum_{i=1}^K r_i},$$

где должно выполняться

$$\begin{aligned} \phi(\mathbf{r}) &= r_0 - \sum_{i=0}^K r_i \leq 0, \\ f_i(r_i) &= -r_i \leq 0, \quad i = \overline{1, K}, \\ g_i(r_i) &= r_i - r_i^{\max} \leq 0, \quad i = \overline{1, K}. \end{aligned} \quad (2.3)$$

Условия Каруша-Куна-Таккера Для решения ЕСР при заданных ограничениях в виде неравенств, обычно применяется подход Каруша-Куна-Таккера (Karush-Kuhn-Tucker, ККТ). Ниже приводится условие стационарности и записывается система уравнений для нахождения оптималь-

ного решения ЕСР

$$\begin{aligned} \frac{\partial U(\mathbf{r})}{\partial r_i} + \sum_{i=1}^K \lambda_i \frac{dg_i(r_i)}{dr_i} + \sum_{i=1}^K \mu_i \frac{df_i(r_i)}{dr_i} + \beta \frac{d\phi(\mathbf{r})}{dr_i} = 0 &\Leftrightarrow \\ \Leftrightarrow \frac{\frac{dp_i}{dr_i} \cdot r - (\sum_{i=1}^K p_i + p_c)}{r^2} + \lambda_i - \mu_i - \beta = 0. \end{aligned}$$

Исходные ограничения могут быть переформулированы как

$$\begin{aligned} r_i - r_i^{\max} &\leq 0, \quad i = \overline{1, K}, \\ r_i &\geq 0, \quad i = \overline{1, K}, \\ \sum_{i=1}^K r_i - r_0 &\geq 0. \end{aligned}$$

Кроме того, условия неотрицательности представлены следующими неравенствами

$$\begin{aligned} \lambda_i &\geq 0, \quad i = \overline{1, K}, \\ \mu_i &\geq 0, \quad i = \overline{1, K}, \\ \beta &\geq 0, \end{aligned}$$

где λ_i , μ_i и β — множители Лагранжа.

Наконец, условия дополняющей нежесткости выливаются следующим образом

$$\begin{aligned} \lambda_i(r_i - r_i^{\max}) &= 0, \quad i = \overline{1, K}, \\ \mu_i r_i &= 0, \quad i = \overline{1, K}, \\ \beta \left(\sum_{i=1}^K r_i - r_0 \right) &= 0. \end{aligned}$$

Таким образом, чтобы получить оптимальное решение для ЕСР при указанных выше ограничениях, необходимо решить систему $3K + 1$ уравнений и $4K + 2$ неравенств. Важно отметить, что область, ограниченная данными неравенствами, должна быть непустой. В противном случае рассматриваемая задача не имеет решения.

Примечательно, что условия ККТ сами по себе не обеспечивают нахождения точек максимума/минимума. Вместо этого они определяют только стационарные точки (где градиент равен нулю), среди которых может находиться точка минимума. Поскольку решение системы многих уравнений и неравенств напрямую — непростая задача, в дальнейшем мы подро-

но изложим наш подход к решению рассматриваемой задачи оптимизации в соответствии с [98; 99].

Решение задачи с ограничениями

Эквивалентная задача с ограничениями (Equivalent Constrained Problem, ECP) может быть сформулирована для произвольной функции преобразования скорости передачи. В целях иллюстрации мы также рассмотрим пример, где связь между достижимой скоростью передачи данных и требуемой мощностью передачи задается теоремой Шеннона-Хартли (2.1). Однако возможны и расширенные формулировки на основе более сложных функций практических беспроводных систем.

Случай задачи без ограничений Для того, чтобы найти стационарные точки, лежащие как в рассматриваемой области, так и вне ее, следует вернуться к более простой проблеме оптимизации с той же целевой функцией (2.2) для $\mathbf{r} \neq \mathbf{0}$, но без каких-либо ограничений. Далее это решение применяется к целевому ECP (2.3).

Эквивалентная задача без ограничений (Equivalent Unconstrained Problem, EUP)

$$\min_{\{r_i\}_{i=1}^K} U(\mathbf{r}) = \min_{\{r_i\}_{i=1}^K} \frac{\sum_{i=1}^K p_i + p_c}{\sum_{i=1}^K r_i}. \quad (2.4)$$

Теорема 1. *Оптимальное решение EUP (2.4) может существовать только в том случае, если трансцендентное уравнение $Xe^X = \alpha$ имеет вещественные корни*

$$\alpha = \frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{k=1}^K w_k} \prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}}.$$

Оптимальный вектор оптимальных мощностей $\mathbf{p} = (p_1, \dots, p_K)$ может быть найден следующим образом

$$p_i = -\frac{1}{\gamma_i} + w_i e \prod_{j=1}^K (w_j \gamma_j)^{\frac{-w_j}{\sum_{k=1}^K w_k}} \exp W(\alpha), \quad i = \overline{1, K}, \quad (2.5)$$

где $W(\alpha)$ – корень трансцендентного уравнения (см. рисунок 2.8) и известен как функция Ламберта [100]. Процедура выбора конкретного корня зависит от значения α и подробно описана в ниже.

Далее докажем Теорему 1 путем применения математической индукции.

Доказательство. Чтобы найти оптимальное решение задачи без ограничений (EUP), необходимо проверить стационарные точки, удовлетворяющие следующему условию

$$\begin{aligned} \frac{\partial U(\mathbf{r})}{\partial r_j} = 0 &\Leftrightarrow \frac{\frac{dp_j}{dr_j} \sum_{i=1}^K r_i - (\sum_{i=1}^K p_i + p_c)}{(\sum_{i=1}^K r_i)^2} = 0 \Leftrightarrow \\ &\Leftrightarrow \frac{\frac{dp_j}{dr_j} \cdot c - (\sum_{i=1}^K p_i + p_c)}{r^2} = 0. \end{aligned} \quad (2.6)$$

Далее мы находим искомые точки. Во-первых, для этого необходимо вычислить производные для отдельных функций распределения мощности

$$\frac{dp_j}{dr_j} = \frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}}, \quad j = \overline{1, K}. \quad (2.7)$$

Подставим выражение (2.7) для производной функции p_j в условие для стационарных точек (2.6)

$$\frac{\partial U(\mathbf{r})}{\partial r_j} = \frac{\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} \sum_{i=1}^K r_i - \left[\sum_{i=1}^K \frac{1}{\gamma_i} (e^{\frac{r_i}{w_i}} - 1) + p_c \right]}{\left(\sum_{i=1}^K r_i \right)^2} = 0, \quad j = \overline{1, K}. \quad (2.8)$$

Следовательно, из (2.8) получаем достаточное условие для стационарных точек:

$$\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} \sum_{i=1}^K r_i - \sum_{i=1}^K \frac{1}{\gamma_i} e^{\frac{r_i}{w_i}} + \left[\sum_{i=1}^K \frac{1}{\gamma_i} - p_c \right] = 0, \quad j = \overline{1, K}.$$

Объединяя элементы, которые зависят и не зависят от j отдельно, получаем для каждого $j = \overline{1, K}$

$$\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} = \frac{\sum_{i=1}^K \frac{1}{\gamma_i} e^{\frac{r_i}{w_i}} - \left[\sum_{i=1}^K \frac{1}{\gamma_i} - p_c \right]}{\sum_{i=1}^K r_i}, \quad j = \overline{1, K}. \quad (2.9)$$

Обозначим правую часть уравнения (2.9), постоянную по индексу j , как D :

$$\frac{\sum_{i=1}^K \frac{1}{\gamma_i} e^{\frac{r_i}{w_i}} - \left[\sum_{i=1}^K \frac{1}{\gamma_i} - p_c \right]}{\sum_{i=1}^K r_i} = D. \quad (2.10)$$

Следовательно, для левой части

$$\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} = D, \quad j = \overline{1, K}. \quad (2.11)$$

Здесь r_j выражается через D следующим образом

$$r_j = w_j \ln(w_j D) + w_j \ln \gamma_j, \quad j = \overline{1, K}. \quad (2.12)$$

Важно заметить, что $\frac{1}{\gamma_j} e^{\frac{r_j}{w_j}} = w_j D$, $j = \overline{1, K}$, тогда подстановка (2.12) в (2.10) приводит к

$$D \sum_{j=1}^K w_j - \left[\sum_{j=1}^K \frac{1}{\gamma_j} - \sum_{j=1}^K p_j^c \right] = D \left(\sum_{j=1}^K w_j \ln(w_j D) + \sum_{j=1}^K w_j \ln \gamma_j \right).$$

Упрощая приведенное выше уравнение, можно получить следующее

$$D \sum_{j=1}^K w_j \ln(w_j D) + D \sum_{j=1}^K w_j \ln \gamma_j - \sum_{j=1}^K w_j D = - \left[\sum_{j=1}^K \frac{1}{\gamma_j} - \sum_{j=1}^K p_j^c \right].$$

Переупорядочивая приведенное выше выражение, устанавливаем, что

$$D \ln \left\{ \prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \frac{D}{e} \right\} = \frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{\sum_{j=1}^K w_j}. \quad (2.13)$$

Введем новую переменную

$$X = \ln \left[\prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \frac{D}{e} \right].$$

Тогда для D выполняется

$$D = e^{(X+1)} \prod_{j=1}^K w_j^{\frac{-w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{-w_j}{\sum_{k=1}^K w_k}}.$$

С учетом замены переменной можно переписать уравнение (2.13) следующим образом

$$X e^X = \frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{j=1}^K w_j} \prod_{i=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}}. \quad (2.14)$$

Из уравнения (2.14) можно получить значение X и, соответственно, выражение для D

$$\begin{aligned} X &= W \left(\frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{j=1}^K w_j} \prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \right), \\ D &= e \left(\prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \right)^{-1} \times \\ &\times \exp \left\{ W \left(\frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{j=1}^K w_j} \prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \right) \right\}, \end{aligned} \quad (2.15)$$

где $W(x)$ — это W -функция Ламберта [100], представляющая решение следующего уравнения

$$Xe^X = \alpha. \quad (2.16)$$

Следует отметить, что трансцендентное уравнение (2.16) имеет ровно один вещественный корень $x = W(\alpha)$ в области $\{-\frac{1}{e}\} \cup [0, \infty)$, а также два вещественных корня, если $\alpha \in (-\frac{1}{e}, 0)$ (это следует из формы функции Ламберта, имеющей несколько ветвей, см. рисунок 2.8). Здесь мы рассмотрим значение из основной (верхней) ветви, что объясняется Утверждением 1. В случае, когда $\alpha < -\frac{1}{e}$, можно заключить, что оптимального решения в R^K в рассматриваемой задаче оптимизации нет. Если существует реальное значение $W(\alpha)$, то можно подставить выражение для D (2.15) в (2.12)

$$\begin{aligned} r_i &= w_i \ln(w_i D) + w_i \ln \gamma_i = w_i \ln \gamma_i + w_i \ln w_i + w_i \ln D = \\ &= w_i \ln \gamma_i + w_i \ln w_i + w_i \left\{ 1 - \sum_{j=1}^K \frac{w_j}{\sum_{k=1}^K w_k} \ln(w_j \gamma_j) \right\} \times \\ &\times W \left(\frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{k=1}^K w_k} \prod_{j=1}^K w_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \prod_{j=1}^K \gamma_j^{\frac{w_j}{\sum_{k=1}^K w_k}} \right). \end{aligned}$$

Окончательное выражение для скорости передачи данных r_i задается как

$$\begin{aligned} r_i &= w_i \ln(\gamma_i w_i) + w_i \left\{ 1 - \frac{\sum_{j=1}^K w_j \ln(w_j \gamma_j)}{\sum_{k=1}^K w_k} \right\} + \\ &+ w_i W \left(\frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{k=1}^K w_k} \prod_{j=1}^K (w_j \gamma_j)^{\frac{w_j}{\sum_{k=1}^K w_k}} \right). \quad (2.17) \end{aligned}$$

В этом случае получена стационарная точка $\mathbf{r} = \{r_i\}_{i=1}^K$. Оптимальный уровень мощности для работы в одном канале может быть найден по формуле (2.1), а окончательная формула для $p_i = w_i D - \frac{1}{\gamma_i}$ может быть выписана как

$$p_i = -\frac{1}{\gamma_i} + w_i e \prod_{j=1}^K (w_j \gamma_j)^{\frac{-w_j}{\sum_{k=1}^K w_k}} \times$$

$$\times \exp \left\{ W \left(\frac{-\sum_{j=1}^K \frac{1}{\gamma_j} + \sum_{j=1}^K p_j^c}{e \sum_{k=1}^K w_k} \prod_{j=1}^K (w_j \gamma_j)^{\frac{w_j}{\sum_{k=1}^K w_k}} \right) \right\}, \quad (2.18)$$

что и требовалось доказать. \square

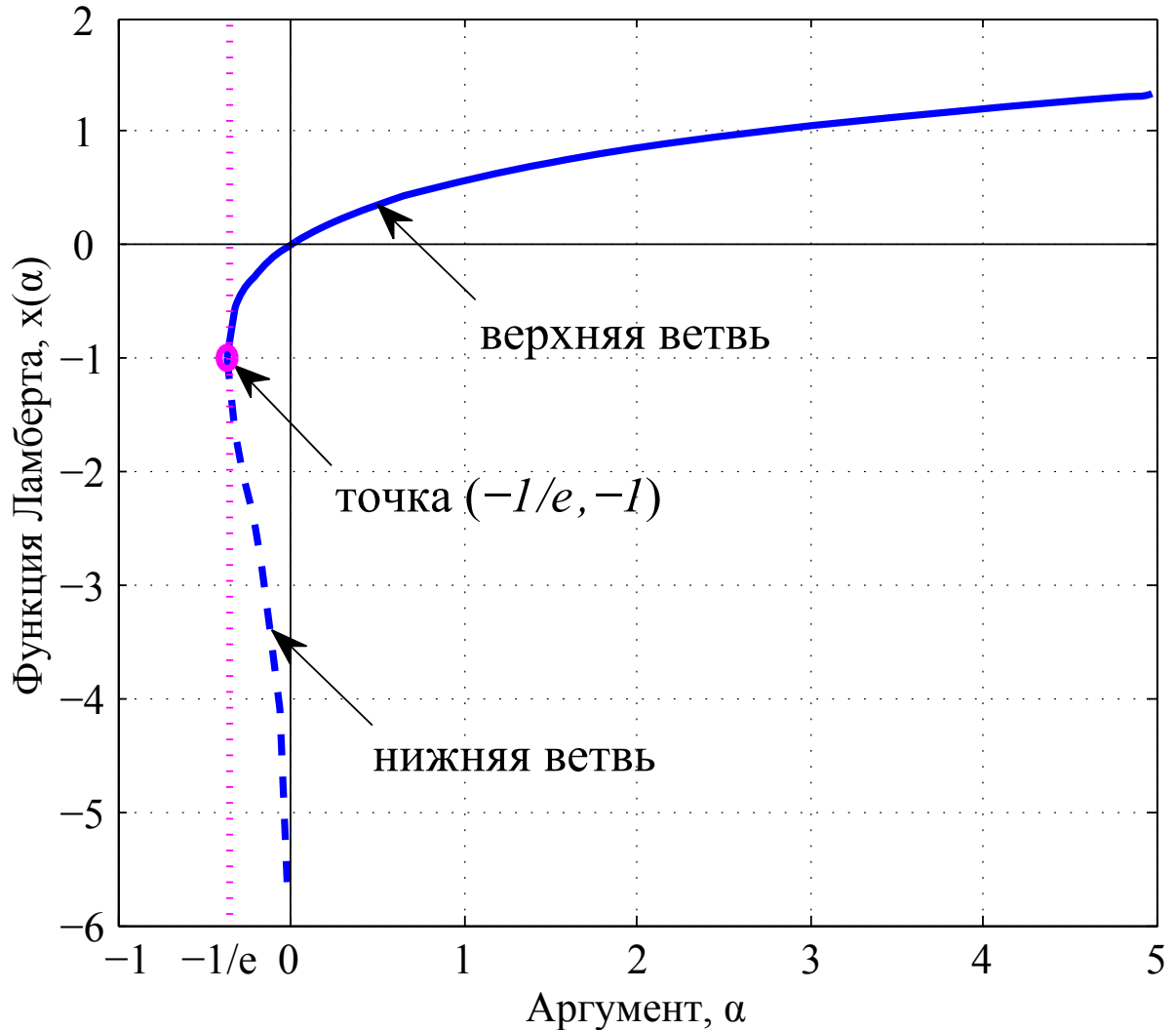


Рисунок 2.8 — График функции Ламберта

Трансцендентное уравнение $Xe^X = \alpha$ имеет вещественные корни тогда и только тогда, когда $\alpha \geq -1/e$. В этом случае может существовать либо один, либо два вещественных корня, в зависимости от выбора ветви функции $W(\alpha)$ [100]. Обсудим теперь, каким образом следует выбрать ветвь функции Ламберта $W(\alpha)$. Обозначим корень, полученный через верхнюю ветвь, как $X_0 = W_0(\alpha)$, а тот, который получен через нижнюю ветвь, как $X_1 = W_1(\alpha)$, для всех $\alpha \in [-1/e, 0)$.

Утверждение 1. Для заданных параметров системы возможны четыре разных случая

1. Если $\alpha < -\frac{1}{e}$, уравнение $Xe^X = \alpha$ не имеет вещественных корней, а стационарной точки функции $U(\mathbf{r})$ не существует.

2. Если $\alpha = -\frac{1}{e}$, то уравнение $Xe^X = \alpha$ имеет один реальный корень $X = -1$, и, следовательно, может существовать не более одной стационарной точки функции $U(\mathbf{r})$.

3. Если $-\frac{1}{e} < \alpha < 0$, то существуют ровно две разные стационарные точки $\mathbf{p}^{(0)}$ и $\mathbf{p}^{(1)}$ соответствующие X_0 и X_1 . Кроме того, для всех $i = \overline{1, K}$: $p_i^{(0)} > p_i^{(1)}$ и по меньшей мере одна из компонент вектора $\mathbf{p}^{(1)}$ отрицательна, т. е. существует такой индекс $i_0 \in \{i\}_{i=1}^K$, что $p_{i_0}^{(1)} < 0$.

4. Если $\alpha \in (0, \infty)$, то существует ровно одно оптимальное решение, т.е. оптимальный вектор мощностей \mathbf{p} имеет только одно значение.

Доказательство. Первый, второй и четвертый случаи тривиальны и следуют из свойств функции Ламберта. Однако все эти случаи не гарантируют наличия стационарных точек. По этой причине третий случай более подробно рассматривается в Утверждении 2 ниже.

Для третьего случая уравнение $Xe^X = \alpha$ имеет два вещественных корня iff $-\frac{1}{e} < \alpha < 0$. Следовательно, существуют ровно две различные стационарные точки $\mathbf{p}^{(0)}$ и $\mathbf{p}^{(1)}$, связанные с корнями уравнения. Если $\alpha \in \{-\frac{1}{e}\} \cup (0, \infty)$, то вектор \mathbf{p} имеет только одно значение в силу свойств функции Ламберта. Теперь можно рассмотреть интервал $-\frac{1}{e} < \alpha < 0$ и показать, что $p_i^{(0)} > p_i^{(1)}$ для всех $i = \overline{1, k}$.

Можно переписать выражение (2.5) для мощности

$$p_i^{(0)} = \frac{w_i e^{W_0(\alpha)+1}}{\prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}}} - \frac{1}{\gamma_i}, p_i^{(1)} = \frac{w_i e^{W_1(\alpha)+1}}{\prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}}} - \frac{1}{\gamma_i}.$$

Поскольку $W_0(\alpha) > W_1(\alpha)$, то можно заключить, что $p_i^{(0)} > p_i^{(1)}$.

Если рассмотреть большее значение $p_i^{(1)}$ в области $-\frac{1}{e} \leq \alpha < 0$ и учесть тот факт, что $W_1(\alpha) < -1$, то можно записать следующее

$$p_i^{(1)} = \frac{w_i e^{W_1(\alpha)+1}}{\prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}}} - \frac{1}{\gamma_i} < \frac{w_i \gamma_i - \prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}}}{\prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}} \gamma_i}. \quad (2.19)$$

Пусть индекс i_0 — такой, что $i_0 = \arg \min_i (w_i \gamma_i)$. Следовательно, $w_{i_0} \gamma_{i_0} \leq w_i \gamma_i$ для любых $i = \overline{1, K}$ и

$$\prod_{k=1}^K (w_k \gamma_k)^{\frac{w_k}{\sum_{j=1}^K w_j}} \geq \prod_{k=1}^K (w_{i_0} \gamma_{i_0})^{\frac{w_k}{\sum_{j=1}^K w_j}} = w_{i_0} \gamma_{i_0}. \quad (2.20)$$

Это означает, что $p_i^{(1)} < 0$, и существует, как минимум, одна отрицательная компонента вектора $\mathbf{p}^{(1)}$, что и завершает доказательство. \square

Исходя из этого свойства, можно опустить рассмотрение нижней ветви функции $W(\alpha)$. Далее отдельно рассматриваются два возможных диапазона α , а именно $\alpha \geq -1/e$, когда есть хотя бы одна стационарная точка, и $\alpha < -1/e$, когда стационарных точек нет.

Случай $\alpha \geq -1/e$ (существует стационарная точка) Используя Утверждение 1, можно сформулировать условие, согласно которому стационарная точка является искомым максимумом.

Утверждение 2. Если вектор \mathbf{p} удовлетворяет условию Теоремы 1 и $\alpha > -1/e$, тогда

1. Соответствующий оптимальный вектор \mathbf{r} , полученный через верхнюю ветвь функции Ламберта, является точкой локального минимума функции $U(\mathbf{r})$ и, соответственно, локального максимума функции $\eta()$.

2. Соответствующий оптимальный вектор \mathbf{r} , полученный через нижнюю ветвь функции Ламберта ($-1/e < \alpha < 0$), является точкой локального максимума функции $U(\mathbf{r})$ и, соответственно, локального минимума функции $\eta(\mathbf{r})$.

Доказательство. Далее рассматриваются случаи 2, 3 и 4 Утверждения 1 ($\alpha \geq -1/e$ во всех случаях). Можно начать с общих рассуждений для $\alpha > -1/e$ и затем разделить объединенные случаи 3 и 4 Утверждения 1 на следующие варианты (1) $\alpha > 0$ или $-1/e < \alpha < 0$, когда рассматривается верхняя ветвь функции Ламберта, и (2) $-1/e < \alpha < 0$, когда рассматривается нижняя ветвь функции Ламберта.

Для доказательства необходимо вычислить матрицу Гессе функции $U(\mathbf{r})$. Таким образом, следует выписать $\frac{\partial^2 U(\mathbf{r})}{\partial r_j \partial r_k}$, $k \neq j$ и затем перейти к $\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2}$. Для обоих вычислений используется выражение для первой производной

$$\frac{\partial U(\mathbf{r})}{\partial r_j} = \frac{\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} \sum_{i=1}^K r_i - \left[\sum_{i=1}^K \frac{1}{\gamma_i} (e^{\frac{r_i}{w_i}} - 1) + p_c \right]}{\left(\sum_{i=1}^K r_i \right)^2}, \quad j = \overline{1, K}.$$

Следовательно, недиагональный элемент матрицы Гессе, т.е. вторая производная по r_j, r_k , $k \neq j$, выглядит следующим образом

$$\frac{\partial^2 U(\mathbf{r})}{\partial r_j \partial r_k} = \frac{\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} - \frac{1}{\gamma_k w_k} e^{\frac{r_k}{w_k}}}{\left(\sum_{i=1}^K r_i \right)^2} = -\frac{\partial^2 U(\mathbf{r})}{\partial r_j \partial r_k} = 0, \quad j, k = \overline{1, K},$$

т.к. $\frac{U(\mathbf{r})}{\partial r_j \partial r_k} = \frac{U(\mathbf{r})}{\partial r_k \partial r_j}$. Следовательно, матрица Гессе является диагональной для обоих случаев, и можно обсудить их отдельно.

Далее вычислим диагональный элемент матрицы Гессе, т.е. вторую производную по r_j

$$\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2} \Big|_{\mathbf{r}} = \frac{\frac{1}{\gamma_j w_j} \frac{1}{w_j} e^{\frac{r_j}{w_j}} \sum_{i=1}^K r_i + 2 \left(\sum_{i=1}^K r_i \right)^{-1} \cdot 0}{\left(\sum_{i=1}^K r_i \right)^2}, \quad j = \overline{1, K},$$

где принимается во внимание тот факт, что оптимальный вектор удовлетворяет равенству $\frac{\partial U(\mathbf{r})}{\partial r_j} = 0$ для всех $j = \overline{1, K}$.

Используя выражение (2.17) в доказательстве Теоремы 1, можно вычислить $\sum_{i=1}^K r_i$

$$\begin{aligned} \sum_{i=1}^K r_i &= \sum_{i=1}^K w_i \ln(\gamma_i w_i) + \sum_{i=1}^K w_i \left\{ 1 - \frac{\sum_{j=1}^K w_j \ln(w_j \gamma_j)}{\sum_{k=1}^K w_k} + W(\alpha) \right\} \\ &= [1 + W(\alpha)] \sum_{i=1}^K w_i, \end{aligned}$$

где α задается Теоремой 1. Далее рассматриваются два разных случая, описанных в формулировке этой теоремы.

1. Известно, что для верхней ветви функции Ламберта $W(\alpha) \geq -1$, если $\alpha \geq -1/e$. Кроме того, $W(\alpha) = -1$ только в точке $\alpha = -1/e$. В связи с этим, выражение $\sum_{i=1}^K r_i$ и, следовательно, $\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2}$ всегда больше нуля в стационарной точке \mathbf{r} до тех пор, пока $\alpha > -1/e$ и необходимо рассматривать верхнюю ветвь функции Ламберта.

Если $\alpha > -1/e$ и верхняя ветвь функции Ламберта рассмотрены, то $\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2} > 0$ и матрица Гессе положительно определена в точке \mathbf{r} , откуда следует, что точка является локальным минимумом функции $U(\mathbf{r})$.

2. Если $-1/e < \alpha < 0$ и рассматривается нижняя ветвь функции Ламберта, то $W(\alpha) < -1$ и, следовательно, собственные значения диагонали матрицы Гессе отрицательны $\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2} < 0$. Это, в свою очередь, делает матрицу Гессе отрицательно определенной, а точку \mathbf{r} — локальным максимумом функции $U(\mathbf{r})$.

Наконец, можно отметить, что если $\alpha = -1/e$, то $W(\alpha) = -1$ и $\frac{\partial^2 U(\mathbf{r})}{\partial r_j^2}$ равен нулю, что приводит к вырожденной матрице Гессе. Однако в этом случае $\sum_{i=1}^K r_i = 0$, и полученная точка явно не лежит в области функции $U(\mathbf{r})$. Более того, этот момент также не интересует нас в терминах начальной задачи оптимизации, так как условие $\sum_{i=1}^K r_i = r_0 > 0$ было задано заранее.

□

Согласно Утверждению 1 в случае, когда $\alpha > -1/e$, существует один и только один локальный минимум функции $U(\mathbf{r})$ (т.е. локальный максимум $\eta()$).

Следствие 1. *Если $\alpha > -1/e$ и рассматривается верхняя ветвь функции Ламберта, то стационарная точка $\mathbf{r} > 0$ представляет собой глобальный максимум функции $\eta(\mathbf{r})$.*

Заметим, что в случае $\alpha = -1/e$ мы получаем $\sum_{i=1}^K r_i = 0$, который не принадлежит рассматриваемой области функции $U(\mathbf{r})$ и не входит в сферу наших интересов в силу требования целевой битовой скорости $r_0 > 0$.

Вышеописанный подход позволяет находить максимум функции, если существует стационарная точка. Однако также может быть, что $\alpha < -1/e$ и функция Ламберта не имеет вещественных корней. Тогда стационарные точки не существуют и рассматриваемая функция не имеет локального минимума.

Случай $\alpha < -1/e$ (стационарных точек не существует) Система неравенств (2.3) задает область определения целевой функции, компактную в силу ограниченности и замкнутости в R^K . Т.к. мы рассматриваем компактное множество в R^K , внутри которого целевая функция ограничена и непрерывна, то в случае отсутствия стационарных точек ($\alpha < -1/e$) максимум/минимум может быть достигнут только на границе этой области.

Утверждение 3. *Если $\alpha < -1/e$, то локального максимума функции $\eta(\mathbf{r})$ (или локального минимума $U(\mathbf{r})$) не существует. В силу того, что все компоненты вектора скорости неотрицательны, т.е. $\mathbf{r} \geq 0$, решение задачи (в данном случае, ограниченное только этим условием $\mathbf{r} \geq 0$) может быть найдено на плоскости, заданной равенством $r_i = 0, 1 \leq i \leq K$. Выбор компонент i подробно описан ниже.*

Доказательство. Зафиксируем индекс i и значения $r_j, j \neq i$ и рассмотрим функцию $\tilde{\eta}(r_i) = \frac{r_i + \sum_{j=1, j \neq i}^K r_j}{p_i(r_i) + \sum_{j=1, j \neq i}^K p_j(r_j)}$. Ниже доказывается, что $\tilde{\eta}(r_i) \rightarrow 0$ в случае $r_i \rightarrow +\infty$ и $\tilde{\eta}(0)$ стремится к положительной константе, если $r_i \rightarrow +0$,

из чего (в силу монотонности $\tilde{\eta}(r_i)$) следует, что максимум следует искать на одной из плоскостей, заданных в $r_i = 0$. Т.к. функция $\frac{dp_i}{dr_i}$ непрерывно дифференцируема и монотонно возрастает в пределах $(0, \infty)$, то согласно правилу Лопиталья

$$\lim_{r_i \rightarrow +\infty} \frac{r_i + \sum_{j=1, i \neq j}^K r_j}{p_i(r_i) + \sum_{j=1, i \neq j}^K p_j(r_j)} = \lim_{r_i \rightarrow +\infty} \frac{1}{\frac{dp(r_i)}{dr_i}} = 0. \quad (2.21)$$

Подобное рассуждение можно повторить для функции $\tilde{\eta}(\mathbf{r}) = \frac{\sum_{i=1}^K r_i}{\sum_{i=1}^K p_i + p_c}$, где множество $\{r_i^{(0)}\}_{i=1}^{k_1}$, $k_1 = \overline{0, K-1}$ фиксировано, и для всех $i = \overline{k_1+1, K}$ предполагается $r_i \rightarrow \infty$. Далее, можно получить

$$\begin{aligned} & \lim_{r_i \rightarrow \infty, i=\overline{k_1+1, K}} \frac{\sum_{i=1}^{k_1} r_i^{(0)} + \sum_{i=k_1+1}^K r_i^{(0)}}{\sum_{i=1}^{k_1} \frac{1}{\gamma_i} e^{\frac{r_i^{(0)}}{w_i}} + \sum_{i=k_1+1}^K \frac{1}{\gamma_i} e^{\frac{r_i}{w_i}} + p_c} \\ &= \lim_{r_i \rightarrow \infty, i=\overline{k_1+1, K}} \frac{\sum_{i=k_1+1}^K r_i^{(0)}}{\sum_{i=k_1+1}^K \frac{1}{\gamma_i} e^{\frac{r_i}{w_i}}} = +0. \end{aligned} \quad (2.22)$$

Выражение выше означает, что когда все или несколько компонент вектора \mathbf{r} стремятся к бесконечности, значение функции $\eta(\mathbf{r})$ бесконечно мало.

Если параметры системы не удовлетворяют условию $\alpha > -1/e$, то градиент $\nabla \eta \neq 0$. Это означает, что неотрицательная функция $\eta(\mathbf{r})$ достигает своего максимума на границе любого компактного подмножества $\tilde{r} = \{\mathbf{r} | r_i \in [0, C_1]\}$, где $C_1 \in R^+$ — любое достаточно большое число. Можно заметить, что хотя бы для одной из неотрицательных функций $p_i(r_i)$: $\frac{dp(r_i)}{dr_i} > 0$. Далее выводится предел функции $\eta(\mathbf{r})$, если $r_i \rightarrow 0$

$$\lim_{r_i \rightarrow +0} \frac{r_i + \sum_{j=1, i \neq j}^K r_j}{p_i(r_i) + \sum_{j=1, i \neq j}^K p_j(r_j)} = \lim_{r_i \rightarrow +0} \frac{1}{\frac{dp(r_i)}{dr_i}} = const > 0. \quad (2.23)$$

Из выражения для производной от $U(\mathbf{r})$ можно заключить, что $\frac{\partial U(\mathbf{r})}{\partial r_j} < 0$ для области $r_j > r_j^*$, $j = \overline{1, K}$, где \mathbf{r}^* — стационарная точка

функции $U(\mathbf{r})$. Следовательно, функция $U(\mathbf{r})$ монотонно убывает в области $r_j > r_j^*, j = \overline{1, K}$.

Из выражений (2.21) и (2.23) можно заключить, что максимальное значение для функции η лежит на одной из границ $r_i = 0$, что завершает доказательство. \square

Таким образом, можно установить, что точка глобального максимума для целевой функции энергоэффективности имеет по крайней мере одну нулевую составляющую. Следовательно, необходимо уменьшить размерность рассматриваемой задачи и в этом случае исключить компоненту с наименьшим вкладом в рост функции $\eta(\mathbf{r})$, т.е. найти

$$i = \arg \max_{j=1, K} \left\{ \frac{dp_j(r_j)}{dr_j} \Big|_{r_j=0} \right\}.$$

Далее, положим $r_i = 0$. В этом случае можно снизить размерность задачи оптимизации до $K - 1$, учитывая при этом следующее

$$\min_{\{r_j\}_{j=1, j \neq i}^K} U(\mathbf{r}) = \min_{\{r_j\}_{j=1, j \neq i}^K} \frac{\sum_{j=1, j \neq i}^K p_j(r_j) + p_i + \sum_{j=1}^K p_j^r}{\sum_{j=1, j \neq i}^K r_j + r_i},$$

при наличии тех же ограничений. Здесь r_i и $p_i = p_i(r_i)$ являются фиксированными значениями для i , тогда как r_i может быть либо равна нулю, либо следовать из другого условия границы (более подробное описание приведено ниже). Можно продолжать процесс решения задачи оптимизации размерности $K - 1$ или уменьшать размерность, если необходимо, до тех пор, пока $K \geq 1$. Если $K = 1$, то такой подход приводит к одномерной задаче с частным решением r_j^* , которое должно быть снова проверено с учетом имеющихся ограничений.

Решение задачи с ограничениями Используя полученное решение r^* задачи EUP без ограничений, можно перейти к решению системы неравенств (2.4). Во-первых, можно предположить, что аргумент функции Ламберта превышает значение $-\frac{1}{e}$, так что целевая функция $U(\mathbf{r})$ задачи ECP с ограничениями (2.3) имеет локальный минимум в соответствующей точке r^* . Очевидно, что r^* может не соответствовать условиям системы (2.4).

Чтобы найти решение задачи ЕСР, необходимо проверить, что локальный минимум r^* удовлетворяет всем неравенствам

$$r_i \geq 0, \quad i = \overline{1, K},$$

$$r_i - r_i^{\max} \leq 0, \quad i = \overline{1, K},$$

$$\sum_{i=1}^K r_i - r_0 \geq 0.$$

Если точка r^* удовлетворяет вышеприведенным соотношениям, то она представляет собой решение системы (2.4), и, следовательно, оптимальное решение задачи (2.3). В противном случае, если хотя бы одно из неравенств не выполняется, локальный минимум функции $U(\mathbf{r})$ будет расположен за пределами области, заданной ограничениями, а оптимальное решение лежит границе этого домена. Общая схема нашего решения подробно описана алгоритмом 2.

В частности, если рассмотреть ограничения $r_i - r_i^{\max} = 0$, аналогичным образом можно уменьшить размерность задачи ЕСР $K = 1$. Разница с ЕУР заключается в том, что при переходе к задаче размерности $K - 1$, если не выполняется ограничение с индексом K (без потери общности можно отсортировать вектор), то фиксируются компоненты при ненулевых значениях $r_K = r_K^{\max}$. Следовательно, мы приходим к слегка иным выражениям, приведенным в Теореме 2 ниже.

Теорема 2. *Если у ЕСР есть решение $r^* \in R^K$ и существует хотя бы один индекс i , такой, что $r_i > r_i^{\max}$, то оптимальное решение должно быть расположено среди точек*

$$p_i = e^{1 - \frac{r_K}{w_K}} \prod_{k=1}^{K-1} (w_k \gamma_k)^{\frac{-w_k}{\sum_{j=1}^{K-1} w_j}} \cdot e^{W(\alpha)} - \frac{1}{\gamma_i}, \quad i = \overline{1, K-1},$$

$$p_K = \frac{1}{\gamma_K} \left(e^{\frac{r_K^{\max}}{w_K}} - 1 \right).$$

```

r* ← FindSolution( $K, \{i\}_{i=1}^K$ )
{FindSolution}{ $K, \{i_j\}_{j=1}^K$ } while  $K \geq 1$  do
    end
    Решить EUP (задача без ограничений) согласно разделу 2.4.3
r* ← стационарная точка if  $\mathbf{r}^* \in \mathbb{R}^K$  then
    if  $r_i^* < 0$  then
        end
         $i_0 \leftarrow \max_{\{i\}_{i=1}^K | r_i^* < 0} \frac{dp_i}{dr_i}$ 
 $r_{i_0}^* \leftarrow 0$ 
 $\{r_i^*\}_{i=1, i \neq i_0}^K \leftarrow \text{FindSolution}(K - 1, \{i\}_{i=1, i \neq i_0}^K)$ 
    end
if  $r_i^* > r_i^{\max}$  then
    end
     $i_0 \leftarrow \max_{\{i\}_{i=1}^K | r_i^* > r_i^{\max}} \frac{dp_i}{dr_i}$ 
 $r_{i_0}^* \leftarrow r_i^{\max}$ 
 $\{r_i^*\}_{i=1, i \neq i_0}^K \leftarrow \text{FindSolutionMax}(K - 1, \{i\}_{i=1, i \neq i_0}^K)$ 
if  $\sum_{i=1}^K r_i^* < r_0$  then
    end
     $r_K^* = r_0 - \sum_{i=1}^K r_i^*$ 
 $\{r_i^*\}_{i=1}^{K-1} \leftarrow \text{FindSolutionSum}(K - 1, \{i\}_{i=1}^{K-1})$ 
return  $r^*$  //  $r^*$  — искомая точка
else
    end
    //  $\mathbf{r}^* \notin R^K$ , локальный максимум не существует
     $r^*$  лежит на границе  $r_i^* = 0$ 
     $i_0 \leftarrow \max_{\{i\}_{i=1}^K} \frac{dp_i}{dr_i}$ 
 $r_{i_0}^* \leftarrow 0$ 
 $\{r_i^*\}_{i=1, i \neq i_0}^K \leftarrow \text{FindSolution}(K - 1, \{i\}_{i=1, i \neq i_0}^K)$ 

{FindSolutionMax}{ $K, \{i_j\}_{j=1}^K$ }
Решать согласно Теореме 2
return  $r^*$ 
{FindSolutionSum}{ $K, \{i_j\}_{j=1}^K$ }
Решать согласно Теореме 3
return  $r^*$ 

```

Алгоритм 2: Решение ECP с ограничениями

Если оптимальное решение не принадлежит требуемому домену, то, решая задачу с ограничениями (ЕСР), можно исходить из рассмотрения всех возможных вариантов расположения границы максимума энергоэффективности. Сначала рассмотрим случай, когда $r_i > r_i^{\max}$. Это означает, что компонента i должна быть установлена на границе, т.е. $R_i = r_i^{\max}$. Если несколько компонент не удовлетворяют неравенствам (2.4), следует выбирать компоненту с наименьшим вкладом в рост энергоэффективности, и, следовательно

$$i = \arg \min_{j=1, K} \left\{ \frac{dp_j(r_j)}{dr_j} \mid r_j > r_i^{\max} \right\}.$$

Далее можно задать $r_i = r_i^{\max}$ и, тем самым, уменьшить размерность задачи оптимизации до $K - 1$, учитывая следующее

$$\min_{\{r_j\}_{j=1, j \neq i}^K} U(\mathbf{r}) = \min_{\{r_j\}_{j=1, j \neq i}^K} \frac{\sum_{j=1, j \neq i}^K p_j(r_j) + p_i(r_i^{\max}) + \sum_{j=1}^K p_j^r}{\sum_{j=1, j \neq i}^K r_j + r_i^{\max}},$$

при условии, что соблюдаются те же ограничения. Мы переставляем порядок переменных r_i без потери общности, таким образом, что r_i^{\max} становится последним в ряду по индексу K , когда $r_K = r_i^{\max}$.

Ниже предлагается доказательство Теоремы 2, в котором устанавливается оптимальный вектор \mathbf{p} в случае, если для решения ЕСР $r^* \in R^K$ существует хотя бы один индекс i , для которого справедливо $r_i > r_i^{\max}$.

Доказательство. Пусть $p_i(r_K^{\max}) = p_K$ и $r_K^{\max} = r_K$, тогда для любого индекса $j = \overline{1, K}, i \neq j$

$$\frac{\partial U(\mathbf{r})}{\partial r_j} = \frac{\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} \sum_{j=1}^K r_j - \left[\sum_{j=1}^K \frac{1}{\gamma_j} (e^{\frac{r_j}{w_j}} - 1) + \sum_{j=1}^K p_j^c \right]}{\sum_{j=1}^{K-1} r_j + r_K}.$$

Таким образом, необходимое условие стационарных точек выглядит как

$$\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} \sum_{j=1}^K r_j - \sum_{j=1}^{K-1} \frac{1}{\gamma_j} e^{\frac{r_j}{w_j}} + \left[\sum_{j=1}^K \frac{1}{\gamma_j} - \sum_{j=1}^K p_j^c - p_K \right] = 0,$$

которое выполняется для всех $i = \overline{1, K-1}$. Следуя той же логике, что и раньше, можно получить выражение для индекса $j = \overline{1, K-1}$

$$\frac{1}{\gamma_j w_j} e^{\frac{r_j}{w_j}} = \frac{\sum_{j=1}^{K-1} \frac{1}{\gamma_j} e^{\frac{r_j}{w_j}} + \frac{1}{\gamma_K} e^{\frac{r_K}{w_K}} - \left[\sum_{j=1}^K \frac{1}{\gamma_j} - \sum_{j=1}^K p_j^c \right]}{\sum_{j=1}^{K-1} r_j + r_K} = D.$$

Можно предположить ради удобства, что

$$\alpha = \frac{\frac{1}{\gamma_K} e^{\frac{r_K}{w_K}} - \sum_{i=1}^K \frac{1}{\gamma_i} + p_c}{e^{\sum_{i=1}^{K-1} w_i}} \prod_{i=1}^{K-1} w_i^{\frac{w_i}{\sum_{j=1}^{K-1} w_j}} \prod_{i=1}^{K-1} \gamma_i^{\frac{w_i}{\sum_{j=1}^{K-1} w_j}} \cdot e^{\frac{r_K}{w_K}}, \quad (2.24)$$

тогда получаем выражение для D

$$D = e^{\left(\prod_{i=1}^{K-1} w_i^{\frac{w_i}{\sum_{j=1}^{K-1} w_j}} \prod_{i=1}^{K-1} \gamma_i^{\frac{w_i}{\sum_{j=1}^{K-1} w_j}} \cdot e^{\frac{r_K}{w_K}} \right)^{-1}} \cdot e^{W(\alpha)}.$$

Подставляя это в выражение для достижимой скорости передачи данных (2.12), можно установить

$$\begin{aligned} r_i &= w_i \ln(w_i D) + w_i \ln \gamma_i = w_i \ln \gamma_i + w_i \ln w_i + w_i \ln D = \\ &= w_i \ln \gamma_i + w_i \ln w_i + w_i \left\{ 1 - \frac{r_K}{w_K} - \sum_{k=1}^{K-1} \frac{w_k}{\sum_{j=1}^{K-1} w_j} \ln(w_k \gamma_k) \right\} \cdot W(\alpha). \end{aligned}$$

Окончательные выражения для скорости передачи данных r_i и мощности p_i для любого канала $i = \overline{1, K-1}$ задаются следующим образом

$$r_i = w_i \ln(\gamma_i w_i) + w_i \left\{ 1 - \frac{r_K}{w_K} - \frac{\sum_{j=1}^K w_j \ln(w_j \gamma_j)}{\sum_{j=1}^K w_j} \right\} W(\alpha), \quad (2.25)$$

$$p_i = e^{1 - \frac{r_K}{w_K}} \prod_{k=1}^{K-1} (w_k \gamma_k)^{\frac{-w_k}{\sum_{j=1}^{K-1} w_j}} \cdot e^{W(\alpha)} - \frac{1}{\gamma_i}, \quad i = \overline{1, K-1}, \quad (2.26)$$

что и требовалось доказать. \square

В случае, когда ограничение $\sum_{i=1}^K r_i - r_0 = 0$ не выполняется, мы применяем аналогичную процедуру и устанавливаем $r_K = r_0 - \sum_{i=1}^{K-1} r_i$. Окончательные выражения приводятся в Теореме 3.

Теорема 3. Если у ЕСП есть решение $r^* \in R^K$ и $\sum_{i=1}^K r_i - r_0 < 0$, то оптимальное решение должно быть расположено среди точек

$$p_i = \frac{1}{\gamma_i} \left(e^{\frac{r_i}{w_i}} - 1 \right), \quad i = \overline{1, K},$$

где вектор $\{r_i\}_{i=1}^K$ задается следующим образом

$$r_i = w_i \frac{r_0 - \sum_{j=1}^K w_j \ln(\gamma_j w_j)}{\sum_{j=1}^K w_j} + w_i \ln(\gamma_i w_i), \quad i = \overline{1, K-1},$$

$$r_K = r_0 - \sum_{j=1}^{K-1} r_j.$$

Далее рассмотрим случай $\sum_{i=1}^K r_i - r_0 < 0$. Мы установили, что оптимальное решение расположено на плоскости $\sum_{j=1}^K r_j = r_0$, поэтому следует учитывать следующую задачу с ограничениями

$$\min_{\{r_j\}_{j=1}^K} U(\mathbf{r}) = \min_{\{r_j\}_{j=1}^K} \frac{\sum_{j=1}^K p_j + \sum_{j=1}^K p_j^r}{\sum_{j=1}^K r_j},$$

где

$$\sum_{j=1}^K r_j = r_0. \quad (2.27)$$

Теорема 3 определяет выражение для вектора оптимальных мощностей \mathbf{p} и вектора оптимальных скоростей \mathbf{r} в случае, если у ЕСП есть решение $\mathbf{r}^* \in R^K$ и $\sum_{i=1}^K r_i^* - r_0 < 0$. Ниже мы приводим доказательство Теоремы 3.

Доказательство. Можно выразить r_K из уравнения (2.27) следующим образом

$$r_K = r_0 - \sum_{j=1}^{K-1} r_j. \quad (2.28)$$

Это позволяет снизить размерность задачи, таким образом, наша система преобразуется в

$$\min_{\{r_j\}_{j=1}^{K-1}} U(\mathbf{r}) = \min_{\{r_j\}_{j=1}^{K-1}} \frac{\sum_{j=1}^{K-1} \frac{1}{\gamma_j} e^{\frac{r_j}{w_j}} + \frac{1}{\gamma_K} e^{\frac{r_0 - \sum_{j=1}^{K-1} r_j}{w_K}} + p_c}{r_0},$$

что эквивалентно следующей задаче

$$\min_{\{r_j\}_{j=1}^{K-1}} U_1(\mathbf{r}) = \min_{\{r_j\}_{j=1}^{K-1}} \left\{ \sum_{j=1}^{K-1} \frac{1}{\gamma_j} e^{\frac{r_j}{w_j}} + \frac{1}{\gamma_K} e^{\frac{r_0 - \sum_{j=1}^{K-1} r_j}{w_K}} \right\}. \quad (2.29)$$

Условие для стационарной точки задачи (2.29)

$$\frac{\partial U_1(\mathbf{r})}{\partial r_i} = \frac{1}{\gamma_i w_i} e^{\frac{r_i}{w_i}} - \frac{1}{\gamma_K w_K} e^{\frac{r_0 - \sum_{j=1}^{K-1} r_j}{w_K}} = 0, \quad i = \overline{1, K-1}.$$

Следуя той же логике, что и раньше, можно обозначить постоянную часть как D

$$\frac{1}{\gamma_i w_i} e^{\frac{r_i}{w_i}} = \frac{1}{\gamma_K w_K} e^{\frac{r_0 - \sum_{j=1}^{K-1} r_j}{w_K}} = D, \quad i = \overline{1, K-1}.$$

Отдельные скорости передачи данных r_i выражаются через D

$$r_i = w_i \ln(D \gamma_i w_i) = w_i \ln(D) + w_i \ln(\gamma_i w_i), \quad i = \overline{1, K-1},$$

и могут быть подставлены в выражение для r_K (2.28)

$$r_0 - \sum_{j=1}^{K-1} w_j \ln(D) - \sum_{j=1}^{K-1} w_j \ln(\gamma_j w_j) = w_K \ln(D \gamma_K w_K).$$

Упрощая это выражение, получим

$$r_0 - \sum_{j=1}^K w_j \ln(\gamma_j w_j) = \sum_{j=1}^K w_j \ln(D).$$

Далее, $\ln D$ выражается из приведенного выше уравнения

$$\ln(D) = \frac{r_0 - \sum_{j=1}^K w_j \ln(\gamma_j w_j)}{\sum_{j=1}^K w_j}.$$

Наконец, можно получить выражение для отдельной скорости передачи данных r_i

$$r_i = w_i \frac{r_0 - \sum_{j=1}^K w_j \ln(\gamma_j w_j)}{\sum_{j=1}^K w_j} + w_i \ln(\gamma_i w_i), \quad i = \overline{1, K-1},$$

что и требовалось доказать. □

В заключение, обсудим ситуацию $p_i^c = 0$ или, иными словами, когда конкретный канал (радиотехнология) не используется (т.е. $P_i = 0$). В разделе 3.4.1 предположим, что $p_i^c = q_i > 0$, где q_i — фиксированная константа. Однако из-за рекурсии в вычислениях мы можем ослабить это условие и рассмотреть функцию $p_i^c = q_i \cdot \mathbf{I}\{p_i = 0\}$, где $\mathbf{I}\{p_i = 0\}$ — индикаторная функция события. Поэтому, получив $p_i = 0$, мы продолжим оптимизацию на следующей итерации, используя $\tilde{p}_c = p_c - p_i^c$, чтобы рассматриваемый радиоинтерфейс оказался неактивен.

2.4.4 Численные результаты

В этом разделе мы предлагаем *три* сценария для исследования достижимой скорости передачи данных, соответствующей мощности передачи и энергоэффективности мобильного пользовательского устройства. Приведенные сценарии предназначены для иллюстрации энергоэффективной работы, достигнутой при помощи использования предлагаемого подхода, и сравнения с более простой эвристической схемой управления мощностью. Ниже мы представляем несколько (абстрактных и реалистичных) вариантов сети для изучения поведения пользовательского устройства.

Симметричный случай. Две радиотехнологии

Здесь мы предполагаем, что в любой момент времени (см. рисунок 2.9) мобильное устройство может использовать две RAT ($K = 2$). На рисунке базовая станция первой RAT расположена в точке $x_1 = 0$, а второй — в точке $x_2 = R$. Пользовательское устройство перемещается вдоль оси x между базовыми станциями, и его текущее местоположение равно $x \in (0, R)$.



Рисунок 2.9 — Геометрия сети для случая двух RAT

Мы используем стандартную модель распространения для микросот в городских районах, где затухание в канале (дБ) определяется как

$$\theta = 40 \log_{10} d + 7.8 - 18 \log_{10} h_s h_u + 2 \log_{10} f_c, \quad \text{если } 10 < d < d_{BP},$$

$$\theta = 22 \log_{10} d + 28 + 20 \log_{10} f_c, \quad \text{если } d_{BP} < d < 5000, \quad (2.30)$$

где h_s и h_u — высоты базовой станции (или точки доступа, access point, AP) и устройства (м); d и d_{BP} — расстояние до базовой станции (точки доступа) и расстояние (break point distance) точки разрыва (м); и f_c — центральная частота (ГГц). Затем, учитывая мощность шума, мы можем рассчитать SNR на единицу мощности как $\gamma_i = \frac{1/\theta}{N_0}$.

Мы исследуем исходную задачу оптимизации с учетом реалистичных параметров, частично заимствованных из [101]. Все соответствующие параметры приведены в Таблице 2.3.

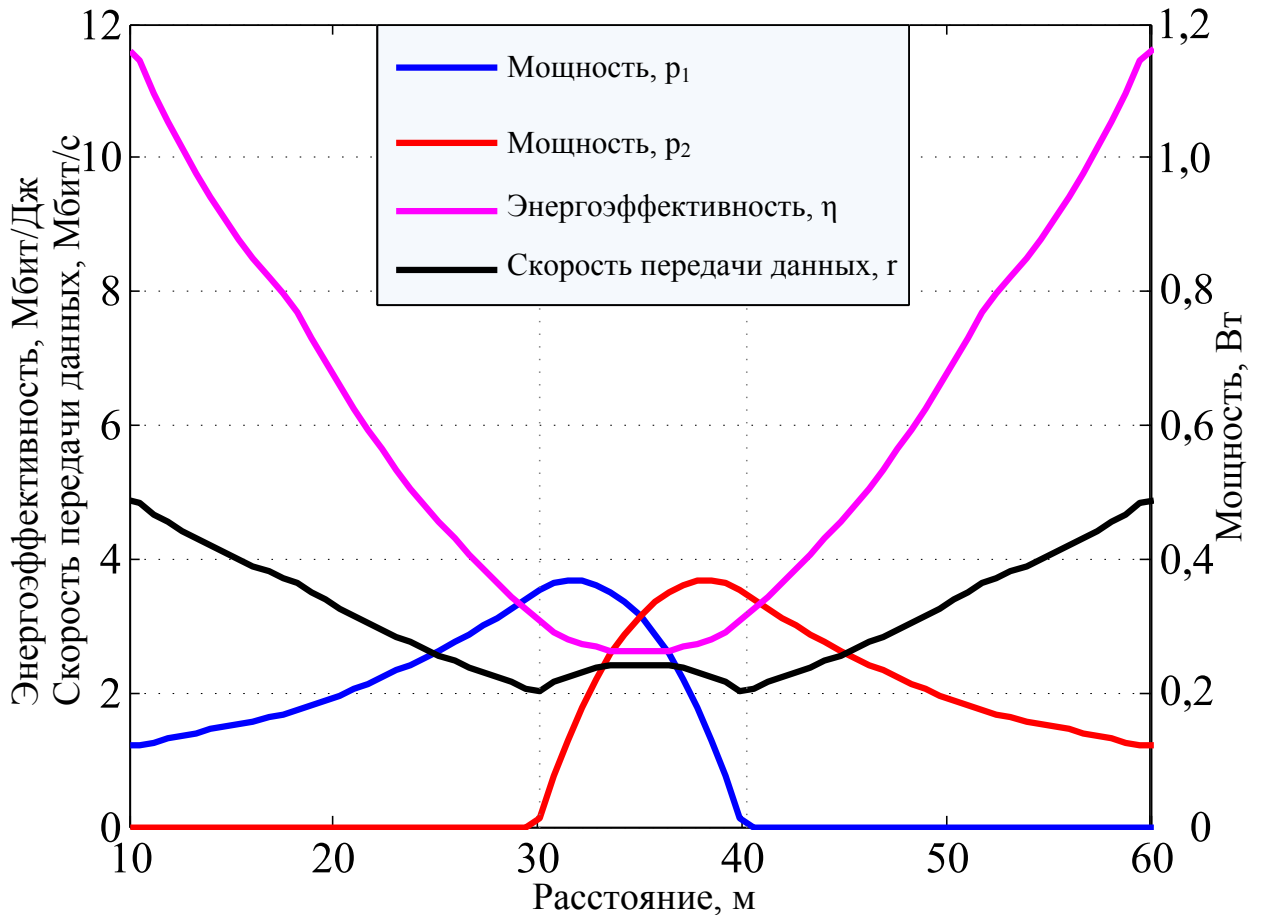


Рисунок 2.10 — Производительность системы для $K = 2$, ограничений нет

В целях иллюстрации мы сначала предоставляем результаты для упрощенного сценария без ограничений максимальной мощности и минимальной скорости передачи. На рисунке 2.10 показана зависимость достижимой скорости передачи данных, мощности передачи и энергоэффективности от текущей координаты устройства пользователя x . Пользовательское устройство начинает с использования единственной RAT и движется от одной базовой станции к другой (слева направо). Мы ясно видим, что наш предлагаемый алгоритм контроля мощности увеличивает мощность передачи, чтобы компенсировать растущее затухание в канале (согласно модели в (2.30)). Начиная с определенной точки ($x = 31$ м на рисунке), пользовательское устройство подключает вторую RAT путем выделения некоторой мощности передачи в соответствующем канале. Следовательно, это позволяет снизить мощность на первом канале, сохраняя при этом максимальную эффективность использования энергии.

Параметры системы

Описание параметра	LTE	WiFi
N_0 Уровень шума	−90 дБ	−90 дБ
f_c Частота	2.0 ГГц	2.4 ГГц
h_s Высота антенны базовой станции	25 м	10 м
h_u Высота антенны пользователя	1.5 м	1.5 м
Число пользователей (на сектор/точку доступа)	10	10

В средней точке уровни мощности передачи по двум каналам совпадают в силу симметрии этого примера. Как правило, скорость передачи данных пользователя снижается по мере ухудшения качества канала. Однако мы также наблюдаем интересный эффект некоторого увеличения скорости передачи данных из-за совместной работы двух RAT. Этот рост скорости передачи данных достигается за счет дополнительной мощности передачи. Следовательно, в конечном итоге это не влияет на оптимальный уровень энергоэффективности, который монотонно уменьшается, пока пользовательское устройство перемещается от начала до середины, а затем симметрично увеличивается.

Симметричный случай. Три технологии радиодоступа

Здесь мы проиллюстрируем поведение предлагаемой процедуры оптимального контроля на плоскости, снова используя рассмотренную модель распространения сигнала (2.30). Теперь мы предполагаем сосуществование трех RAT ($K = 3$), доступных для использования в любой момент времени. В результате рассматриваемая система содержит три аналогичные (например, 3GPP LTE) базовые станции по разным каналам: в точках $x_1 = (0, R)$ и $x_2 = (R, 0)$ и в центре координаты $x_3 = (0, 0)$. Пользовательскому устройству разрешено перемещаться по плоскости между всеми базовыми станциями (см. рисунок 2.11), а управление мощностью устройства определяется алгоритмом 2.

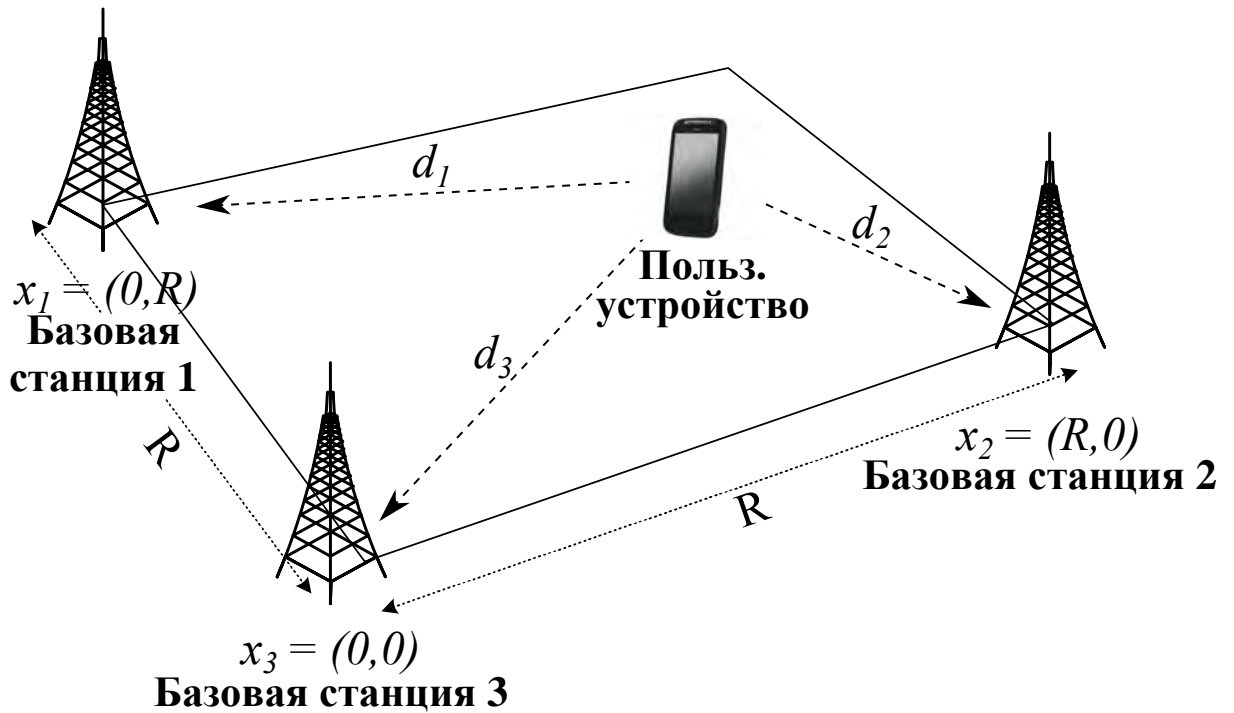


Рисунок 2.11 — Геометрия сети для случая трех RAT

На рисунке 2.12 мы выделяем области совместной работы RAT, которые определяются оптимальными уровнями мощности, установленными предлагаемым алгоритмом. Мы наблюдаем несколько типов регионов: три области, в которых пользовательское устройство передает по одному каналу, три области между парами базовых станций (с несколькими радиоинтерфейсами, активированными для поддержки целевой скорости передачи данных) и одна область поддержки, в которой задействованы все три радиоинтерфейса.

Рисунки 2.13, 2.14 и 2.15 демонстрируют зависимость достижимой скорости передачи данных, мощности передачи и энергоэффективности от координат пользовательского устройства. Ситуация, как правило, аналогична вышеупомянутому двумерному случаю ($K = 2$): скорость передачи данных уменьшается, когда пользовательское устройство удаляется от базовой станции. Мы также можем четко видеть эффект повышенной скорости передачи данных в областях совместной работы нескольких RAT, что связано с уменьшением передаваемой мощности, обеспечиваемой нашим алгоритмом оптимального контроля.

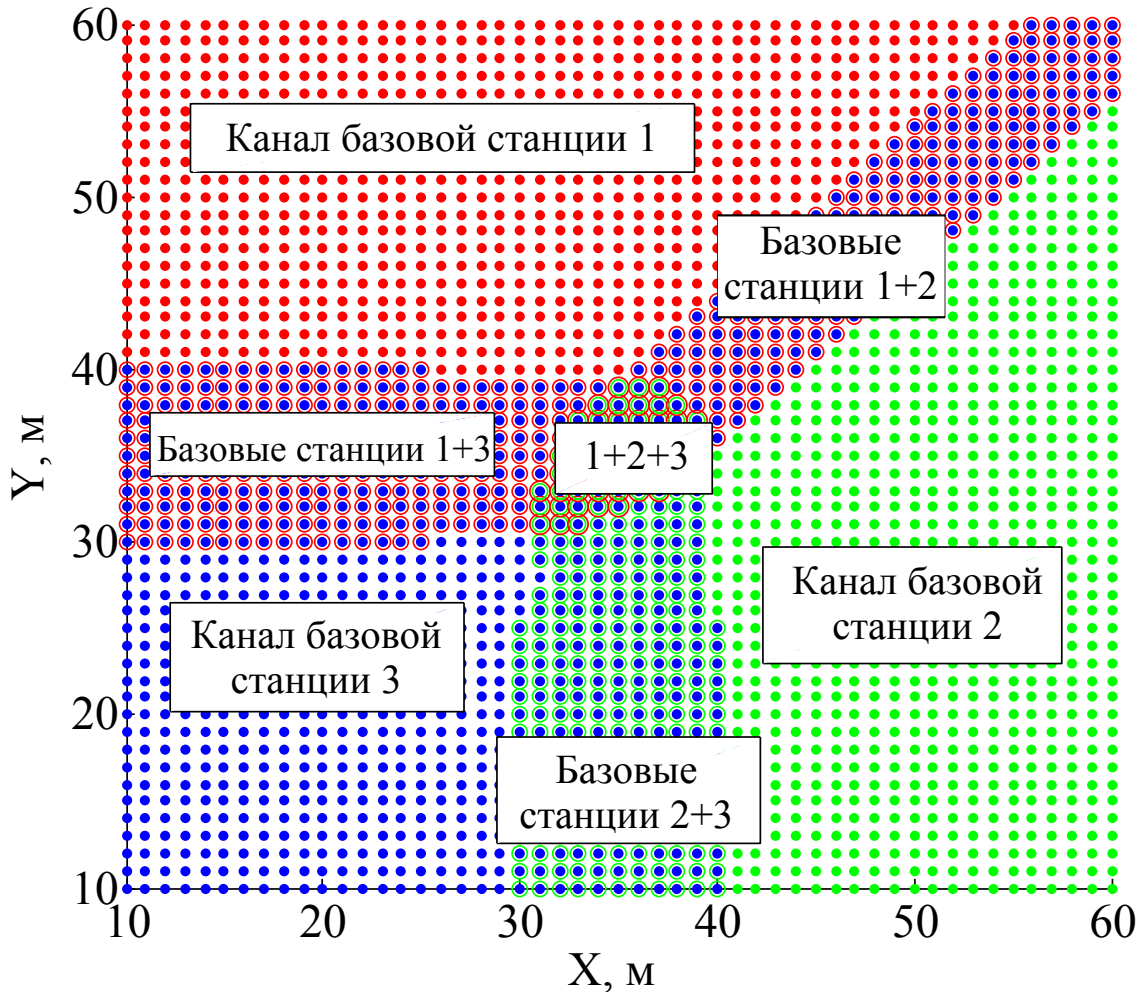
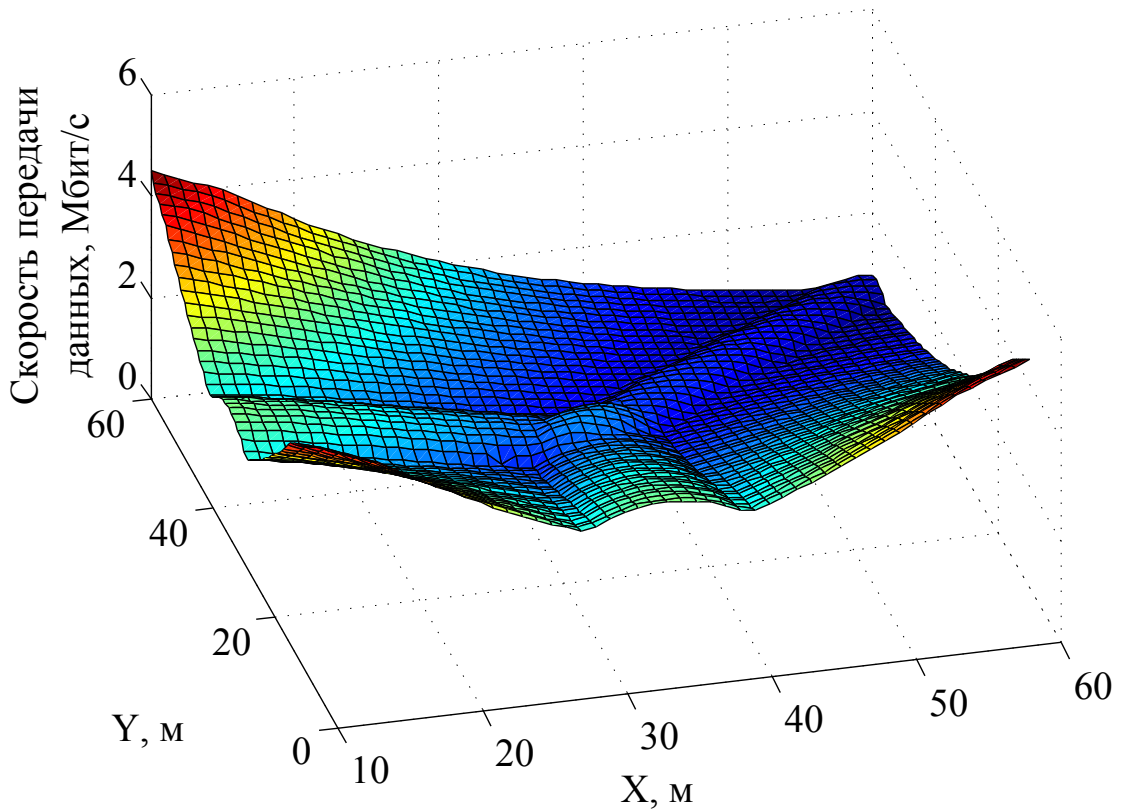


Рисунок 2.12 — Области передачи данных по разным каналам

Асимметричный случай. Две технологии радиодоступа

В дальнейшем мы рассмотрим две различные радиотехнологии: 3GPP LTE и WiFi, предполагая, что конкретный пользователь перемещается между базовой станцией LTE и точкой доступа WiFi. Для простоты мы предполагаем, что работа WiFi (согласно технологии IEEE 802.11-2012) контролируется функцией координации точек (PCF) и, следовательно, round-robin планировщиком (например, при коммерческом развертывании WiFi на светофорах или на уличных столбах).

Рисунок 2.16 описывает случай, когда действуют ограничения на максимальную мощность и минимальную скорость передачи данных. В области до точки (A) отправляет данные только интерфейс LTE, а мощность передачи увеличивается из-за растущего затухания; от точки (A) до точки

Рисунок 2.13 — Скорость передачи для $K = 3$

Ⓔ мощность передачи WiFi быстро растет до максимального уровня (из-за того, как работает WiFi PHY). Оба радиointерфейса поддерживают максимальный уровень мощности до точки Ⓒ, после чего использование LTE снижается из-за низкого качества канала. В точке Ⓓ скорость передачи данных приближается к минимальному уровню, и, следовательно, уменьшение мощности передачи LTE несколько замедляется, чтобы поддерживать целевую скорость передачи. Когда WiFi становится более эффективным в точке Ⓔ, LTE больше не используется. Далее увеличиваются и энергоэффективность, и общая скорость передачи данных в силу улучшения WiFi соединения до точки Ⓕ, и, следовательно, наша схема достигает максимального уровня скорости передачи данных, когда дальнейшее увеличение SNR не приводит к увеличению скорости передачи.

Кроме того, для сравнения предложенного оптимального алгоритма управления мощностью с возможными (более простыми) альтернативами, мы рассматриваем две упрощенные политики управления мощностью, где мощность передачи фиксирована. В частности, мы устанавливаем выделенный уровень мощности равным максимально допустимому значению.

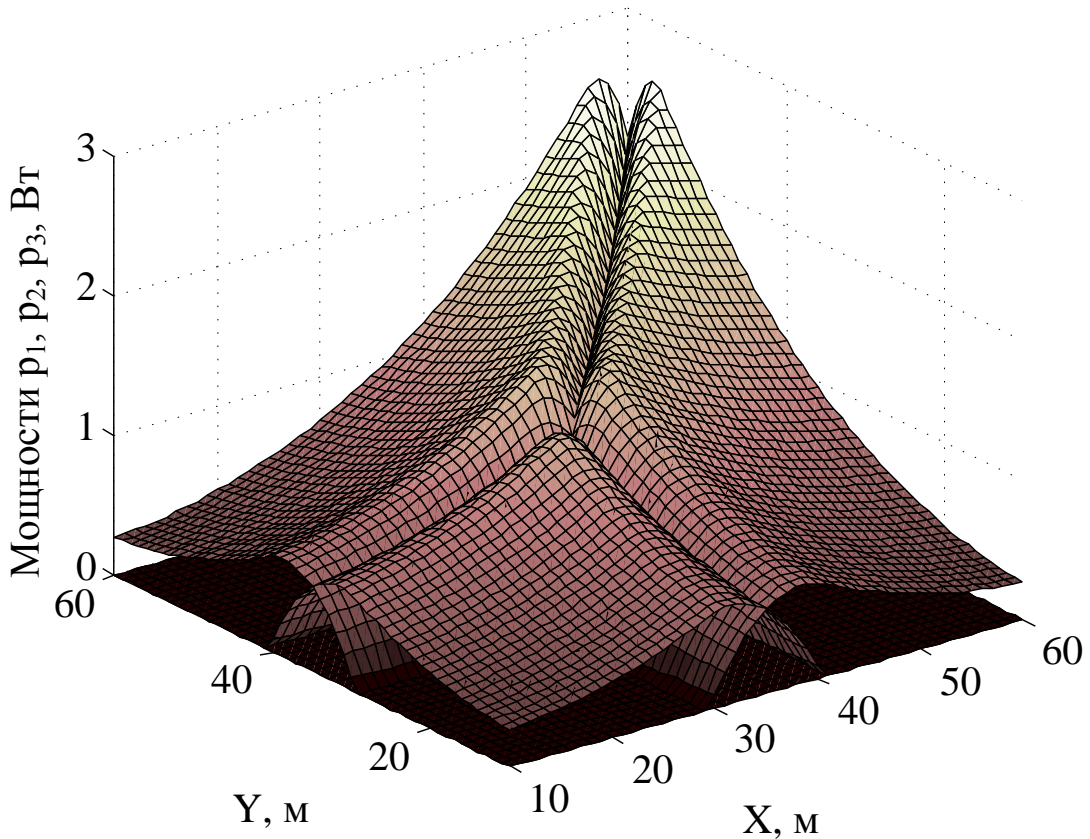


Рисунок 2.14 — Мощность передачи $K = 3$

Кроме того, в качестве более сложной стратегии управления мощностью ниже мы вводим интуитивную эвристическую политику передачи.

- *Простая политика 1.* Пользовательское устройство передает на один канал с максимальным качеством, выделяя ему фиксированную мощность. Мы определяем лучший канал, выбирая самое высокое значение $\gamma_i w_i$, которое учитывает как SNR, так и пропускную способность канала.
- *Простая политика 2.* Пользовательское устройство передает по всем каналам одновременно путем задания фиксированной мощности в каждом канале.
- *Интуитивно понятная эвристическая политика.* Пользовательское устройство следует стратегии эвристического распределения мощности, используя сначала канал наилучшего качества, а затем, при необходимости, используя остальную требуемую скорость передачи (вверх к целевому значению r_0) на других каналах в порядке уменьшения качества канала (подробнее см. Алгоритм 3). Доступ-

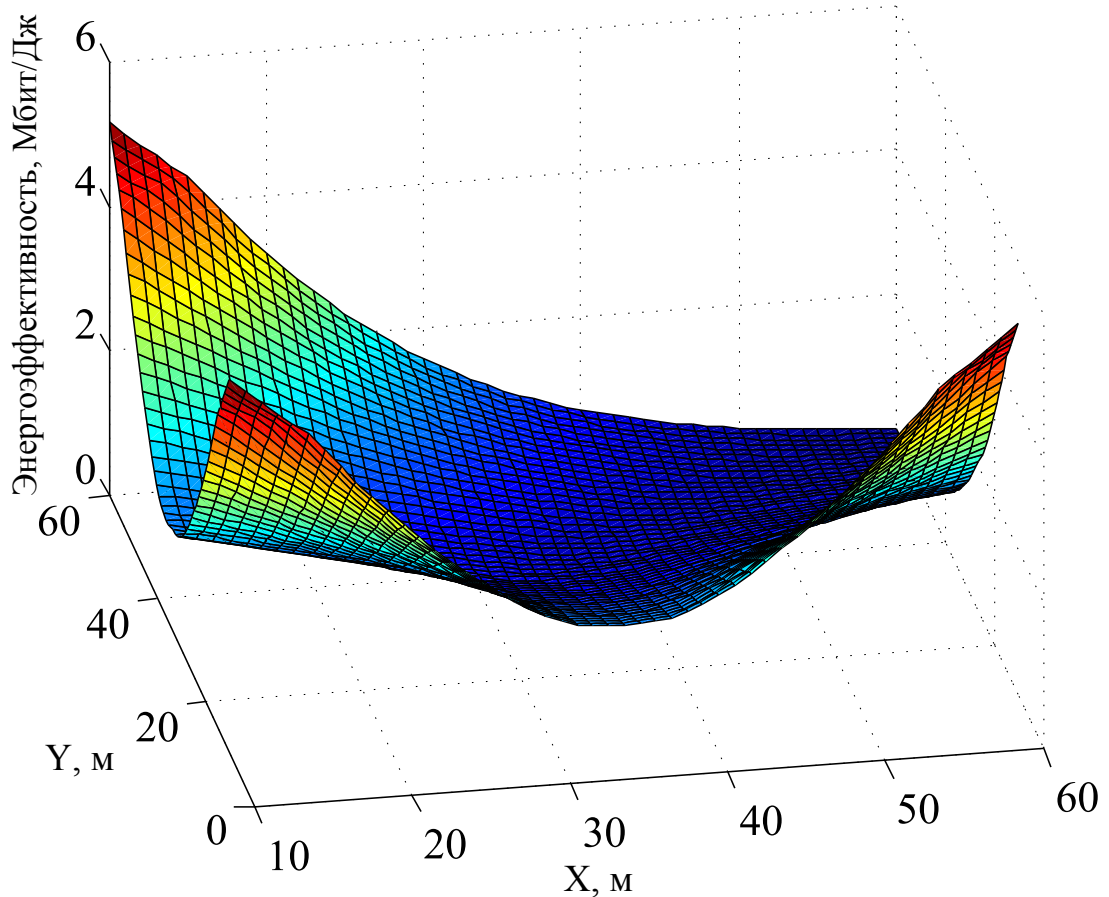


Рисунок 2.15 — Энергоэффективность для $K = 3$

ные каналы могут сравниваться по тому же критерию, что и выше, то есть предпочтительным будет канал с более высоким значением $\gamma_i w_i$.

На рисунке 2.17 эффективность энергопотребления пользовательского устройства и скорость передачи данных, достигаемая с помощью предлагаемого оптимального управления мощностью, сравниваются с производительностью трех более простых стратегий управления мощностью, описанных выше. Скорость передачи данных в первой стратегии является самой низкой в средней области, где качество обоих каналов невелико, что приводит к нарушению QoS, когда целевая битовая скорость не может быть удовлетворена даже при максимальной мощности передачи. Вторая стратегия обычно обеспечивает более высокие скорости передачи данных за счет чрезмерного потребления энергии. Наш оптимальный подход к энергоэффективности позволяет повысить скорость пользовательских данных практически до уровня эффективности второй стратегии, гарантируя целе-

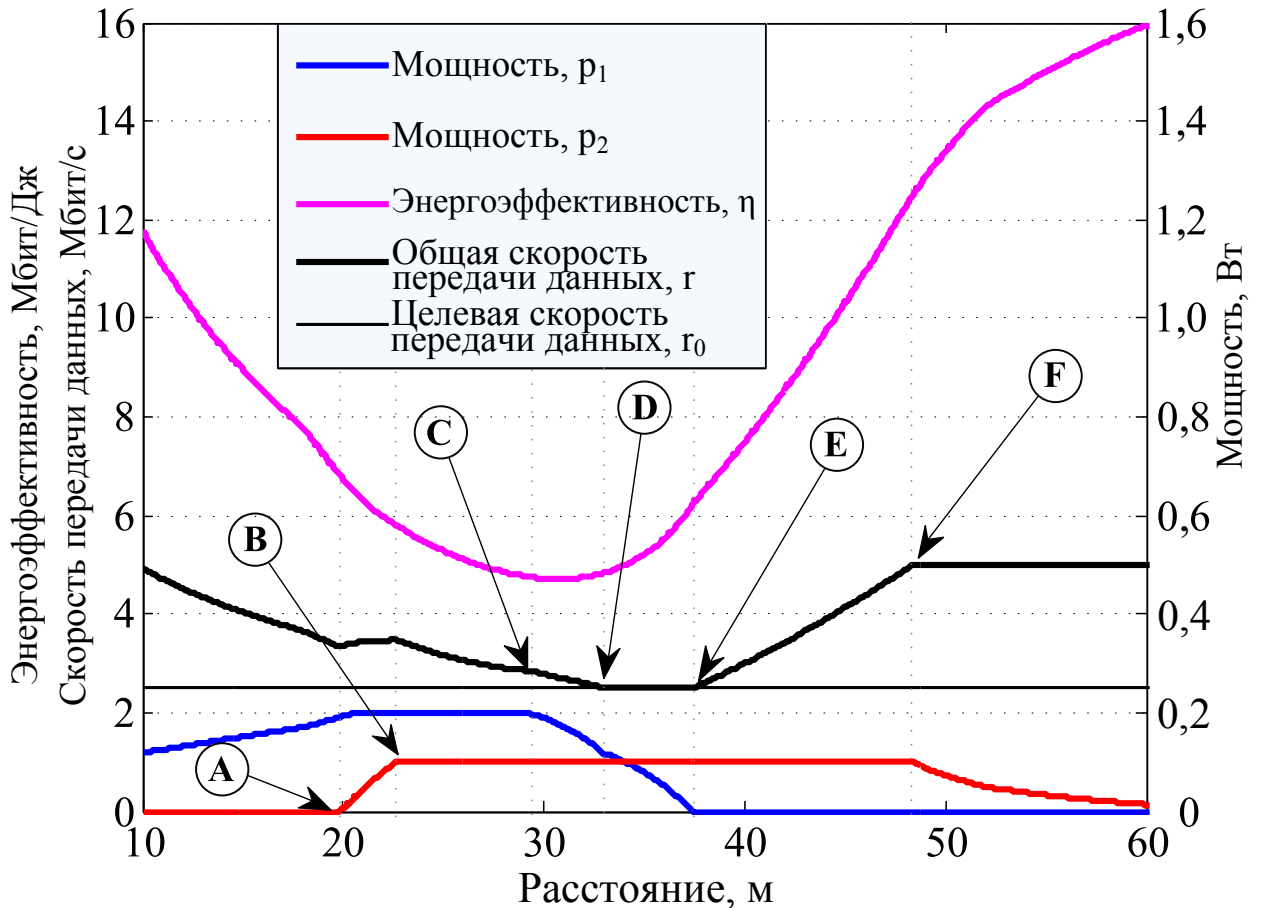


Рисунок 2.16 — Производительность для $K = 2$ с ограничениями по максимальной мощности и минимальной скорости передачи бит

вую скорость передачи данных r_0 . Сокращая ненужное потребление энергии, наша схема в явном виде приводит к максимальной эффективности использования энергии. Самое интересное, что третья (интуитивная эвристическая) политика обеспечивает почти оптимальную эффективность использования энергии. Однако на определенных расстояниях она не оптимальна с точки зрения энергоэффективности.

Наконец, на рисунке 2.18 мы оцениваем относительное снижение энергоэффективности при использовании одной из трех альтернативных политик. Мы заключаем, что всякий раз, когда поддерживается целевая битовая скорость, предлагаемая оптимальная схема обеспечивает гораздо более высокую энергетическую эффективность, чем две схемы примитивного распределения мощности. Третья (эвристическая) стратегия приближается к нашему оптимальному решению почти на всем пути, но несмотря на это оставляет место для улучшения.

```

 $r_{rest} \leftarrow r_0$ 
 $I = \{i\}_{i=1}^K$  while  $r_{rest} > 0$  or  $K > 0$  do
    end
    Найти канал с наилучшим качеством  $i_0 \leftarrow \max_{i \in I} \gamma_i$  if
 $r_{rest} > w_{i_0} \log(1 + p_{i_0}^{\max} \gamma_{i_0})$  then
        end
         $p_{i_0} \leftarrow p_{i_0}^{\max}$ 
 $r_{i_0} \leftarrow w_{i_0} \log(1 + p_{i_0}^{\max} \gamma_{i_0})$  else
        end
         $r_{i_0} \leftarrow r_{rest}$ 
 $p_{i_0} \leftarrow 1/\gamma_{i_0} (e^{r_{i_0}/w_{i_0}} - 1)$ 
 $r_{rest} \leftarrow r_{rest} - r_{i_0}$ 
 $K \leftarrow K - 1$ 
 $I \leftarrow I \setminus \{i_0\}$ 

```

Алгоритм 3: Эвристическая стратегия распределения ресурсов

2.4.5 Заключение

В этом разделе рассматривается энергоэффективное управление мощностью устройства, использующего несколько технологий радиодоступа в гетерогенных беспроводных сетях. Аналитически решается задача максимизации энергоэффективности устройства для произвольного количества RAT при нескольких практических ограничениях, таких как минимальная целевая битовая скорость и максимальная допустимая мощность передачи. Результаты иллюстрируются числовыми примерами для двух и трех RAT и подтверждают, что предлагаемая схема управления мощностью снижает расходы на энергию мобильного устройства и в то же время поддерживает требуемый уровень скорости пользовательских данных.

В отличие от существующих подходов, предлагаемое решение устанавливает области, в которых два или более RAT работают совместно, что приводит к более энергоэффективной работе по сравнению с другими методами управления мощностью. Представленные результаты показывают, что предлагаемая стратегия управления мощностью может стать привле-

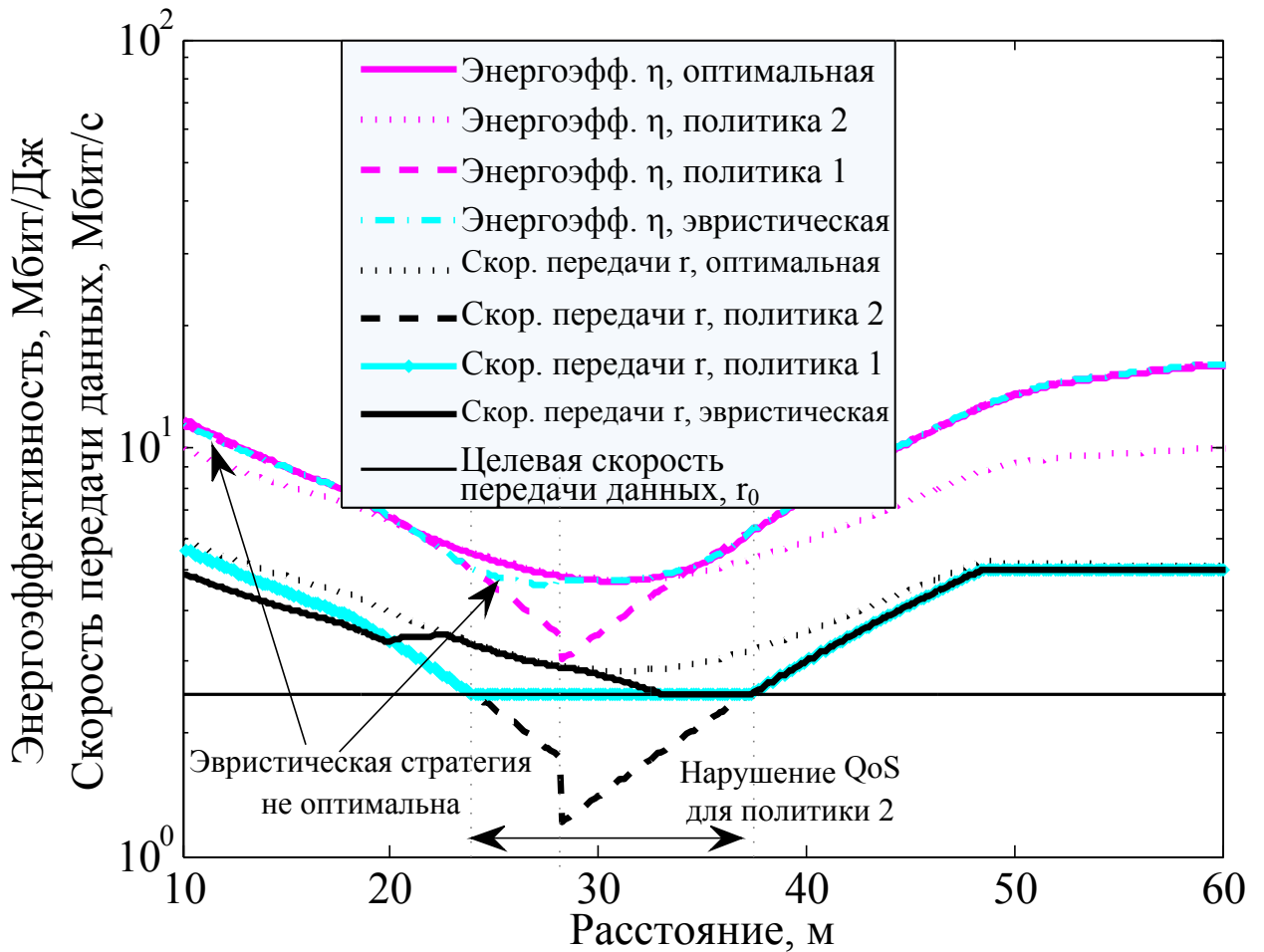


Рисунок 2.17 — Сравнение производительности политик управления энергопотреблением: оптимальные и более простые подходы

кательным выбором для будущих интегрированных беспроводных систем 5G и внести вклад в соответствующие исследования. Выбор более адекватных эвристических схем распределения мощности, которые позволяли бы лучше приблизиться к оптимальной производительности, а также учет динамических моделей трафика, могут служить дальнейшими направлениями развития такого класса задач.

2.5 Выводы

1. Анализ особенностей перехода к сетям связи пятого поколения показал, что эти сети количественно и качественно будут принципи-

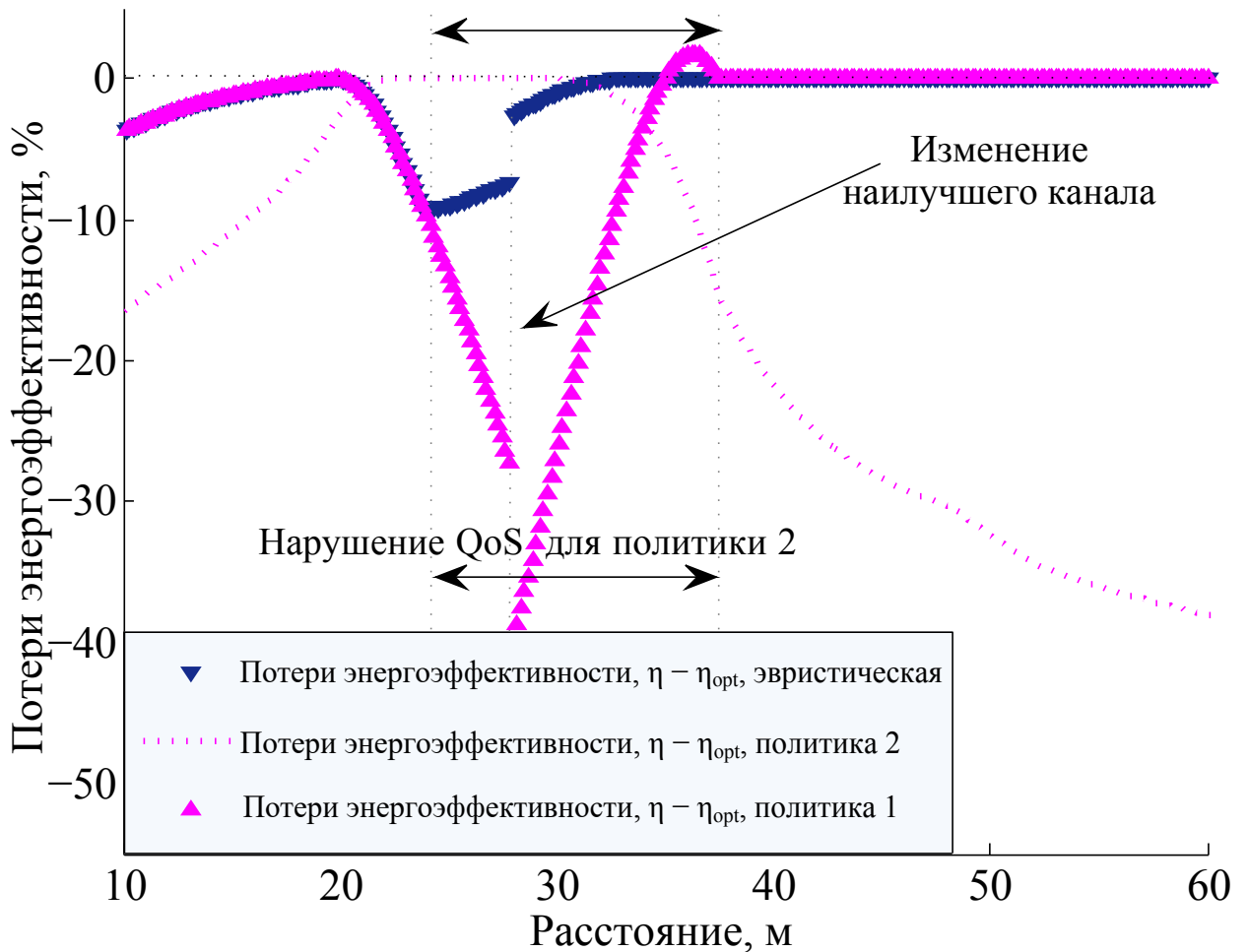


Рисунок 2.18 — Относительная потеря энергоэффективности: оптимальная схема управления мощностью и более простые подходы

ально отличаться от всех существующих сетей связи как по плотности размещения устройств сети, так и по требуемым задержкам. При этом сети связи пятого поколения характеризуются как сверхплотные сети (до 1 млн устройств на кв.км) и сети с ультра малыми задержками. Это требует разработки новой методологии распределения ресурсов в таких сетях и новых технологий взаимодействия устройств на относительно небольших расстояниях.

2. Предложена новая классификация гетерогенных беспроводных сетей связи, подразделяющая все гетерогенные беспроводные сети на три типа.
3. Предложен новый шлюз входа в сеть доступа (AAGW), который в отличие от предыдущих архитектурных решений не требует обеспечения доступа ни в операторскую магистральную сеть, ни в ин-

тернет и отражает функциональность решений 3GPP во WLAN радио технологиях RAT (например, WiFi) и наоборот. Это, в свою очередь, позволяет пользовательскому оборудованию максимально использовать доступные радио интерфейсы без необходимости применения отдельных решений для мобильности для сетей доступа не-3GPP.

4. Разработан алгоритм выбора мощности передачи мобильного устройства в условиях наличия двух и более радиотехнологий, способных функционировать одновременно. Показано, что предлагаемый алгоритм максимизирует энергоэффективность мобильного устройства при обеспечении минимально требуемой скорости передачи и ограничении сверху на мощность передачи.

Раздел 3. Разработка и исследование моделей и методов при использовании кооперативных коммуникаций для распределения ресурсов в сетях M2M

Для современных сотовых сетей связи одной из перспективных концепций с точки зрения увеличения производительности сети (в частности, пропускной способности и энергоэффективности) являются кооперативные коммуникации. Мобильный пользователь, чей беспроводной канал не гарантирует надежную передачу в силу низкого качества соединения, может дополнительно использовать возможности близлежащих устройств, которые, в свою очередь, готовы передавать данные от его имени. В этом разделе рассмотрена трехузловая система *клиентской ретрансляции* с ограничением на число повторных передач [102]. Принимая во внимание подобное ограничение, мы получаем аналитические оценки для значений пропускной способности, средней задержки, расхода энергии и энергоэффективности.

Далее, рассмотрим беспроводную сотовую сеть с поддержкой приложений межмашинные коммуникации M2M (Machine-to-Machine). В соответствии с текущими предложениями стандарта IEEE 802.16p, беспроводное M2M устройство может выступать в качестве агрегирующей точки и передавать пакеты от имени других M2M устройств, у которых или отсутствует интерфейс с сотовой сетью, или связь с сетью недостаточно хорошая. В этом разделе мы применяем схему клиентской ретрансляции для увеличения надежности линий передачи и энергетической эффективности устройств с плохими условиями связи. Производительность предложенной схемы оценивается посредством анализа и имитационного моделирования нескольких метрик и позволяет учесть пропускную способность, задержку и потребление энергии.

В конечном счете, набор доступных для ретрансляций устройств может быть объединен в некоторое «облако ретрансляции», которое может осуществлять кооперативную повторную передачу данных пользователей с более низким качеством канала. Ключевая идея облака ретрансляции состоит в обеспечении распределенного гибкого управления кооперативной

передачей посредством самих беспроводных пользователей. В противовес централизованному управлению это предложение минимизирует ресурсы на сигнализацию, вовлеченные в процесс кооперации, и обеспечивает более простое внедрение. Далее в этом разделе разрабатывается модель, ориентированная на источник пакетов данных, и исследуется производительность облака ретрансляции с точки зрения описанных выше метрик производительности – пропускной способности, задержки пакетов и энергоэффективности.

3.1 Введение и предпосылки

3.1.1 Кооперативные коммуникации и ретрансляция

На ослабление негативных последствий замираний в многолучевых каналах связи (с целью увеличения надежности беспроводной передачи) направлены разнообразные приемы разнесения сигналов (*diversity techniques*). В частности, одним из наиболее усовершенствованных методов является *пространственное разнесение передатчиков*, которое позволяет использовать две или более передающие антенны для улучшения качества линии связи [103]. Однако, мобильные терминалы с несколькими антеннами могут оказаться экономически невыгодными вследствие большего размера, а также технических ограничений оборудования. Исходя из сказанного, в целях обеспечения пространственного разнесения для мобильных устройств со стандартными антеннами была предложена концепция *кооперативных коммуникаций*, предполагающая поддержку так называемой *кооперативной передачи*.

Базовая идея кооперативных коммуникаций впервые была предложена в фундаментальной работе [104], где простая трехузловая модель системы, включающая в себя *передатчик*, *приемник* и *ретранслятор*, была подробно исследована. Анализ производительности канала ретрансляции

был проведен позже в работе [105]. Эти первые исследования, посвященные простому трехузловому случаю, предложили ряд стратегий ретрансляции. Также были установлены достижимые регионы и верхние границы производительности канала ретрансляции, который мы теперь называем классическим.

В последние годы появился новый всплеск интереса к этому вопросу, и вырос спрос на соответствующее расширение исследований по кооперативным коммуникациям. Например, работа [106] адресована некоторым новым информационно-теоретическим аспектам ретрансляции, позволяющим более подробно изучить особенности кооперативных коммуникаций.

Более детально кооперативная передача была рассмотрена в [107] как относительно новый класс пространственно разнесенной передачи, обеспечивающий ретрансляцию и кооперацию. В [108] авторы предложили стратегию эффективной кооперации и также исследовали внедрение этой концепции совместно с рядом практических вопросов, связанных с имплементацией. Хороший обзор литературы, посвященной кооперативным коммуникациям, можно найти в [109].

Ряд существующих работ адресован более сложным моделям с несколькими беспроводными пользователями (клиентами), которые могут быть выбраны в качестве ретрансляторов. Проблема выбора ретранслятора (данные источника информации могут теоретически иметь более одного потенциального ретранслятора) были исследованы в [110], при этом предполагалась доступность централизованного контроллера кооперации. Все вышеизложенное переводит концепцию кооперации в сферу исследований беспроводных сотовых сетей с *базовой станцией*, управляющей активностью своих клиентов.

Далее, в [111] для снижения потерь мультиплексирования в сетях ретрансляции были предложены несколько эффективных протоколов для выбора ретранслятора, требующих дополнительной обратной связи. Большинство текущих работ по изучению кооперации связаны с перспективой использования централизованного управления, что увеличивает число протоколов сигнализации, вовлеченных в процесс кооперации, и в результате усложняет процесс внедрения для существующих систем. В противовес из-

ложенному, в диссертации предложены более практичные сценарии с гибким и распределенным управлением кооперацией, использующие возможности самих клиентов беспроводной сети.

По мере разработки телекоммуникационных стандартов [112], [103] беспроводные сети используют все более широкие полосы пропускания. Однако, в случае, когда большое количество пользователей делит между собой ограниченный спектральный ресурс [113], задача эффективного использования ресурсов приобретает первостепенную важность с точки зрения современных беспроводных сетей. В связи с растущей мобильностью пользователей, в более поздних исследованиях происходит переход от оптимизации пропускной способности [73] к улучшению энергоэффективности на всех уровнях беспроводной системы: от их архитектуры [75] до коммуникационных протоколов [114].

В настоящее время задача повышения энергоэффективности получает все больше внимания со стороны исследовательского сообщества [71] и, в частности, делает упор на совместном рассмотрении MAC и PHY уровней [78], [115] и [116]. Кроме того, подходы к межуровневой кооперации получили широкое признание в научной литературе [111], [117]. Такие подходы используют вариативность информации о состоянии канала (Channel State Information, CSI), что позволяет добиться увеличения производительности, и открывают перспективное направление исследований.

В то время как все больше и больше пользователей делят между собой ограниченные беспроводные ресурсы, а в сотовых сетях [112], [103] происходит постепенное смещение в сторону более агрессивных сценариев переиспользования частот (frequency reuse), интерференция становится одним из важнейших ограничивающих факторов, сдерживающих рост производительности сети. Беспроводная передача данных одного пользователя влияет на процесс передачи ряда других абонентов и постепенно приводит к снижению энергоэффективности системы в целом. Однако, пользователи могут увеличить свою энергоэффективность, используя кооперацию [118], позволяющую снизить уровень мощности и, следовательно, интерференции в сети. Такая сфера пространственного управления ресурсами становится все более важной с точки зрения увеличения производительности пользо-

вателей, находящихся на границе соты и имеющих более низкое качество беспроводного канала [119].

В то время как технология ретрансляции для базовых станций была стандартизована комитетом IEEE в стандарте 802.16j [120], кооперации между самими беспроводными пользователями (client relay), несмотря на важность ее изучения, было уделено весьма ограниченное внимание. Однако, в связи с тем, что расход энергии для гарантированной надежной передачи данных быстро увеличивается с ростом расстояния [121], предпочтительнее передавать данные через промежуточные, более короткие с точки зрения расстояния, соединения. Таким образом, клиентская ретрансляция представляет собой перспективную концепцию, которая позволит повысить производительность современных беспроводных сотовых сетей.

Исследование производительности клиентского ретранслятора требует создания адекватной модели беспроводной системы [122]. Поскольку комплексное исследование является довольно сложным, разделим эту задачу на две. Во-первых, будем использовать аналитические методы для получения адекватной модели системы. Однако, многоплановое взаимодействие между очередями во время кооперации пользователей препятствует прямому использованию известных результатов из теории массового обслуживания. Поэтому, для того, чтобы исследовать пропускную способность, энергозатраты, энергоэффективность и среднюю задержку пакетов в системе, должна быть сформулирована расширенная модель. Средняя задержка пакетов находится в центре внимания этого раздела, которая является расширением прежних работ по теме диссертации [123] и [119] на случай более реалистичного сценария с потерями, когда сброс пакетов возможен из-за большого количества неудачных повторных передач.

Кроме того, для учета таких факторов, влияющих на производительность клиентских ретрансляторов [124], таких как реалистичный трафик, заданные параметры качества обслуживания, факторы, влияющие на качество беспроводного канала и т.д., может быть использовано имитационное моделирование. Имитационное моделирование позволяет вникнуть в суть многих интересных процессов в кооперативных беспроводных сетях и получить важные выводы относительно предпочтительной области

применения технологии клиентской ретрансляции. [125]. Для будущих беспроводных сотовых сетей, таких как сети стандартов IEEE 802.16m [112] и LTE-Advanced [103], снижение мощности передачи при помощи использования ретрансляции также позволит снизить общую интерференцию между соседними сотами. Следовательно, использование методов клиентской ретрансляции может оказаться предпочтительным для сценариев, где интерференция играет основную роль. Кроме того, подобное взаимодействие пользователей представляет собой перспективный метод увеличения производительности пользователей, находящихся на границе соты.

3.1.2 Межмашинные коммуникации

В соответствии с [126], коммуникации M2M могут быть определены как обмен информацией между некоторой абонентской станцией SS (Subscriber Station) и сервером в сети (core network) через базовую станцию BS (Base Station), который осуществляется без какого-либо участия человека. Индустриальные отчеты прогнозируют появление в сети миллионов устройств и отмечают высокий потенциал этого рынка с ожидаемой прибылью в 300 миллиардов долларов США [127]. В связи с этим, несколько сотовых стандартов фокусируются на разработке необходимых расширений беспроводных интерфейсов с целью поддержки M2M коммуникаций.

Например, предложения по стандарту IEEE 802.16p [126] описывают расширение возможностей IEEE 802.16m для поддержки приложений M2M. 3GPP LTE также имеет несколько рабочих пунктов по M2M, в первую очередь, имеющих отношение к управлению перегрузками [128]. Отчеты по стандарту IEEE 802.16 M2M покрывают несколько сценариев M2M в таких широких категориях, как измерения, безопасность и наблюдение, удаленное обслуживание и управление. С деталями этих и других сценариев M2M можно ознакомиться в [126].

Ключевым сценарием M2M в настоящее время является использование умных счетчиков (англ. Smart metering), которые автономно информируют энергосистему об *использовании* и *сигнализации* с целью снижения эксплуатационных затрат и регулирования использования коммунальных услуг с помощью ценообразования, основанного на информации о загрузке энергосистемы. Мы ожидаем, что беспроводные технологии, такие как IEEE 802.16 и LTE, будут играть ключевую роль в поддержке подобных приложений.

В диссертации исследуется типовой для M2M сценарий Smart Metering в контексте беспроводной сотовой сети IEEE 802.16. Исследования сконцентрированы на увеличении производительности устройств M2M, которые могут находиться на краю зоны обслуживания и использовать канал связи недостаточного качества. Кроме того, в диссертации предложена простая и реализуемая на практике схема клиентской ретрансляции, направленная на увеличение производительность соединения. Также формулируется аналитический подход, позволяющий оценивать и прогнозировать производительность предложенной схемы. В частности, анализируются такие ключевые метрики производительности, как пропускная способность, задержка пакетов и энергетическая эффективность [71].

Наш анализ показывает, что задержка и энергопотребление устройств M2M, находящихся на границе соты, могут быть значительно понижены даже в ситуации всплеска активности большого числа устройств, что вызывает перегрузки в сети. Подобные всплески возможны тогда, когда большое число сенсоров осуществляет синхронные попытки соединения с сетью, например, в момент восстановления системы после отключения подачи электроэнергии. Предложенная схема клиентской ретрансляции помогает также уберечь другие сотовые устройства в сети от негативных последствий неконтролируемых попыток доступа со стороны M2M устройств.

Несмотря на то, что в диссертации исследуется механизм клиентской ретрансляции в контексте сценария Smart Metering, эта концепция применима и для других приложений, таких как, например, работа сенсоров или оборудования для наблюдения, находящихся внутри зданий и, возможно, имеющих плохое соединение с сетью. Кроме того, основные выводы при-

менимы и к другим беспроводным беспроводным системам с поддержкой M2M, в частности, к сетям 3GPP LTE.

Далее продолжим наши исследования, рассматривая кооперативную систему с несколькими клиентами, которые могут выступать в качестве ретрансляторов. Источник данных может оппортунистически устанавливать соединение с некоторыми ретрансляторами в целях увеличения качества своей передачи по восходящему каналу. Агрегированный набор выбранных ретрансляторов в некоторый момент времени формирует так называемое *облако ретрансляций* (relay cloud). Таким образом, целью исследования в данной области является изучение преимуществ облачной ретрансляции, а также разработка и анализ алгоритмов, которые могли бы максимизировать эффект кооперативных коммуникаций.

3.2 Трехузловая система клиентской ретрансляции с ограничением повторных передач

3.2.1 Описание системы и предположения

Рассмотрим беспроводную сотовую сеть, дополненную возможностями клиентской ретрансляции в соответствии с базовой методологией, приведенной в [123], и расширим ее следующим образом. Сконцентрируемся на простейшей топологии (см. рисунок 3.1), включающей два передатчика и один приемник. Будем называть передающее устройство *A* *источником*. Источник генерирует собственные пакеты данных со средней интенсивностью поступлений λ_A . Далее для второго передающего устройства *R* введем термин *ретранслятор*. Ретранслятор генерирует собственные пакеты данных со средней интенсивностью поступлений λ_R . Кроме того, ретранслятор имеет возможность перехватывать пакеты, исходящие от источника, и сохранять их для дальнейшей передачи. Узел *B* является *базовой станцией*

и получает пакеты данных и от источника, и от ретранслятора. Ниже мы детализируем модель системы и описываем основные ее допущения.

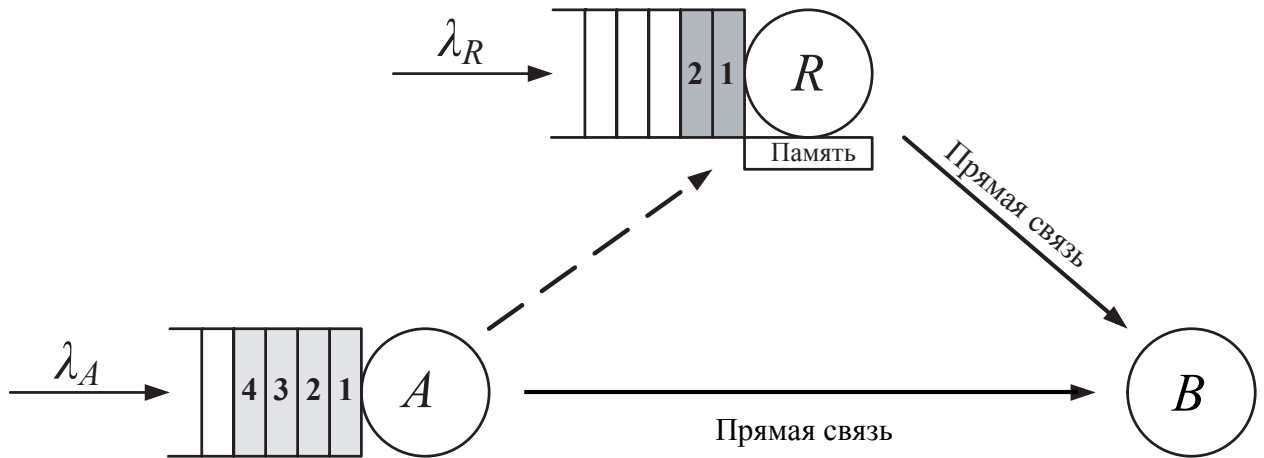


Рисунок 3.1 — Модель трехузловой системы клиентской ретрансляции

Предположение 1. Системное время разбивается на *слоты*. Все пакеты передаваемых данных имеют одинаковый размер, и передача каждого из них занимает ровно один слот.

Предположение 2. Число новых пакетов данных, появляющихся у источника или у ретранслятора во время текущего слота, независимо и одинаково распределено со средними значениями λ_A и λ_R соответственно. Для аналитического упрощения предполагается, что входящий поток новых пакетов является пуассоновским. Базовая станция не имеет своего трафика.

Предположение 3. И источник, и ретранслятор имеют неограниченный буфер для хранения своих собственных пакетов. Кроме того, ретранслятор имеет дополнительный буфер для хранения в точности одного пакета данных от источника с целью его последующей передачи. Ниже мы покажем, что для предлагаемой схемы клиентской ретрансляции, вполне достаточно одной ячейки памяти.

Предположение 4. Система связи централизована и управляется базовой станцией. На базовой станции используется *стохастическое* справедливое (fair) расписание с алгоритмом round-robin, таким образом, альтернативные узлы-источники имеют доступ к беспроводному каналу с равной вероятностью. В частности, если и источник, и ретранслятор имеют в

очереди пакеты данных, то слот следующий слот будет предоставлен одному из них с вероятностью 0,5. Непередающие узлы в течение этого слота бездействуют. Если любой из двух передатчиков имеет пустую очередь пакетов данных, то текущий слот предоставляется другому с вероятностью 1. Такой механизм обеспечивает эффективное использование времени системы. Также предположим, что информация о расписании незамедлительно доступна всем через выделенный канал и при этом не использует ресурсов системы [93].

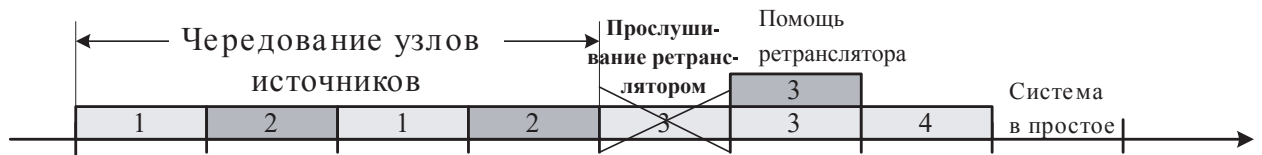


Рисунок 3.2 — Пример функционирования системы клиентской ретрансляции

Предположение 5. Канал связи является каналом с ошибками и основывается на модели многопакетного приема [117]. Передаваемый пакет данных может быть успешно получен с некоторой постоянной вероятностью, которая зависит от типа канала и того, сколько пользователей передают одновременно.

Определим следующие ненулевые вероятности успешной передачи

1. $p_{AB} = \Pr\{\text{пакет от } A \text{ получен на } B \mid \text{передает только } A\}$
2. $p_{RB} = \Pr\{\text{пакет от } R \text{ получен на } B \mid \text{передает только } R\}$
3. $p_{AR} = \Pr\{\text{пакет от } A \text{ получен на } R \mid \text{передает только } A\}$
4. $p_{CB} = \Pr\{\text{пакет от } A \text{ получен на } B \mid A \text{ и } R \text{ кооперируют}\}$

Ожидается, что $p_{AR} > p_{AB}$ и $p_{CB} > p_{AB}$.

Также предположим, что обратная информация об успешном/неуспешном приеме пакета базовой станцией незамедлительно доступен через выделенный канал и так же, как упоминалось выше, не использует ресурсов системы. Если пакет не принят успешно, то источник осуществляет его повторную передачу.

Максимальное число доступных попыток передачи ограничено числом N_A для пакета от A и N_R для пакета от R . Если пакет от A (или R)

сталкивается с N_A (N_R) последовательными неуспешными попытками передач, то он немедленно удаляется из очереди для соответствующего узла. Заметим, что когда $N_A \rightarrow \infty$ и $N_R \rightarrow \infty$ система становится системой без потерь [119].

Предположение 6. При первой попытке передачи пакета от источника ретранслятор будет пытаться перехватить пакет с вероятностью 1. В соответствии с предположением 5 эта попытка будет успешной с вероятностью p_{AR} . Если пакет успешно получен ретранслятором, то он сохраняется в памяти R , вытесняя ее предыдущее содержание. Также в соответствии с предположением 5, в текущем слоте с вероятностью p_{AB} этот пакет может быть также получен и базовой станцией. Если доставка пакета на базовую станцию неуспешна, то источник повторяет передачу этого же пакет в следующем доступном слоте.

Предположение 7. При любой попытке новой передачи пакета от источника к базовой станции ретранслятор будет функционировать одним из следующих способов. Если пакет, который передается, уже был внесен в память ретранслятора, то с вероятностью 1 ретранслятор передает его одновременно с источником (рисунок 3.2), увеличивая при этом производительность источника. В соответствии с предположением 5 этот пакет может быть успешно получен на базовой станции с вероятностью p_{CB} . В другом случае, если пакет не был перехвачен ретранслятором, ретранслятор будет продолжать прослушивать канал с вероятностью 1 (предположение 6).

Заметим, что в соответствии с предположениями 6 и 7 одна ячейка памяти ретранслятора для пакетов, получивших отказ при передаче, вполне достаточна для функционирования рассматриваемой схемы с клиентской ретрансляцией. Кроме того, источник не осведомлен о кооперативной помощи ретранслятора, а ретранслятор не передает источнику никаких подтверждений в отличие от подхода, изложенного в [117]. Это позволяет применить предложенную модель клиентской ретрансляции к современным стандартам сотовых связи [112], [103]. Отметим, что ретранслятор увеличивает пропускную способность источника, жертвуя своей собственной энергоэффективностью, поскольку на перехват пакетов и на одновремен-

ную передачу пакета вместе с источником затрачивается дополнительная энергия ретранслятора.

3.2.2 Оценка производительности

Основные положения Предлагаемый аналитический подход к оценке производительности рассматриваемой трехузловой схемы клиентской ретрансляции основывается на понятии времени обслуживания пакета. Время обслуживания пакета — это интервал времени с момента, когда меченый пакет готов для начала его обслуживания и до момента окончания обслуживания [129]. Более точно, в рассматриваемой системе длительность обслуживания рассматриваемого пакета начинается в тот момент, когда для своего источника этот пакет становится первым в очереди, и заканчивается, когда успешно завершается его передача. Заметим, что учитываются характеристики только успешно переданных пакетов.

Обозначим длительность обслуживания пакета, пришедшего на A , как $T_{AR}(\lambda_A, \lambda_R) \triangleq T_{AR}$, где \triangleq читается как равно по определению. Кроме того, введем понятие среднего времени обслуживания пакета на A , обозначив ее как $\tau_{AR}(\lambda_A, \lambda_R) \triangleq \tau_{AR} = E[T_{AR}]$, при этом если $\lambda_R = 0$, то время обслуживания в узле A равно $\tau_{AR}(\lambda_A, 0) \triangleq \tau_{A0}$. Точно также введем соответствующие характеристики T_{RA} , τ_{RA} и τ_{R0} для R .

Заметим, что для обеих систем с кооперацией (когда $p_{AR} > 0$) и без кооперации (когда $p_{AR} = 0$) выполняется следующее

$$\tau_{R0} = \frac{1 - (1 - p_{RB})^{N_R}}{p_{RB}},$$

в то время как следующее справедливо только для системы без кооперации

$$\tau_{A0} = \frac{1 - (1 - p_{AB})^{N_A}}{p_{AB}}.$$

Вывод τ_{A0} для системы с кооперацией является более сложной задачей и будет рассмотрен далее.

Обозначим число пакетов данных в очередях A и R в начале слота t как $Q_A^{(t)}$ и $Q_R^{(t)}$, соответственно. Будем исследовать схему клиентской ретрансляции в установившемся режиме, поэтому можно опустить верхний индекс t в переменных $Q_A^{(t)}$ и $Q_R^{(t)}$.

Наконец, обозначим коэффициент загрузки очереди [130] A как $\rho_{AR}(\lambda_A, \lambda_R) \triangleq \rho_{AR}$. По определению имеем $\rho_{AR} = \Pr\{Q_A \neq 0\} = \lambda_A \tau_{AR}$. В частности, для случая $\lambda_R = 0$, коэффициент загрузки очереди A , можно установить как $\rho_{AR}(\lambda_A, 0) \triangleq \rho_{A0} = \lambda_A \tau_{A0}$. Для системы без кооперации ρ_{A0} упрощается следующим образом

$$\rho_{A0} = \frac{\lambda_A}{p_{AB}} \cdot \left[1 - (1 - p_{AB})^{N_A} \right].$$

Коэффициент загрузки очереди ρ_{RA} и ρ_{R0} для R вводятся аналогично. Для обеих систем и с кооперацией, и без кооперации ρ_{R0} может быть упрощено далее

$$\rho_{R0} = \frac{\lambda_R}{p_{RB}} \cdot \left[1 - (1 - p_{RB})^{N_R} \right].$$

Системы без кооперации Предложенный в диссертации подход к оценке производительности рассматриваемой трехузловой системы клиентской ретрансляции применим для определения средней интенсивности обслуживания пакетов (иными словами, пропускной способности) на A и R . Например, пропускная способность A определяется следующим образом

$$\eta_A = \begin{cases} \lambda_A (1 - P_{LA}), & \text{без насыщения} \\ (1 - \lambda_R \tau_{R0}) \tau_{A0}^{-1}, & \text{насыщение для } A \\ (2\tau_{A0})^{-1}, & \text{насыщение для } A, R, \end{cases}$$

где P_{LA} вероятность потери пакета A , определенная как $P_{LA} = (1 - p_{AB})^{N_A}$. Пропускная способность R может быть получена подобным образом с учетом вероятности потерь пакетов для узла R $P_{LR} = (1 - p_{RB})^{N_R}$. Здесь условия насыщения определяются следующим образом

- Для A : $(\lambda_A \tau_{A0} + \lambda_R \tau_{R0} > 1)$ и $(\lambda_R \tau_{R0} < 0.5)$.

- Для R : $(\lambda_A \tau_{A0} + \lambda_R \tau_{R0} > 1)$ и $(\lambda_A \tau_{A0} < 0.5)$.
- Для A и R : $(\lambda_A \tau_{A0} > 0.5)$ и $(\lambda_R \tau_{R0} > 0.5)$.

Далее, изучим поведение источника A с точки зрения теории массового обслуживания. Рассмотрим систему, ассоциированную с A . Учитывая тот факт, что очереди для A и R взаимозависимы, формула Полячека-Хинчина [130] не может быть использована для определения *точного значения* средней длины очереди для A . Однако, можно использовать эту формулу для определения приближенного значения величины средней длины очереди A как

$$q_A \cong \lambda_A E[T_{AR}] + \frac{\lambda_A^2 E[T_{AR}^2]}{2(1 - \lambda_A E[T_{AR}])},$$

где $E[T_{AR}] = \tau_{AR}$ — среднее время обслуживания пакета на A , и $E[T_{AR}^2]$ — второй момент времени обслуживания.

Продemonстрируем далее, как определить неизвестные компоненты в приведенном выше уравнении. Рассмотрим время обслуживания меченого пакета от A . Вспомним, что для каждого слота, для которого $Q_R \neq 0$ и $Q_A \neq 0$ пакет от A включен в расписание системы с вероятностью 0.5. Введем следующую дополнительную вероятность

$$\gamma_A \triangleq \Pr\{Q_R \neq 0 | Q_A \neq 0\} = \frac{\Pr\{Q_R \neq 0, Q_A \neq 0\}}{\Pr\{Q_A \neq 0\}}.$$

Более точно, планировщик либо назначает последующий слот для передачи R с вероятностью $0.5\gamma_A$, либо предоставляет этот слот A с дополнительной вероятностью $1 - 0.5\gamma_A$. Используя формулу полной вероятности и предположение 2, можно установить

$$0.5\gamma_A = 1 - \frac{\rho_{A0}}{\rho_{AR}}.$$

При этом распределение времени обслуживания пакета от A может быть получено следующим образом

$$\begin{aligned} \Pr\{T_{AR} = n < N_A\} &= p_{AB}(1 - 0.5\gamma_A) \times (1 - p_{AB}(1 - 0.5\gamma_A))^{n-1}, \\ \Pr\{T_{AR} = N_A\} &= (1 - p_{AB}(1 - 0.5\gamma_A))^{N_A-1}. \end{aligned}$$

Вычисляя первый и второй моменты времени обслуживания ($E[T_{AR}]$ и $E[T_{AR}^2]$) на основе формулы Полячека - Хинчина и формулы Литтла, можно аппроксимировать среднюю задержку пакета на A как

$$\delta_A \cong \frac{\rho_{AR}}{\lambda_A} + \frac{\lambda_A(2-p_{AB}(1-0.5\gamma_A))}{2(1-\rho_{AR})p_{AB}^2(1-0.5\gamma_A)^2} - \frac{\lambda_A(1-p_{AB}(1-0.5\gamma_A))^{N_A}}{2(1-\rho_{AR})} \times \\ \times \left[\frac{2N_A}{p_{AB}(1-0.5\gamma_A)} + \frac{2-p_{AB}(1-0.5\gamma_A)}{p_{AB}^2(1-0.5\gamma_A)^2} \right].$$

Метрики производительности для R могут быть вычислены аналогично по причине симметричной природы соответствующих прямых соединений. Кроме того, можно получить точное значение среднего уровня энергозатрат, например, для A

$$\varepsilon_A = P_{TX}\eta_A\tau_{A0} + P_I(1 - \eta_A\tau_{A0}),$$

для R результат аналогичен.

Здесь P_{TX} — средняя мощность, затрачиваемая устройством при передаче пакета, а P_I — средняя мощность, затрачиваемая в состоянии покоя. Таким образом, средняя энергоэффективность для A и R в системе без кооперации записывается следующим образом $\varphi_A = \eta_A/\varepsilon_A$ и $\varphi_R = \eta_R/\varepsilon_R$.

Система с кооперацией Для математического представления системы с кооперацией сначала рассмотрим очень важный специальный случай, когда очередь для узла R всегда пуста. Все соответствующие метрики для системы с кооперацией будем маркировать символом «*» по тексту.

Случай I. Очередь в узле R всегда пуста ($\lambda_R = 0$).

По аналогии с рассуждениями в предыдущей части можно выразить искомое распределение времени обслуживания пакета от узла A следующим образом

$$\Pr\{T_{A0}^* = n < N_A\} = \frac{X(1 - p_{CB})^{n-1} - Y[(1 - p_{AB})(1 - p_{AR})]^{n-1}}{1 - (1 - p_{AB})(1 - p_{AR})}, \\ \Pr\{T_{A0}^* = N_A\} = \frac{X(1 - p_{CB})^{N_A-1} - Y[(1 - p_{AB})(1 - p_{AR})]^{N_A-1}}{1 - (1 - p_{AB})(1 - p_{AR})},$$

где $X = \frac{p_{AR}(1-p_{AB})p_{CB}}{1-p_{CB}-(1-p_{AB})(1-p_{AR})}$ и $Y = X - p_{AB}$.

Переходя к среднему времени обслуживания, получаем

$$\tau_{A0}^* = \frac{X(1-(1-p_{CB})^{N_A})}{p_{CB}^2} - \frac{Y(1-[1-(1-p_{AB})(1-p_{AR})]^{N_A})}{[1-(1-p_{AB})(1-p_{AR})]^2}.$$

Случай II. Очередь узла R не всегда пуста ($\lambda_R > 0$).

Обобщим вышеизложенные рассуждения на более сложный кооперативный случай $\lambda_R > 0$. Опуская длинные, но тривиальные выкладки, представим соответствующее распределение времени обслуживания пакета от A следующим образом

$$\begin{aligned} \Pr\{T_{AR}^* = n < N_A\} &= X(1 - p_{CB}(1 - 0.5\gamma_A^*))^{n-1} - \\ &\quad - Y(1 - p_A(1 - 0.5\gamma_A^*))^{n-1}, \\ \Pr\{T_{AR}^* = N_A\} &= \frac{X(1-p_{CB}(1-0.5\gamma_A^*))^{N_A-1}}{p_{CB}(1-0.5\gamma_A^*)} - \frac{Y(1-p_A(1-0.5\gamma_A^*))^{N_A-1}}{p_A(1-0.5\gamma_A^*)}, \end{aligned}$$

где $0.5\gamma_A^* = 1 - \rho_{A0}^*/\rho_{AR}^*$ и $p_A = p_{AB} + p_{AR} - p_{AB} \cdot p_{AR}$ для краткости.

Коэффициенты загрузки очередей A и R (ρ_{AR}^* и ρ_{RA}^*) могут быть вычислены аналогично соответствующим параметрам для системы без кооперации с учетом того факта, что $\rho_{A0}^* \triangleq \lambda_A \tau_{A0}^*$, а выражение для τ_{A0}^* дано выше.

Наконец, вычисляя второй момент времени обслуживания, получаем окончательное выражение для аппроксимации средней задержки пакета на A

$$\begin{aligned} \delta_A^* &\cong \frac{\rho_{AR}^*}{\lambda_A} + \frac{\lambda_A}{2(1-\rho_{AR}^*)(1-0.5\gamma_A^*)^2} \left[X \cdot \frac{2-p_{CB}(1-0.5\gamma_A^*)}{p_{CB}^3} - Y \cdot \frac{2-p_A(1-0.5\gamma_A^*)}{p_A^3} \right] - \\ &\quad - \frac{\lambda_A X}{2p_{CB}(1-\rho_{AR}^*)} \cdot \left[\frac{2N_A}{p_{CB}(1-0.5\gamma_A^*)} + \frac{2-p_{CB}(1-0.5\gamma_A^*)}{p_{CB}^2(1-0.5\gamma_A^*)^2} \right] \times \\ &\quad \times (1 - p_{CB}(1 - 0.5\gamma_A^*))^{N_A} + \\ &\quad + \frac{\lambda_A Y}{2p_A(1-\rho_{AR}^*)} \cdot \left[\frac{2N_A}{p_A(1-0.5\gamma_A^*)} + \frac{2-p_A(1-0.5\gamma_A^*)}{p_A^2(1-0.5\gamma_A^*)^2} \right] (1 - p_A(1 - 0.5\gamma_A^*))^{N_A}, \end{aligned}$$

где X и Y были также определены выше в части, относящейся к пустой очереди R .

Заметим, что аппроксимация δ_A^* может быть улучшена с помощью следующей модификации

$$0.5\gamma_A^* = 1 - \frac{\rho_{A0}^*}{\rho_{AR}^*} \cdot (1 - P_{LA})^{N_A} \text{ и}$$

$$0.5\gamma_R^* = 1 - \frac{\rho_{R0}^*}{\rho_{RA}^*} \cdot (1 - P_{LR})^{N_R}.$$

С учетом τ_{A0}^* , окончательная аппроксимация для средней задержки пакета δ_R^* на R и выражения для пропускной способности η_A^* и η_R^* в системе с кооперацией аналогичны соответствующим метрикам для системы без кооперации, полученным в предыдущем разделе.

Аналогично, средние энергозатраты узла A в рассматриваемом случае будут составлять

$$\varepsilon_A^* = P_{TX}\eta_A^*\tau_{A0}^* + P_I(1 - \eta_A^*\tau_{A0}^*),$$

а средние энергозатраты для R могут быть определены как

$$\varepsilon_R^* = P_{TX} \left(\eta_R^*\tau_{R0} + \eta_A^* \cdot \frac{1 - \frac{p_{AB}\tau_{A0}^*}{1 - (1 - p_{AB})^{N_A}}}{p_{CB} - p_{AB}} \right) +$$

$$+ P_{RX} \left(1 - \eta_R^*\tau_{R0} - \eta_A^* \cdot \frac{1 - \frac{p_{AB}\tau_{A0}^*}{1 - (1 - p_{AB})^{N_A}}}{p_{CB} - p_{AB}} \right),$$

где P_{RX} — средняя мощность, затрачиваемая узлом при приеме пакета. Как выше было отмечено, средняя энергоэффективность A и R задается выражениями $\varphi_A^* = \eta_A^*/\varepsilon_A^*$ и $\varphi_R^* = \eta_R^*/\varepsilon_R^*$, соответственно.

3.3 Схема с клиентской ретрансляцией для M2M коммуникаций

3.3.1 Архитектура M2M

На рисунке 3.3 приведена архитектура системы, основанная на M2M архитектуре согласно стандарту IEEE 802.16 [126]. M2M устройство является абонентской станцией стандарта IEEE 802.16 с функциональностью

M2M. *M2M сервер* представляет собой сущность, обеспечивающую взаимодействие одного или более M2M устройств стандарта IEEE 802.16 с помощью базовой станции IEEE 802.16. M2M сервер имеет интерфейс, который может быть доступен потребителю M2M услуг (например, компании, оказывающей коммунальные услуги). Заметим, что архитектура системы M2M позволяет M2M устройству выступать в качестве *узла агрегации* для *не-IEEE 802.16 M2M устройств* (сенсоров или счетчиков) без задействования интерфейса сотовой сети. Подобные не-IEEE 802.16 M2M устройства могут, в свою очередь, использовать другие радиointерфейсы, например, такие как IEEE 802.11, IEEE 802.15, PLC и т.д.

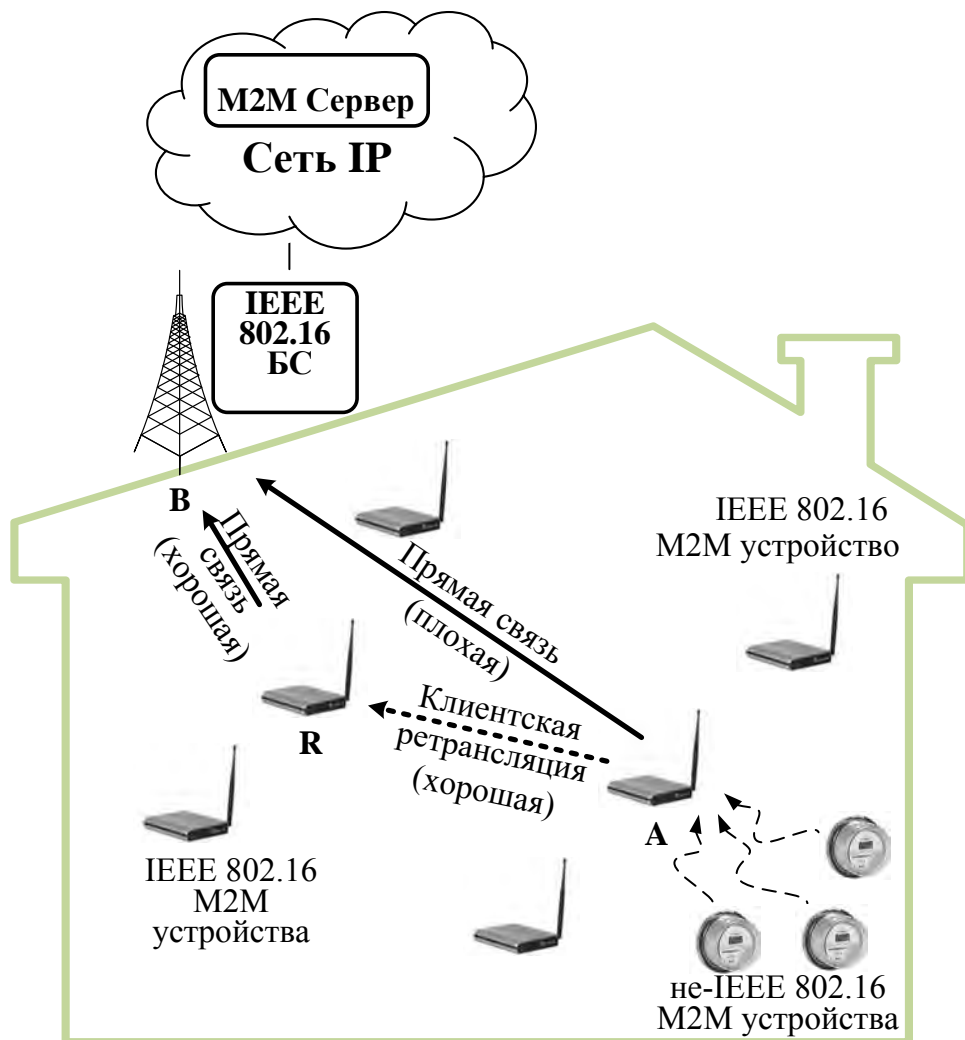


Рисунок 3.3 — Архитектура для передачи трафика M2M с кооперацией

Важно отметить, что M2M устройство стандарта IEEE 802.16 может также исполнять роль кооператора для других IEEE 802.16 M2M устройств. Это означает, что M2M устройство стандарта IEEE 802.16 R может ретранслировать трафик от имени, например, устройства A , чей беспроводной канал связи обладает недостаточным качеством, улучшая, таким образом, его производительность. В этом случае можно ожидать, что радиointерфейс IEEE 802.16 будет поддерживать функциональность *клиентской ретрансляции*. В частности, если обратиться к рисунку 3.3, можно отметить, что функционирование R должно обеспечить возможность прослушивания/перехвата трафика A . В ранних работах автора [119] был предложен простой протокол клиентской ретрансляции, который может быть использован в сотовых сетях. Ниже мы пересмотрим схему клиентской ретрансляции из [119] для случая M2M.

3.3.2 Модель системы и предположения

Рассмотрим M2M-совместимую систему с беспроводными сотовыми сетями, дополненную возможностями клиентской ретрансляции в соответствии с базовой методологией, приведенной в работе автора [119], и расширим ее ниже, имитируя работы протокола IEEE 802.16r. Сконцентрируем внимание на реалистичной топологии сети (рисунок 3.3), включающей в себя множество исходящих M2M и не-M2M устройств, а также один шлюз (базовую станцию). Оценим производительность меченого узла агрегации M2M трафика A и назовем ее *источником*. Источник агрегирует пакеты данных от N не-IEEE 802.16 M2M устройств со средней интенсивностью поступления λ_A . Назовем другое M2M устройство R *ретранслятором*. Ретранслятор генерирует собственные пакеты данных со средней интенсивностью поступления λ_R . Кроме того, ретранслятор прослушивает передачу данных от источника и может временно сохранять пакеты A для последующей ретрансляции. Устройство B представляет собой *базовую станцию* и получает пакеты данных и от источника, и от ретранслятора, а также от

других абонентов. Базовая станция не имеет исходящего трафика. Ниже представлена детализированная модель.

Предположение 1. Система. Системное время дискретно, будем называть единицу системного времени *фрейм*. Все передаваемые от М2М устройств пакеты данных имеют достаточно короткую длину, чтобы быть переданными в соответствии с их запросами в соответствии с процедурой *случайного доступа* [131]. В связи с этим, нет необходимости в явном расписании передачи пакетов данных, и устройства следуют процедуре, основанной на конфликтах согласно стандарту [112]. Для уменьшения расходов на управление, каждый фрейм имеет только одну возможность для случайного доступа — так называемый слот конкуренции или слот для захвата доступа в сеть (contention slot).

В качестве примера рассмотрим рисунок 3.4, где передача от R всегда успешна во фреймах под номерами 2, 4 и 6. Сфокусируем далее внимание на неудачных событиях для A . Поскольку ретранслятор не способен одновременно осуществлять и передачу, и прием, предположим, что исходящие устройства чередуются в доступе к каналу с вероятностью 0.5 (например, они могут быть разделены на две группы мультикаста). Изначально (в первом фрейме) A не удается успешно выполнить процедуру случайного доступа. Когда он предпринимает попытку повторной передачи далее во фреймах 3 и 5, можно наблюдать неудачный исход вследствие низкого качества канала. В свою очередь, в третьем фрейме R не может прослушать пакет от A , но затем успешно перехватывает его во фрейме номер 5. И, наконец, R передает перехваченные данные *одновременно* с A во фрейме номер 7 (создавая виртуальную ММО линию), и передача становится успешной благодаря усилению, вызванному использованию кооперации (более детально с процедурой можно ознакомиться в [119]). Таким образом, предложенная схема отличается от известных исследований в данной области [111], и достаточно проста для реализации.

Предположение 2. Трафик. Агрегация трафика от N счетчиков к источнику A имеет следующие реалистичные свойства (рисунок 3.5). Каждый счетчик создает специфический поток пакетов, поэтому предположим для упрощения, что потоки от различных счетчиков взаимно независи-

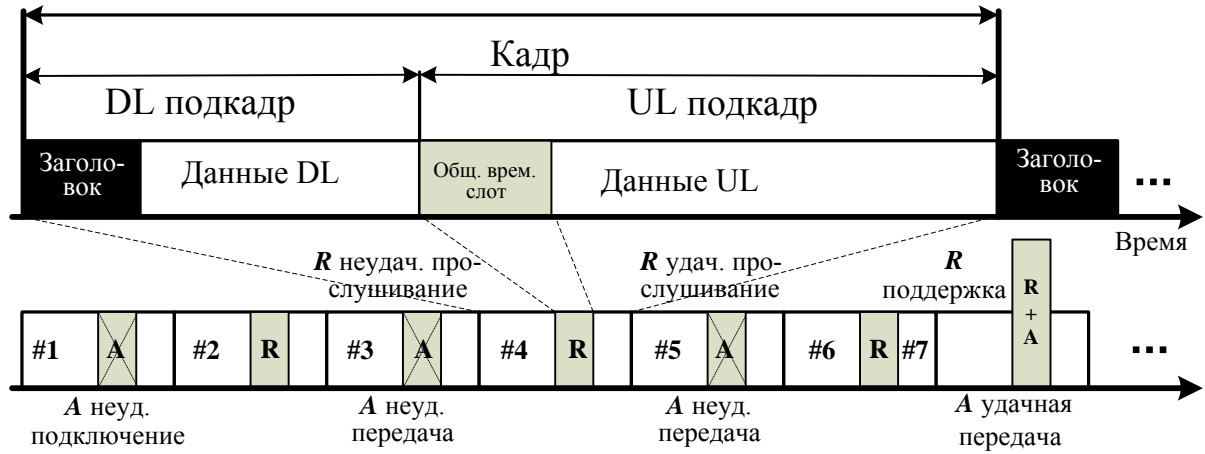


Рисунок 3.4 — Пример функционирования системы клиентской ретрансляции

мы. Каждый поток принадлежит к одному из видов потоков i , где $i = \overline{1, L}$. Здесь L — общее число типов источников трафика (пользовательские счетчики, аварийные счетчики и т.д.). Тип источника трафика определяется двумя параметрами: длиной периода без передачи информации T_i и длительностью передачи T_i^{ON} (или ON-периодом). На рисунке 3.5 приведен пример для $N = 3$, при этом все три типа источников трафика различны. Заметим, что счетчик имеет один ON-период в течение периода T_i . Как предложено в рекомендациях [132], случайный интервал времени между двумя последовательными ON-периодами (T_i^{OFF}) соответствует или равномерному $U[0, T_i - T_i^{ON}]$, или бета-распределению $Be(3, 4)$. Каждому счетчику с вероятностью p_i ставится в соответствие один из существующих типов i (используется случайное распределение вида $\{p_i\}_{i=1}^L$).

Собственный трафик ретранслятора предполагается пуассоновским. Заметим, что в более ранних работах автора [119] предполагалось, что пуассоновский характер входящего потока мог быть использован и для источника, и для ретранслятора. Однако, здесь приводится расширение модели для учета реальных характеристик M2M трафика для A . Предложенная методология позволяет также рассматривать агрегацию трафика на R , предположение о пуассоновском распределении всего лишь упрощает представление аналитических результатов. Интересно отметить, что в случае если число N счетчиков достаточно велико, агрегированный тра-

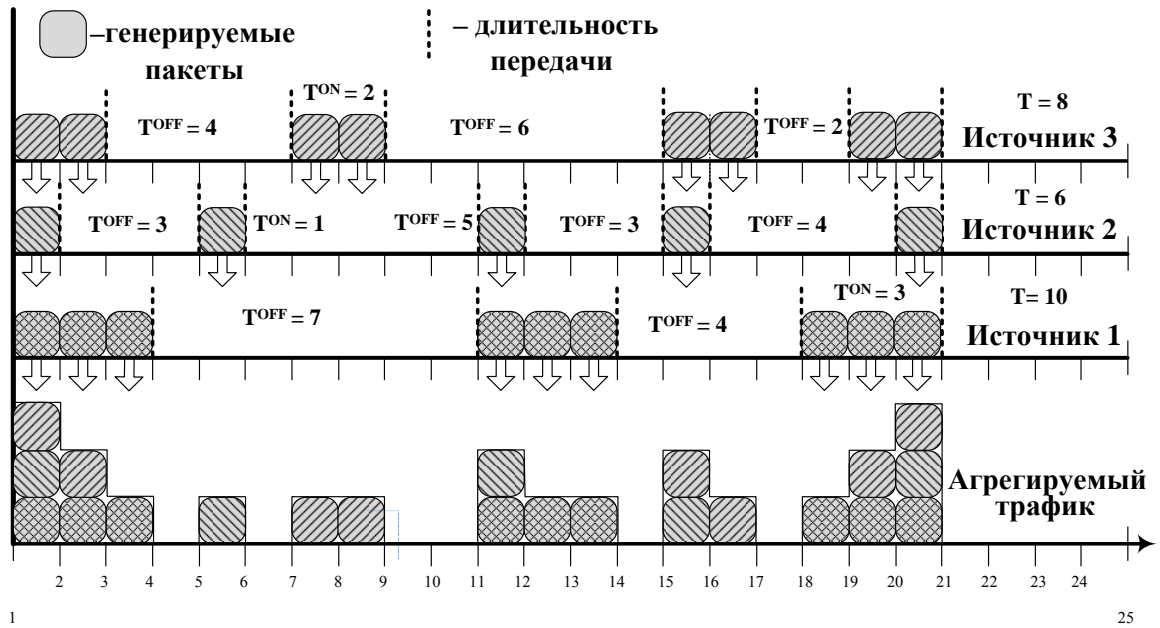


Рисунок 3.5 — Пример агрегации процессов поступления

фик демонстрирует высокую степень *самоподобия*. Оценим здесь параметр Херста [133] самоподобного процесса на A . Если большая часть источников N имеет ON-период T_i^{ON} больше 1, параметр Херста отличен от 0,5 (последнее характерно для пуассоновского трафика). Для примера, предположение о том, что все источники имеют $T^{ON} = 2$, приводит к трафику с параметром Херста H около 0,65. В случае, если $T^{ON} = 3$, значение H близко к 0,75. Это подтверждает тот факт, что трафик очевидно не является пуассоновским.

Специфика входящих потоков трафика на других типах устройств в системе не является важной для последующего анализа, поскольку влияет на A и R только через вероятность успеха в захвате доступа P_S (contention success probability). Последнее есть вероятность того, что в слоте конкуренции устройство получит доступ к каналу при условии осуществления попытки доступа; данная вероятность предполагается константой. На практике, на эту вероятность оказывает влияние множество факторов, таких как, например, конфигурация и параметры системы, и ее оценка является самостоятельной исследовательской задачей. Вместо этого далее изучим, каким образом метрики производительности зависят от этой вероятности.

Обозначения аналитической модели

Обозначения	описание параметров
λ_X	Средняя интенсивность поступления пакетов на X (A или R)
p_{AB}	Вероятность приема на B , если передает A
p_{RB}	Вероятность приема на B , если передает R
p_{AR}	Вероятность приема на R , если передает A
p_{CB}	Вероятность приема от A и R на B при кооперации
τ	Среднее время обслуживания пакета
ρ	Коэффициент загрузки очереди
q_X	Средняя длина очереди на X
δ_X	Средняя задержка X
η_X	Средняя пропускная способность X
ϵ_X	Среднее энергопотребление X
ϕ_X	Средняя энергоэффективность X
H_A	Параметр Херста самоподобного процесса для A
P_S	Вероятность успешного доступа в канал

Предположение 3. Клиенты. И источник, и ретранслятор имеют неограниченный буфер для собственных пакетов. Для обеих очередей предполагается дисциплина FIFO. Кроме того, ретранслятор имеет дополнительную память для пакетов данных от источника с целью последующей кооперативной передачи. Ранее в работах автора [119] было продемонстрировано, что одной ячейки памяти достаточно для функционирования схемы с клиентской ретрансляцией.

Предположение 4. Канал. Канал связи подвержен ошибкам и соответствует модели с приемом множества пакетов [117]. Передача пакетов данных успешна на приемной стороне с постоянной вероятностью, определяемой только типом линии (прямая или ретранслирующая) и тем, какие из узлов передают информацию одновременно. Ненулевые вероятности успешной передачи также, как и другие параметры модели системы, сведены в таблицу 3.1. Ожидается, что $p_{AR} > p_{AB}$, а также $p_{CB} > p_{AB}$. Если пакет не принят успешно, то он будет передан заново своим же источником. Максимальное число попыток передачи не ограничено.

3.3.3 Аналитические результаты

В первую очередь, аналитический подход подходит для оценки точного значения средней интенсивности обслуживания пакетов (пропускной способности) на A и R . Как было показано ранее, пропускная способность A определяется как

$$\eta_A = \begin{cases} \lambda_A, \text{ без насыщения} \\ (1 - \lambda_R \tau_{R0}) \tau_{A0}^{-1}, \text{ насыщение для } A \\ (2\tau_{A0})^{-1}, \text{ насыщение для } A, R. \end{cases}$$

Пропускная способность узла R может быть определена аналогично. Условия насыщения определяются следующим образом

- Для A : $(\lambda_A \tau_{A0} + \lambda_R \tau_{R0} > 1)$ и $(\lambda_R \tau_{R0} < 0.5)$.
- Для R : $(\lambda_A \tau_{A0} + \lambda_R \tau_{R0} > 1)$ и $(\lambda_A \tau_{A0} < 0.5)$.
- Для A и R : $(\lambda_A \tau_{A0} > 0.5)$ и $(\lambda_R \tau_{R0} > 0.5)$.

Среднее время обслуживания пакета на A при условии, что $\lambda_R = 0$, есть

$$\tau_{A0} = \frac{p_{AR}(1-p_{AB})}{1-p_{CB}-(1-p_{AB})(1-p_{AR})} \frac{1}{p_{CB}} - \left(\frac{p_{AR}(1-p_{AB})p_{CB}}{1-p_{CB}-(1-p_{AB})(1-p_{AR})} - p_{AB} \right) \frac{1}{[1-(1-p_{AB})(1-p_{AR})]^2}.$$

Заметим, что среднее время обслуживания пакета на R при условии, что $\lambda_A = 0$, может быть оценена тривиально как $\tau_{R0} = p_{RB}^{-1}$.

Далее, можно найти точное значение среднего энергопотребления A следующим образом

$$\varepsilon_A = P_{TX} \eta_A \tau_{A0} + P_I (1 - \eta_A \tau_{A0})$$

и для R как

$$\begin{aligned} \varepsilon_R = & P_{TX} \left(\eta_R \tau_{R0} + \eta_A \frac{1 - \tilde{p}_{AB} \tau_{A0}}{\tilde{p}_{CB} - \tilde{p}_{AB}} \right) + P_{RX} \left(\eta_A \tau_{A0} - \eta_A \frac{1 - \tilde{p}_{AB} \tau_{A0}}{\tilde{p}_{CB} - \tilde{p}_{AB}} \right) \\ & + P_I (1 - \eta_R \tau_{R0} - \eta_A \tau_{A0}). \end{aligned}$$

Здесь P_{TX} — средняя мощность, затрачиваемая на передачу пакета, P_{RX} — средняя мощность, затрачиваемая на прием, а P_I представляет собой сред-

нюю мощность затрачиваемую в случае простаивания устройства. Средняя энергоэффективность A и R задается равенствами $\varphi_A = \eta_A \varepsilon_A^{-1}$ и $\varphi_R = \eta_R \varepsilon_R^{-1}$, соответственно.

С учетом вышеприведенного анализа, $\tilde{\rho}_{AB} = P_S p_{AB} \rho_{A0} \rho_{AR}^{-1}$, а $\tilde{\rho}_{CB} = P_S p_{CB} \rho_{A0} \rho_{AR}^{-1}$. Коэффициент загрузки очереди на A при условии, что $\lambda_A = 0$, равен $\rho_{A0} = \lambda_A \tau_{A0}$, а коэффициент загрузки очереди на R при условии, что $\lambda_R = 0$, есть $\rho_{R0} = \lambda_R \tau_{R0}$. Если мы теперь рассмотрим для примера $\rho_{A0} > \rho_{R0}$, то коэффициент загрузки очереди на A переписется как $\rho_{AR} \cong \rho_{A0}(1 - \rho_{R0})^{-1}$, а коэффициент загрузки очереди на R — как $\rho_{RA} = \rho_{AR} - \rho_{A0} + \rho_{R0} \cong \rho_{A0}(1 - \rho_{R0})^{-1} - \rho_{A0} + \rho_{R0}$ (можно обратиться к работе автора [119] для более детальной информации).

Наконец, рассмотрим среднюю задержку пакета для источника A . Вследствие самоподобной природы трафика невозможно использовать предыдущий подход [119], основанный на формуле Полячека-Хинчина, поскольку он применим только для пуассоновских входных потоков. Однако, принимая во внимание значение параметра Херста H_A , можно усовершенствовать аналитическую модель, как предлагается в [133]. Для кооперативных систем аппроксимация средней задержки пакета для узла A определяется следующим образом

$$\delta_A \cong \tau_{AR} + \frac{\rho_{AR}^{\frac{0.5}{1-H_A}}}{\tau_{AR}} \frac{1}{2(1-\rho_{AR})^{\frac{H_A}{1-H_A}}} \left[\frac{X}{\tilde{\rho}_{CB}} \frac{2-\tilde{\rho}_{CB}}{\tilde{\rho}_{CB}^2} - \frac{Y}{\tilde{\rho}_A} \frac{2-\tilde{\rho}_A}{\tilde{\rho}_A^2} \right],$$

где $\tau_{AR} = \rho_{AR} \lambda_A^{-1}$, и введены дополнительные переменные $X = \frac{\rho_{AR} P_S (1-\tilde{\rho}_{AB}) \tilde{\rho}_{CB}}{1-\tilde{\rho}_{CB}-(1-\tilde{\rho}_{AB})(1-\rho_{AR})}$ и $Y = X - \tilde{\rho}_{AB}$. Также для краткости $\tilde{\rho}_A = \tilde{\rho}_{AB} + \rho_{AR} P_S - \tilde{\rho}_{AB} \rho_{AR} P_S$.

Средняя задержка пакета ретранслятора R может быть определена аналогично средней задержке источника A в случае без кооперации (когда $\rho_{AR} = 0$). При этом параметр Херста H_A может быть оценен с помощью известной процедуры [134]. Как и ожидалось, для специального случая, когда $T_i^{ON} = 1$, агрегированный поток становится пуассоновским ($H_A = 0.5$) и статистические тесты (критерии Пирсона и Колмогорова-Смирнова) подтверждают экспоненциальное распределение промежутков времени между

поступлениями пакетов. Таким образом, приведенная выше формула δ_A сводится к соответствующему уравнению из [119].

3.4 Клиентская ретрансляция в беспроводных сотовых сетях

3.4.1 Модель системы

В этом разделе исследуем следующую модель: беспроводная сотовая сеть содержит базовую станцию B и несколько мобильных клиентов (см. рисунок 3.6 для топологии сети и таблицы 3.1 и 3.2 для обозначений).

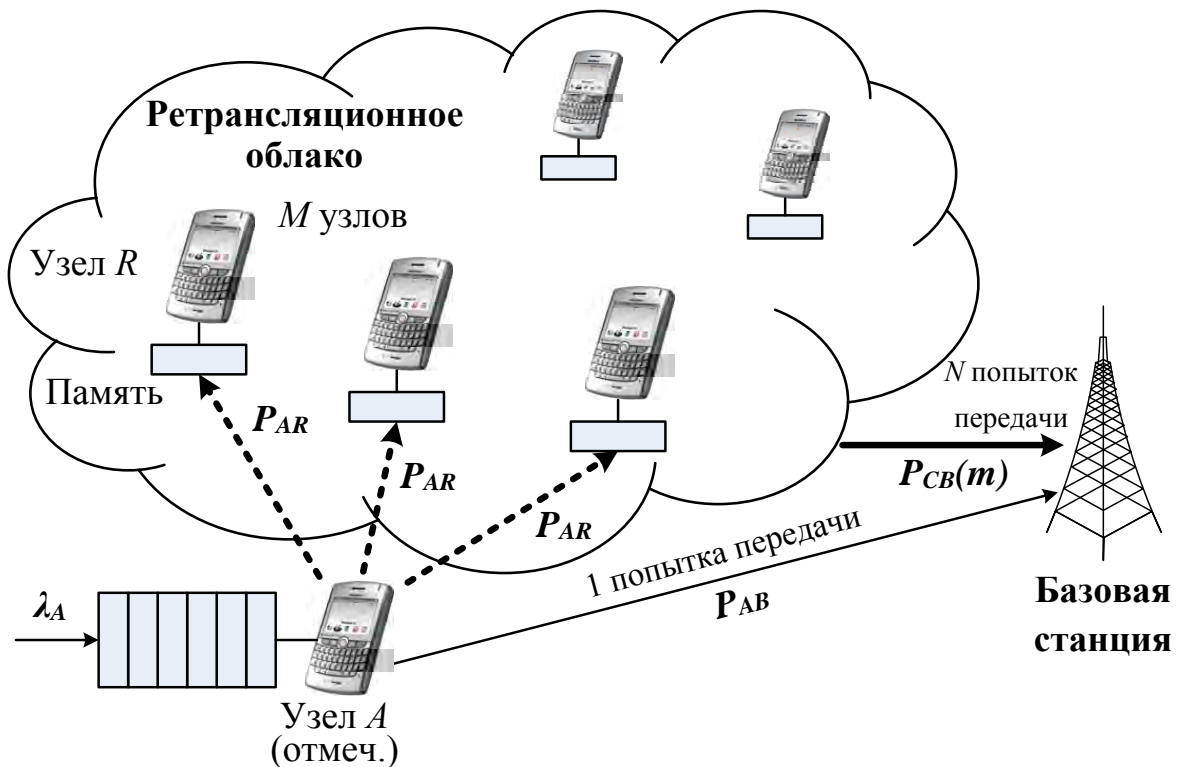


Рисунок 3.6 — Топология ретрансляционного облака

Беспроводные клиенты, выступающие как узлы ретрансляции, могут прослушивать пакеты данных от источника. Как ранее отмечалось, агрегированный набор доступных клиентских ретрансляторов формирует облако ретрансляции. В случае успешного прослушивания пакета от источника

Обозначения аналитической модели

Обозначения	параметры
N	Максимальное число попыток облачной передачи, продиктованное B
M	Число узлов ретрансляции в облаке
m	Число узлов ретрансляции в группе
$p_{CB}(m)$	Вероятность успешного приема на B при облачной кооперации
P_{lossA}	Вероятность потери пакета от A
P_{TX}	Уровень мощности передачи
P_{RX}	Уровень мощности прослушивания
P_I	Уровень мощности простаивания

участники облака могут оппортунистически передавать пакеты от его имени с целью увеличения производительности. Базовая станция обеспечивает информацию о временных ресурсах (слотах) для такой кооперативной передачи, в то время как оппортунистическое управление осуществляется с клиентской стороны.

Далее, для упрощения и без потери общности будем рассматривать пропускную способность меченого устройства A (источник данных). Предположим для примера, что A представляет собой мобильного пользователя, расположенного на краю соты, и имеющего, таким образом, низкое качество восходящего канала до базовой станции. Остальные M близлежащих беспроводных устройств (образующие облако ретрансляции) потенциально могут обеспечить кооперацию в качестве ретрансляторов.

В то время как источник передает исходный пакет данных, каждый из узлов ретрансляции в облаке может перехватить этот пакет и сохранить для последующей передачи. Размер дополнительной памяти для каждого из ретрансляторов предполагается равным единице для каждой сессии ретрансляции, в то время как размер буфера исходящего источника неограничен. В случае, когда первичная передача пакета от источника не успешна, а перехват пакета выполнен удачно, узел ретрансляции R принимает вероятностное решение о том, будет ли осуществляться кооперация с его стороны или нет.

Узлы ретрансляции, которые принимают решение о кооперации, будем называть группой ретрансляции. Отметим, что предложенная схема не требует явного централизованного управления со стороны базовой станции и, таким образом, минимизирует расходы на сигнальные процедуры. Базовая станция может быть совершенно не осведомлена о том, какие узлы находятся в группе ретрансляции в конкретный момент времени. Рассмотрим более детально системную модель.

Предположения о трафике. Для оценки производительности системы и сохранения аналитичности модели, рассматриваем простую стохастическую модель трафика. В качестве первого шага для такого исследования предположим, что случайные интервалы между поступлениями заявок на источник A независимы, одинаково распределены и отвечают экспоненциальному распределению. Сконцентрируем внимание на трафике одного источника, абстрагируясь от анализа собственного трафика в облаке ретрансляции, который может быть более сложным. Базовая станция в данном случае также не создает исходящего трафика.

Предположения о планировщике. Время системы разбито на слоты. Размеры пакетов одинаковы, и передача пакета занимает в точности один слот. Информация о расписании при этом доступна незамедлительно для всех участников (например, посредством выделенных нисходящих каналов).

Рассмотрим функционирование такого планировщика. Если канал между A и получателем имеет низкое качество, очень вероятно, что некоторые пакеты могут быть переданы неудачно. В связи с этим, предположим, что только одна попытка передачи может быть осуществлена источником исходя из соображений экономии энергии. Если источник A имеет пакеты для передачи, то следующий слот отдается ему. Во время передачи потенциальный ретранслятор может перехватить и сохранить пакет от A .

В случае, если A терпит неудачу в передаче своего пакета, базовая станция может отдать следующий слот облаку ретрансляции с целью помочь источнику в передаче пакета. Такое переназначение слотов может повторяться либо до успешной доставки пакета, либо до определенного числа повторных передач, осуществленных облаком ретрансляции, не превыша-

ющего некоторое максимальное значение N (параметр устанавливается базовой станцией). В последнем случае, все участники облака могут обнулить свою память, и система рассматривает такой пакет потерянным.

Предположения о канале. Предполагаем, что существует немедленная обратная связь по выделенному каналу (например, по нисходящей линии). Также будем учитывать вероятности успешной доставки p_{AB} , $p_{CB}(m)$ и такую же вероятность для каждого ретранслятора p_{AR}

- $p_{AB} = Pr\{\text{пакет от A получен на B} \mid \text{передает только узел A}\}$,
- $p_{CB}(m) = Pr\{\text{пакет от A получен на B} \mid \text{передают точно } m \text{ ретрансляторов}\}$,
- $p_{AR} = Pr\{\text{пакет от A получен заданным ретранслятором} \mid \text{передает только A}\}$.

Проиллюстрируем функционирование планировщика на примере, когда $N = 2$, как это показано на рисунке 3.7. Вначале передача (нулевого) пакета успешна, и система находится в равновесном состоянии. Далее, источник генерирует новый пакет с номером 1 и осуществляет попытку его передачи, которая заканчивается неудачно с вероятностью $1 - p_{AB}$. Участники облака ретрансляции независимо перехватывают пакет, каждый с вероятностью p_{AR} . В следующем слоте успешные ретрансляторы принимают решение о кооперативной поддержке источника с вероятностью p_{tx} . Те ретрансляторы, которые приняли положительное решение, одновременно передают пакет на базовую станцию. Таким образом, создается некая виртуальная MIMO линия с лучшим качеством обслуживания за счет пространственного разнесения передачи [119], и пакет будет передан успешно с вероятностью $p_{CB}(m)$.

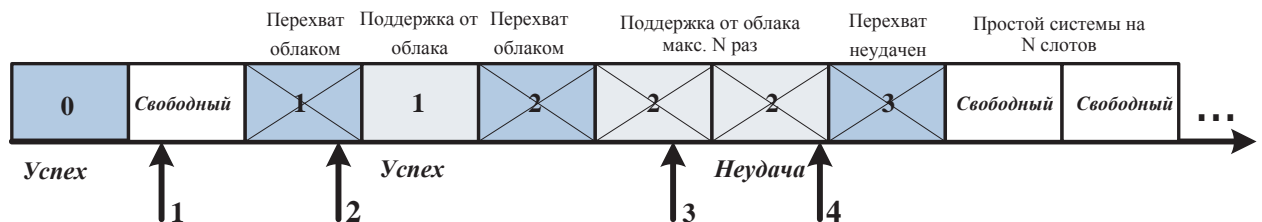


Рисунок 3.7 — Пример временной диаграммы для облака ретрансляций

Отметим, и это важно, что вследствие выигрыша от разнесения, вероятность $p_{CB}(m)$ представляет собой неубывающую функцию для группы ретрансляторов размером m . Здесь m зависит от вероятностей p_{AR} и p_{tx} . Также будем полагать, что качество канала на участке “источник-базовая станция” низкое, в то время как качество канала “источник-облако ретрансляторов” существенно лучше вследствие множества доступных ретрансляторов вблизи клиентов.

Следующий слот отдается источнику A (пакет номер 2), и во время его неудачной передачи ретрансляторы вновь перехватывают пакет. Однако в это время облако ретрансляторов — также неудачно — пытается передать пакет $N = 2$ раза подряд. Таким образом, базовая станция рассматривает этот пакет как потерянный и предоставляет следующий слот источнику. Далее, если неудачен перехват переданного пакета (пакет номер 3) или все успешные ретрансляторы решают не передавать перехваченный пакет, $N = 2$ слотов предоставляются облаку ретрансляторов в любом случае, но система простаивает. Это является негативным следствием использования распределенного управления в клиентской ретрансляции.

В соответствии с вышеизложенным, исследуем среднюю задержку пакета, пропускную способность и вероятность потери пакета. При этом в результате должны быть получены простые и точные выражения в замкнутом виде.

3.4.2 Оценка производительности

В этой части проанализируем систему облачной ретрансляции с точки зрения основных метрик производительности, таких как среднее *число повторных передач*, *пропускная способность*, *вероятность потери пакета* и *средняя задержка пакета*.

Сначала введем следующие определения:

Определение 1. Время обслуживания определяется как временной интервал между началом первой попытки передачи и моментом достиже-

ния пакетом базовой станции. В случае потери пакета его длительность обслуживания принимается равной N , тем самым, потерянные пакеты вносят свой вклад в среднее время обслуживания.

Определение 2. Предельная пропускная способность определяется как предел пропускной способности системы при неограниченном росте входной нагрузки [135].

Определение 3. Задержка пакета определяется как длительность, в течение которой пакет достигает точки назначения после его поступления в систему (включая и длительность ожидания в очереди, и длительность обслуживания).

Определение 4. Энергоэффективность определяется как количество энергии, требуемое для успешной передачи одного пакета данных.

Предлагаемый аналитический подход базируется на понятии времени обслуживания. Пусть T_A — случайная величина, которая является временем обслуживания пакета от A . Будем рассматривать систему вида $M/G/1$, что соответствует свойствам входящего трафика. Зададим дисциплину обслуживания и продолжим исследования для получения в замкнутой форме выражений для первого и второго моментов случайной величины. Заметим, что первый момент представляет собой среднее число попыток передачи пакетов.

Зная оба момента, можно найти среднюю задержку пакета, используя формулу Полячека-Хинчина и формулу Литтла. Другие метрики, вызывающие интерес, такие как пропускная способность, вероятность потери пакета, энергопотребление и энергоэффективность, можно будет определить из полученных выражений.

После тщательного анализа всех возможностей для передачи пакетов, сформулируем следующую дисциплину обслуживания для A

$$Pr\{T_A = 1\} = p_{AB},$$

$$Pr\{T_A = n\} = (1 - p_{AB}) \sum_{m=1}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \times \\ \times \left\{ \sum_{j=1}^m \binom{m}{j} p_{tx}^j (1 - p_{tx})^{(m-j)} (1 - p_{CB}(j))^{(n-1)} p_{CB}(j) \right\},$$

$$n \leq N,$$

$$Pr\{T_A = N + 1\} = (1 - p_{AB}) \sum_{m=1}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} (1 - p_{tx})^m + \\ + (1 - p_{AB}) \sum_{m=1}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \times \\ \times \left\{ \sum_{j=1}^m \binom{m}{j} p_{tx}^j (1 - p_{tx})^{(m-j)} (1 - p_{CB}(j))^N \right\} + \\ + (1 - p_{AB}) (1 - p_{AR})^M.$$

Далее, избегая масштабных преобразований, приводим выражения только для первых двух моментов случайной величины T_A

$$\tau_A = E[T_A] = p_{AB} + (N + 1)(1 - p_{AB})(1 - p_{AR})^M + \\ + (N + 1)(1 - p_{AB}) \cdot S_1 + (1 - p_{AB}) \cdot S_2, \quad (3.1)$$

$$E[T_A^2] = p_{AB} + (N + 1)^2(1 - p_{AB})((1 - p_{AR})^M + S_1) + (1 - p_{AB}) \cdot S_3, \quad (3.2)$$

где компоненты S_1 , S_2 и S_3 приведены ниже.

Средняя загрузка очереди для меченого устройства A может быть оценена как

$$\rho_A = \lambda_A \tau_A. \quad (3.3)$$

Средняя пропускная способность A может вычислена следующим образом

$$\eta_A = \lambda_A(p_{AB} + (1 - p_{AB})S_4), \quad (3.4)$$

где компоненты S_4 также приведены ниже.

Задавая первый и второй моменты, используем формулу Полячека-Хинчина для получения точных значений величины средней задержки пакета

$$\delta_A = \tau_A + \frac{E[T_A^2]\lambda_A}{2(1 - \rho_A)}. \quad (3.5)$$

Имея теперь базовую формулу для первых двух моментов, можно также найти другие важные метрики. Например, вероятность потери пакета будет определена так

$$P_{lossA} = 1 - \eta_A\tau_A. \quad (3.6)$$

Далее, давайте найдем выражения для потребления энергии. Если уровень мощности P_i соответствует мощности в состоянии i , тогда нормированное энергопотребление на слот равно $P_i\pi_i$. Здесь $\vec{\pi}$ есть стационарное распределение для мощностей состояний и $i \subset G$, где G — набор возможных состояний. В рассматриваемой модели три состояния являются важнейшими с точки зрения анализа мощности:

- устройство передает данные с мощностью P_{TX} ;
- получает данные с мощностью P_{RX} ;
- простаивает/находится в состоянии ожидания с мощностью P_I .

Таким образом, энергопотребление для A и группы ретрансляторов может быть определены следующим образом

$$\epsilon_A = P_{TX}\lambda_A + P_I(1 - \lambda_A\tau_A), \quad (3.7)$$

$$\begin{aligned} \epsilon_R = P_{RX}M\lambda_A + P_{TX}p_{AR}p_{tx}\lambda_A(\tau_A - 1)M + \\ + P_iM(1 - p_{AR}p_{tx}\lambda_A(\tau_A - 1) - \lambda_A), \end{aligned} \quad (3.8)$$

Тогда общее энергопотребление для системы составит

$$\epsilon = \epsilon_A + \epsilon_R. \quad (3.9)$$

Как было отмечено выше, энергоэффективность определяется как

$$\phi = \frac{\eta_A}{\epsilon}. \quad (3.10)$$

3.4.3 Вспомогательные переменные

Рассмотрим вспомогательные переменные, позволяющие упростить аналитические выражения, полученные выше.

$$S_1 = \sum_{m=0}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \times \left\{ \sum_{j=0}^m \binom{m}{j} p_{tx}^j (1 - p_{tx})^{(m-j)} (1 - p_{CB}(j)) + (1 - p_{tx})^m \right\}^N \quad (3.11)$$

и

$$S_2 = \sum_{m=0}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \cdot X \cdot b, \quad (3.12)$$

где

$$X = a^N - \frac{a^{(N+1)}}{(1-a)^2} - \frac{(a-2)}{(1-a)^2} - a^N \frac{(N+2)}{(1-a)}. \quad (3.13)$$

а

$$S_3 = \sum_{m=0}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \cdot Z \cdot b, \quad (3.14)$$

где

$$Z = X + \frac{2(2a - a^{(N+1)}(N+2))}{(1-a)^2} + \frac{2(a^2 - a^{(N+2)})}{(1-a)^3} + \frac{2 - a^N(N+1)(N+2)}{(1-a)}. \quad (3.15)$$

Следующая переменная представляет собой вероятность успешной передачи ретрансляторами

$$S_4 = \sum_{m=0}^M \binom{M}{m} p_{AR}^m (1 - p_{AR})^{(M-m)} \cdot Y \cdot b, \quad (3.16)$$

где

$$Y = \frac{(1 - a^N)}{(1 - a)}. \quad (3.17)$$

Здесь для краткости вводятся переменные a и b

$$a = \sum_{j=0}^m \binom{m}{j} p_{tx}^j (1 - p_{tx})^{(m-j)} (1 - p_{CB}(j)) + (1 - p_{tx})^m, \quad (3.18)$$

$$b = \sum_{j=0}^m \binom{m}{j} p_{tx}^j (1 - p_{tx})^{(m-j)} p_{CB}(j). \quad (3.19)$$

3.5 Численные результаты

3.5.1 Система трехузловой клиентской ретрансляции

В этой части верифицируем только предложенные в диссертации аппроксимации средней длительности задержки (А-I и А-II), поскольку другие метрики оценены точно. Используем разработанную под руководством автора систему имитационного моделирования WINTERsim, детально рассмотренную в [136] и [125]. Частично следуя [117], в качестве параметров моделирования используем $p_{AB} = 0,3$, $p_{RB} = 0,7$, $p_{AR} = 0,4$, $p_{CB} = 0,5$, $N_A = N_R = 8$, $\lambda_R = 0,15$, где λ_A варьируется в области устойчивого функционирования системы. В соответствии с [137] обозначим нормализованную потребляемую энергию как P_{TX} , и положим $P_{TX} = 1,00$, $P_{RX} = 0,85$ и

$P_I = 0,70$. На рисунке 3.8 проведено сравнение двух сценариев: с кооперацией и без таковой.

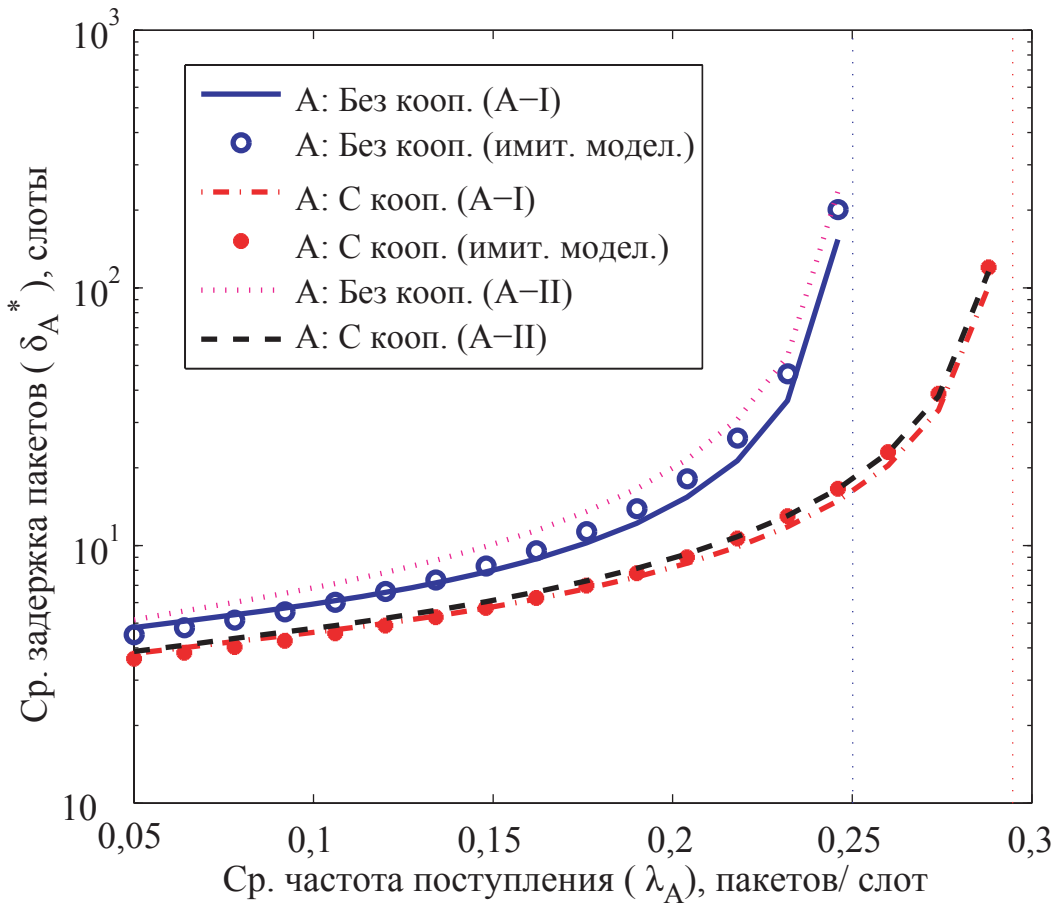


Рисунок 3.8 — Сравнение аналитического и имитационного моделирования для средней задержки пакета

Как видим, обе предложенные аппроксимации А-I и А-II показывают превосходное согласие с результатами моделирования (рисунок 3.8). Далее, принимая во внимание мало реалистичное предположение о пуассоновском характере входящего потока, исследуем также пачечную модель входящего потока (рисунок 3.9). Заметим, что в каждом слоте средняя скорость входящего потока, равная λ_A , или нулевая или сверх предельной и принимает значение 0,3 с вероятностью P . При этом вероятность прихода пачки показывает, что предложенные аппроксимации все еще демонстрируют очень хорошие результаты.

Широкое распространение беспроводных сетей открывает все новые научно-исследовательские направления, включая клиентскую кооперацию,

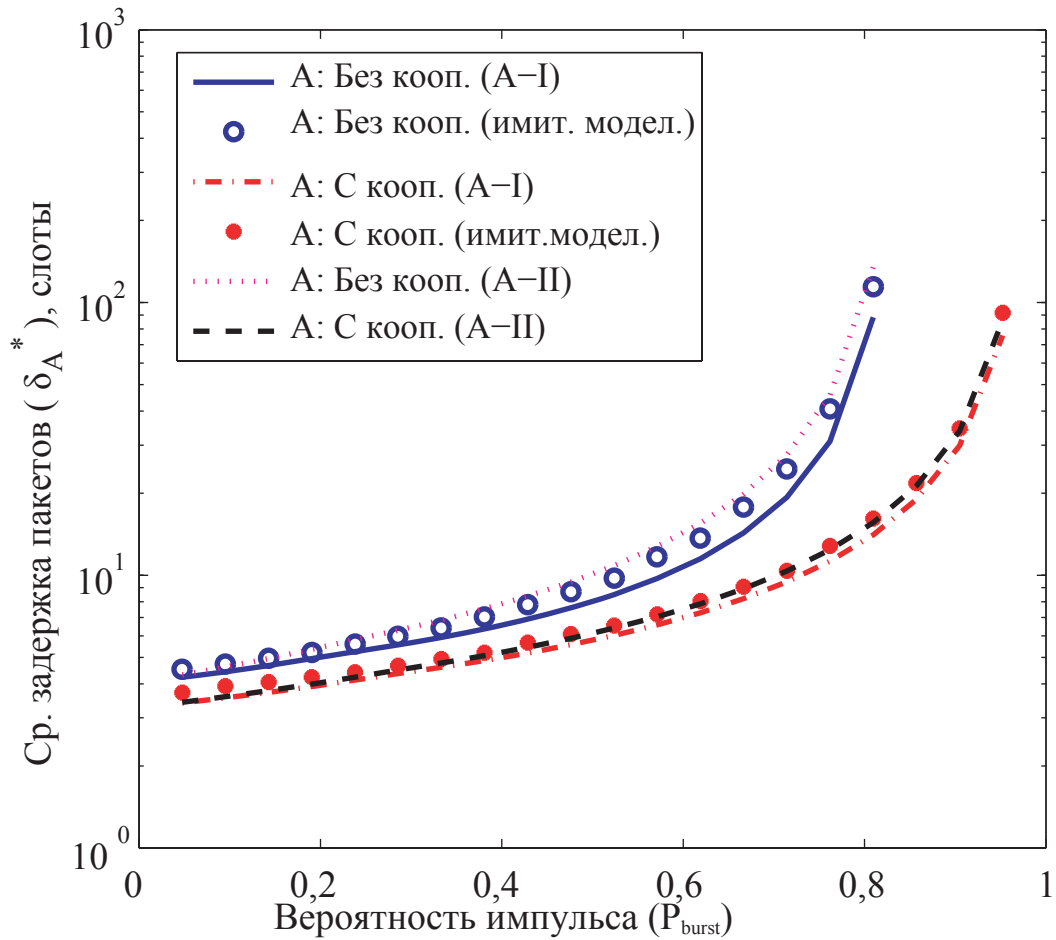


Рисунок 3.9 — Пачечное поступление заявок

энергетически эффективные взаимодействия, сосуществование различных технологий, агрегацию спектра и другие. Эти направления пока еще недостаточно покрываются традиционной методологией моделирования и существующими аналитическими методами, поскольку последние нацелены на статические или почти статические условия использования сетей связи [119]. Кроме того, известные модели не учитывают многие реалистичные факторы, такие как входящие потоки трафика, предопределенные характеристики качества обслуживания, деградацию беспроводного канала и т.п.

3.5.2 Клиентская ретрансляция для взаимодействия машина-машина

Методология моделирования Для верификации полученных аналитических результатов использовалась система имитационного моделирования WINTERsim. Частично следуя [117], параметры моделирования были представлены следующим набором: $p_{AB} = 0,3$, $p_{RB} = 0,7$, $p_{AR} = 0,4$, $p_{CB} = 0,5$, где λ_R фиксирована на уровне 30 пакетов в секунду. Кроме того, дополнительно используются значения потребления энергии из [137]: $P_{TX} = 1,65$ В, $P_{RX} = 1,40$ В и $P_I = 1,15$ В. Агрегированный трафик от большого числа счетчиков моделируется в соответствии с [132] (для $N = 1000, 3000, 5000$). Используется также реалистичное предположение $T_i^{ON} = 1,2,3$ и принимаются во внимание значения T_i , как это предложено в [138]: $T_1 = 900$ с, $T_2 = 300$ с, $T_3 = 60$ с и $T_4 = 10$ с.

Набор типов источников выбирается в количестве $L = 7$. Суммируя, имеем следующие периоды

$$T_i = \{T_1, T_1, T_1, T_2, T_2, T_3, T_4\},$$

длины ON-периодов

$$T_i^{ON} = \{3, 2, 1, 2, 1, 1, 1\},$$

и типы вероятностей

$$\{p_i\}_{i=1}^L = \{0,20; 0,01; 0,01; 0,75; 0,01; 0,01; 0,01\}.$$

Для конкретной топологии беспроводной системы можно оценить число счетчиков с заданной интенсивностью поступления данных и т.п. Фиксируем также интенсивность поступления λ_A как 30 пакетов в секунду для источника. Для численного примера оценим число источников, используя

$$\lambda_A = \frac{N}{W} \sum_{i=1}^L \frac{T_i^{ON} p_i}{T_i},$$

где T_i^{ON} , p_i и T_i — параметры системной топологии, непосредственно связанные с видами источников, W — матожидание для выбранного распределения. Будем рассматривать $\alpha = 3$ и $\beta = 4$ для бета-распределения [132], тогда $W = \frac{1}{2}$ для равномерного распределения и $W = \frac{3}{7}$ для бета-распределения $Be(3,4)$.

Результаты моделирования На рисунке 3.10 и рисунке 3.12 зафиксированы значения скорости поступления $\lambda_A = 30$ пакетов в секунду, а вероятность коллизий $P = 1 - P_S$ варьируется. На рисунке 3.11 и рисунке 3.13 переменной является интенсивность поступления пакетов и число источников данных при фиксированной вероятности коллизий 0.1.

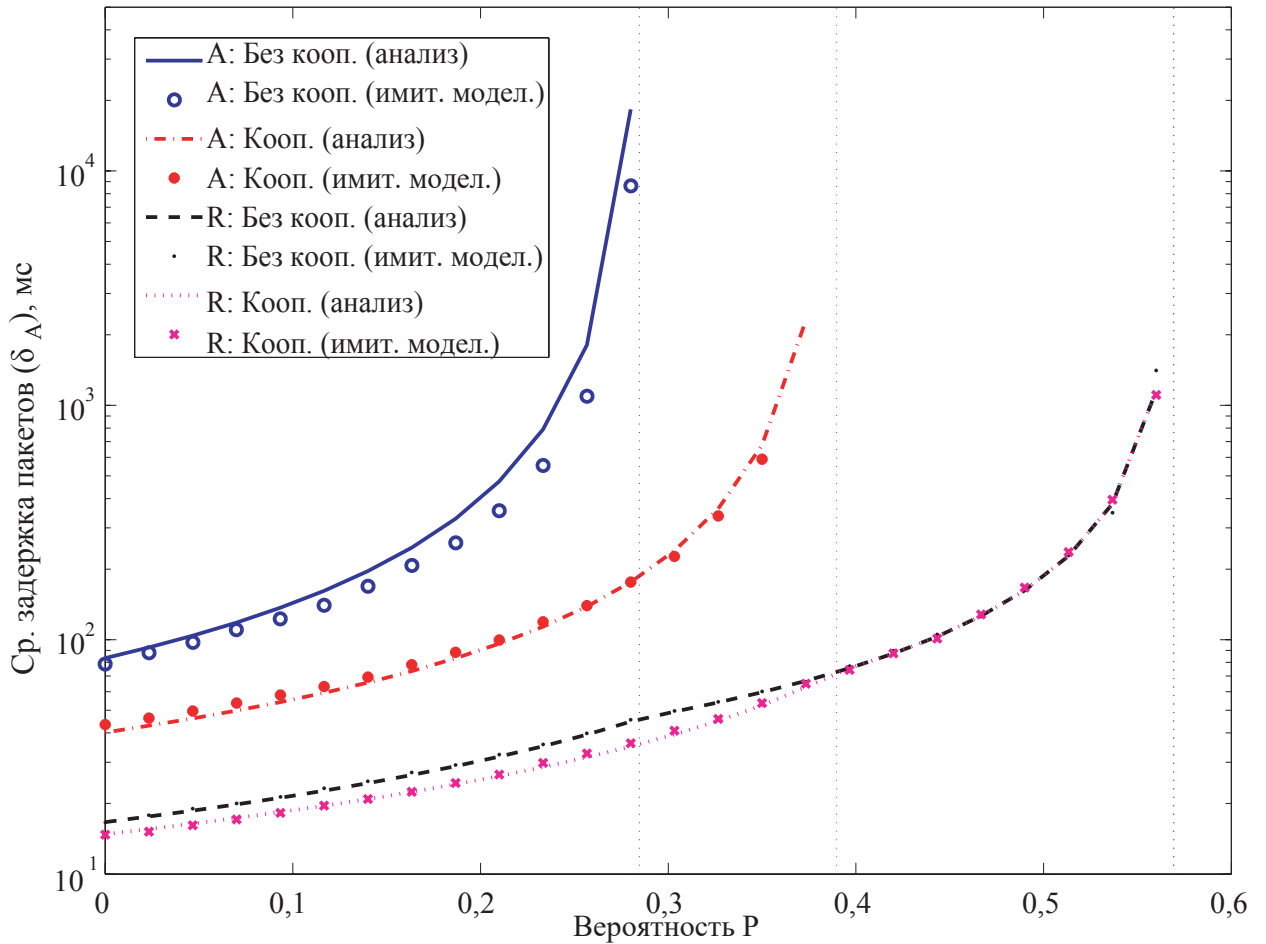


Рисунок 3.10 — Средняя задержка пакета в зависимости от вероятности коллизии

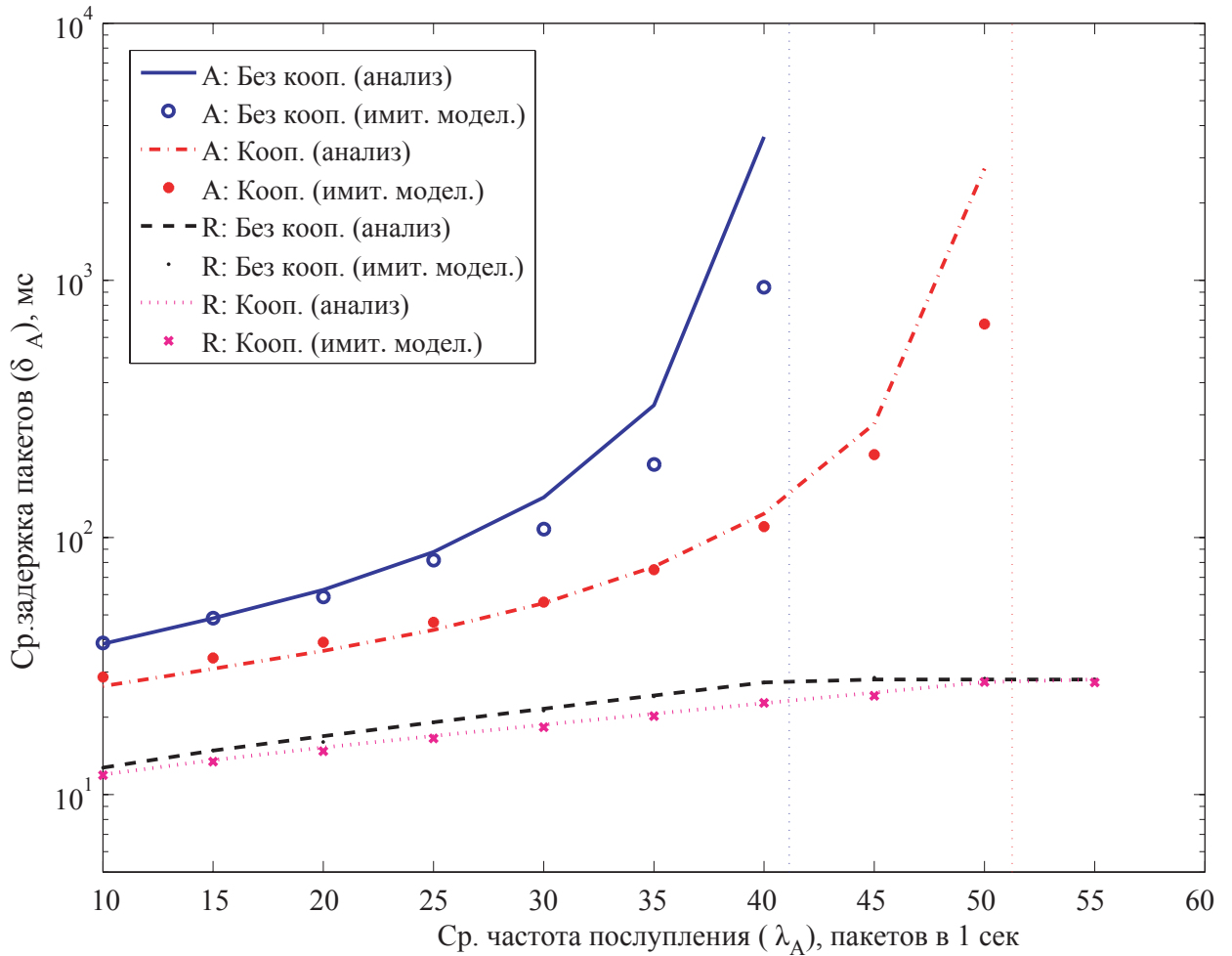


Рисунок 3.11 — Средняя задержка пакета в зависимости от средней интенсивности поступления пакетов

Рисунок 3.10 показывает, что с ростом вероятности коллизий в канале происходит увеличение средней длительности задержки для пакетов источника и ретранслятора. Каждая из величин стремится к соответствующей асимптоте с ростом вероятности P . Асимптотические значения для узла A составляют 0,28 (некооперативные системы) и 0,39 (кооперативные). Асимптотическое значение для R равно 0,57 (и для некооперативных систем, и для кооперативных), поскольку кооперация не запрещает R передачу собственного трафика. Рассмотрим область второй (кооперативной) асимптоты для источника.

Разделим рисунок 3.12 на четыре сегмента по горизонтали. Первый сегмент $[0; 0,28]$ содержит четыре монотонно возрастающие функции (поскольку канал является активно используемым, A пытается передавать

пакеты многократно и расход его энергии увеличивается). Второй сегмент $[0,28; 0,39]$ объясняет тот факт, что вероятность успешной конкуренции снижается, и источник имеет меньше шансов на успешную передачу пакета, чем в первом сегменте. В следующем сегменте $[0,39; 0,57]$ мы можем увидеть подобную ситуацию для расхода энергии в случае наличия ретранслятора. Что касается последней асимптоты для ретранслятора, система представляет собой полностью заполненный буфер. Интересно, что тренды для источника в обоих режимах и для ретранслятора в режиме без кооперации сходятся. Это результат справедливого доступа к каналу между A и R .

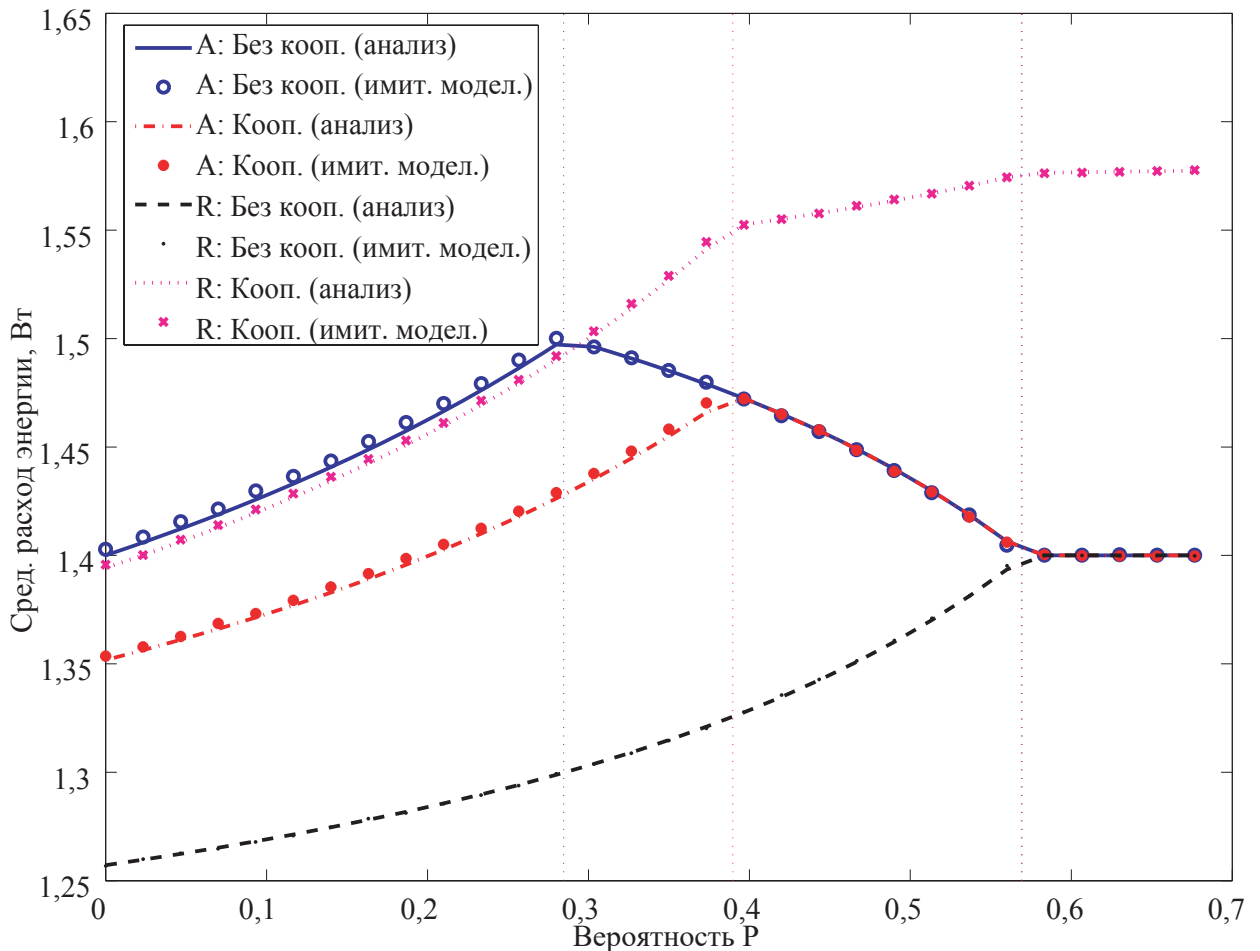


Рисунок 3.12 — Среднее количество потраченной на передачу энергии в зависимости от вероятности коллизии

Подобным образом можно рассмотреть зависимость значения средней задержки пакета от числа источников (рисунок 3.13).

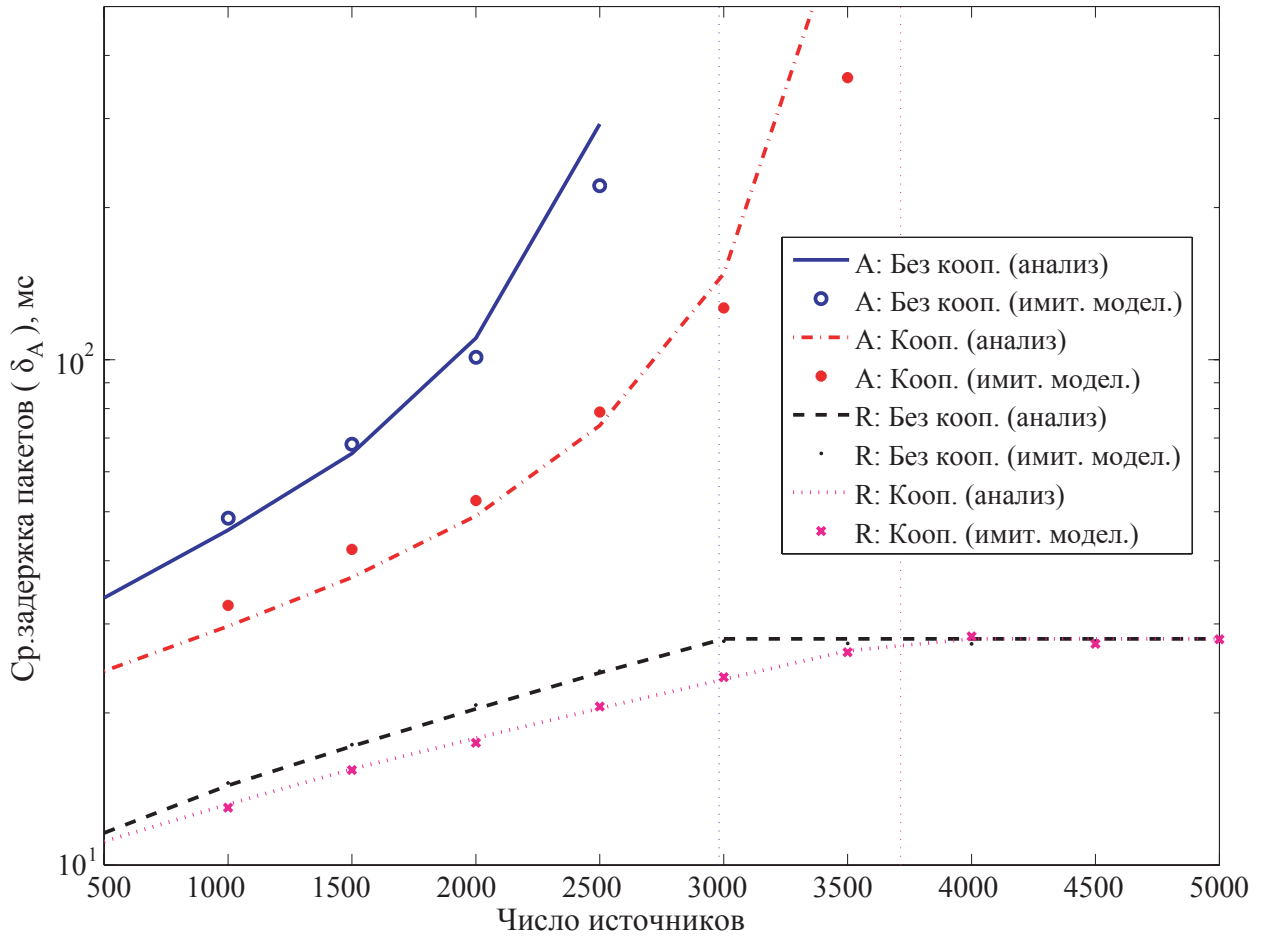


Рисунок 3.13 — Средняя задержка пакета в зависимости от числа источников

3.5.3 Облако клиентской ретрансляции

В этой части также используем систему имитационного моделирования WINTERsim для верификации полученных аналитических результатов. Используем значения величин мощности из [98]: $P_{TX} = 1,65$ В, $P_{RX} = 1,40$ В и $P_I = 1,15$ В. Предположим также, что размер каждого слота равен 5 мс.

В качестве основных параметров моделирования используем набор $p_{AB} = 0,3$, $p_{AR} = 0,7$. Вектор вероятностей успешной доставки p_{CB} , для примера, представим как случайную неубывающую линейную функцию (в реальных условиях эта функция имеет более сложную форму). Заметим,

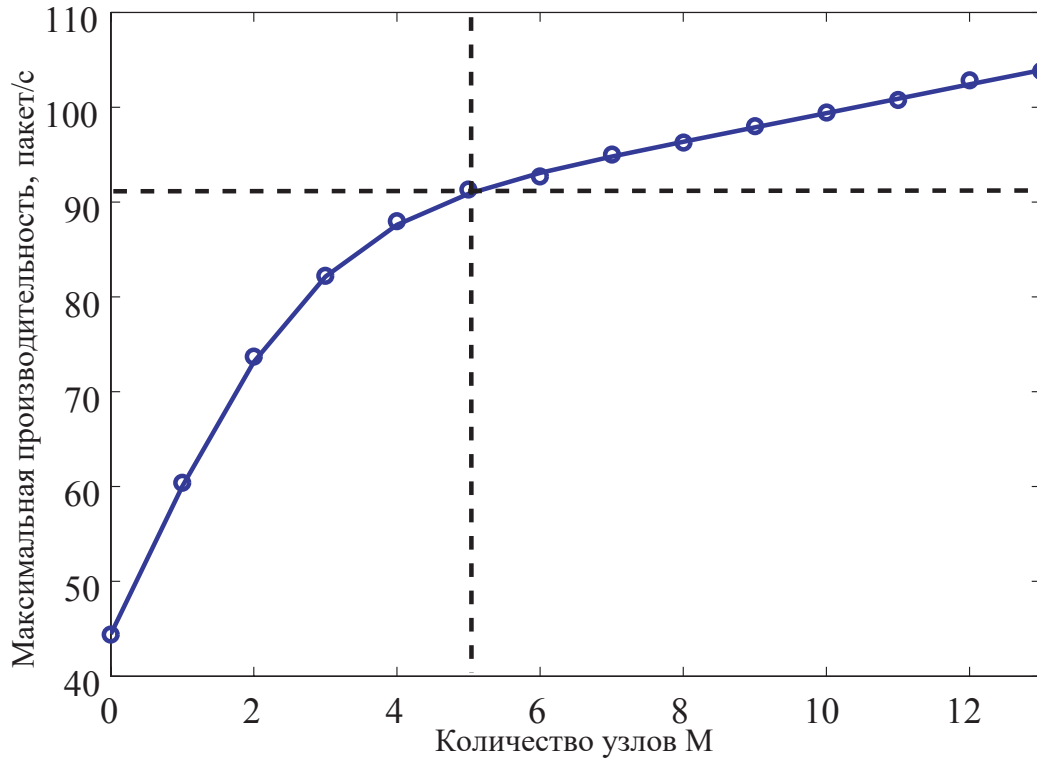


Рисунок 3.14 — Пропускная способность в зависимости от количества узлов в облаке ретрансляций

что сплошные кривые представляют аналитические результаты, а маркеры — результаты моделирования.

На рисунке 3.14 проанализируем поведение предельной пропускной способности для различного числа узлов ретрансляции в облаке. Как видим, наблюдается монотонно возрастающая функция для M . Кроме того, начиная с точки $M = 5$, зависимость является практически линейной. В дальнейшем тогда будем использовать $M = 5$.

Далее, изучим среднюю длительность задержки пакета на источнике для различного числа доступных узлов ретрансляции. С этой целью будем изменять скорость входящего потока λ_A . На рисунке 3.15 для сравнения представлены зависимости для различных значений M (с учетом асимптотических значений). Естественным образом, задержка существенно снижается с ростом числа доступных ретрансляторов.

Изучим также зависимость энергоэффективности от, например, вероятности p_{AB} . Соответствующие зависимости приведены на рисунке 3.16. Здесь в значительной степени проявляется разница для систем без коопе-

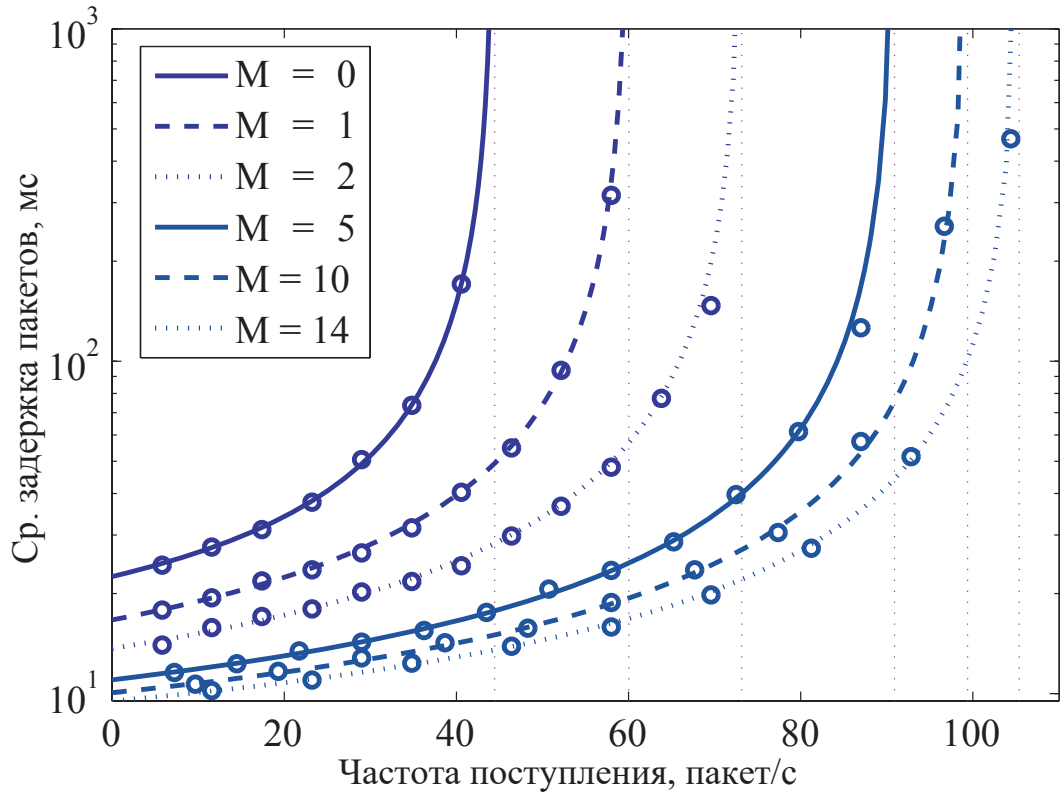


Рисунок 3.15 — Средняя задержка пакета в зависимости от интенсивности поступления пакетов λ_A для различных значений M

рации (нет кооператоров) и систем с кооперацией при $M = 1$ и $M = 5$. Как можно заметить, помощь со стороны облака ретрансляции приводит к некоторому увеличению энергопотребления, которое является ценой увеличения производительности источника.

В заключении отметим, что проведено глубокое и системное исследование кооперативной техники для беспроводных сотовых сетей. Предложен эффективный механизм клиентской ретрансляции, который существенно увеличивает пропускную способность на границе сот для стандартов IEEE 802.16m [112] и LTE-Advanced [103]. Предложенная модель клиентской кооперации приводит к существенному прогрессу в соответствующей области исследований. Ожидается, что новая модель и ее расширения станут очень важными для дальнейшего прогресса в области технологий беспроводной связи.

Предложена также простая схема клиентской ретрансляции для улучшения показателей задержки и энергетической эффективности для гранич-

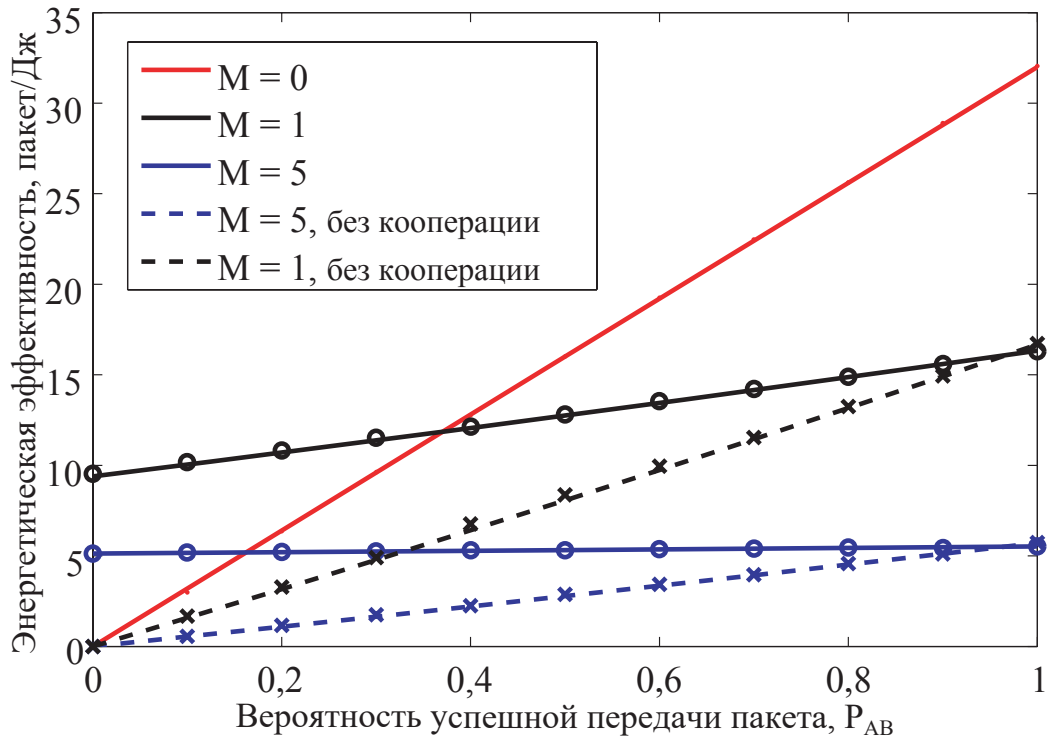


Рисунок 3.16 — Энергоэффективность в зависимости от вероятности успешной передачи пакета

ных ячеек M2M устройств в условиях плохих линий связи. Предложенный аналитический подход дает существенный выигрыш в пропускной способности, что было подтверждено имитационным моделированием. Это особенно важно для последующих исследований в области разработки стандарта. Кроме того, важным приложением полученных результатов является рынок умных счетчиков, поддерживаемый международными правительственными организациями, компаниями жилищно-коммунального хозяйства и производителями оборудования.

Наконец, рассмотрены беспроводные сотовые сети в условиях распределенного управления кооперативными коммуникациями посредством облака клиентской ретрансляции. Основная цель такой облачной структуры состоит в увеличении пропускной способности путем поддержки мобильных клиентов, находящихся на границе покрытия с плохими каналами связи. Определены основные характеристики производительности для предложенной структуры, такие как пропускная способность, задержки и энергетическая эффективность. Получены аналитические выражения в за-

мкнутом виде и верифицированы путем моделирования системного уровня. Результаты показали существенную пользу от применения облака ретрансляции, которая выражается в восстановлении пропускной способности для мобильных клиентов с деградировавшими беспроводными линиями. В будущих исследованиях предполагается изучить модели с иными реалистичными входящими потоками и предложить эффективные алгоритмы кооперации для этих случаев. Кроме того, важным представляется также оценка практических сценариев, когда облако клиентской ретрансляции способствует увеличению производительности беспроводных систем, и когда нет.

3.6 Выводы

1. Анализ кооперативных коммуникаций вкпе с анализом особенностей машино-ориентированных коммуникаций и взаимодействий машина-машина M2M показал, что до настоящего времени возможности использования клиентской ретрансляции для трафика M2M практически не изучены.
2. Исследована трехузловая модель клиентской ретрансляции, а также возможности ее облачной реализации, что дало возможность предложить такие модели для обеспечения приемлемого уровня качества обслуживания в сетях M2M.
3. Разработаны модель и метод облачной клиентской ретрансляции, учитывающие особенности трафика для устройств M2M, расположенных на границе соты, позволяющие увеличить производительность системы в 3 - 4 раза при уменьшении в несколько раз задержки пакетов.

Раздел 4. Разработка и исследование моделей и методов распределения ресурсов при использовании для выгрузки трафика технологии D2D с сетевой поддержкой

4.1 Введение

Пытаясь всегда создать некий оптимальный баланс между покрытием и емкостью в архитектуре сотовой сети в течение нескольких десятилетий, начиная со скандинавских сетей мобильной связи (NMT) и сетей стандарта GSM, доминируют сотовый технологический ландшафт, обеспечивающий беспрецедентную функциональность мобильной телефонии. Поскольку базовые станции всегда имели связь с *высоко централизованным ядром сети*, сотовая архитектура была полностью совместима с существующими телефонными сетями общего пользования (ТфОП), гарантируя при этом покрытие почти всего земного шара и достаточно высокий уровень качества обслуживания.

На всем своем 40-летнем существовании сотовые сети эволюционировали в направлении как можно более высокой спектральной эффективности, но при этом не смогли обеспечить масштабирование емкости сети до уровней, которые соответствовали бы постоянно растущему потребительскому спросу. Это произошло, прежде всего, потому, что емкость в сотовой системе распределяется между всеми пользователями в зоне покрытия, а при росте потребительского спроса эта область не изменяется. Было установлено, что увеличение плотности сот на единицу области сотовой сети может устранить проблему [50]. Поэтому, существующие сотовые сети эволюционировали в сторону развертывания многочисленных малых сот, также называемых пико (с покрытием около 100 метров) [139] и фемтосот (с WiFi-подобным покрытием) [140]. Кроме того, кооперативные сети и распределенные антенные системы были введены в сеть для более гибкого ее построения. С учетом этого тренда, известного как *уплотнение сети*, большое число инфраструктурных сетевых узлов стало доступно непосред-

ственно пользователям, увеличивая число сот в расчете на пользователя и, таким образом, расширяя емкость сети [141].

Однако, при увеличении плотности сети с одновременным ростом пропускной способности на пользователя увеличиваются и расходы на инфраструктуру, располагаемую теперь в непосредственной близости от пользователя. В то же время еще не достигнут консенсус между существующими сетевыми технологиями и технологиями сетей связи пятого поколения 5G, при этом плотность сетевой инфраструктуры не может возрастать бесконечно. Поэтому, среди множества предложений для сетей 5G появились такие, в которых рассматривалась не сотовая структура. Это, в первую очередь, прямые взаимодействия устройство-устройство (device-to-device, D2D), одно из самых перспективных направлений развития сетей связи в последние годы. Рассматриваемая изначально под различными брендами, такими как ad-hoc и ячеистые сети [142], кооперативные сети, клиентские сети с ретрансляцией, а также даже и когнитивное радио, D2D технология стала одной из фундаментальных задач для сотовой инфраструктуры. Особенно важным является то, что эта технология предполагает взаимодействие клиентов сотовой сети без использования централизованной инфраструктуры, уменьшая нагрузку на нее.

Теоретически D2D предоставляет для связи самые короткие линии, лучшую энергетическую эффективность и даже более естественные потоки трафика. Вследствие этого за прошедшие несколько лет D2D взаимодействия привлекли существенное внимание как производителей оборудования, так и исследовательских организаций, почти как панацея от всех проблем сотовых сетей. На практике, естественно, D2D приносит ряд своих проблем, но потенциальное использование *ресурсов нелицензируемого спектра* выглядит исключительно прибыльным. Как будет установлено далее, это направление — один из самых простых способов повысить пропускную способность сотовых систем без необходимости переосмысления всей концепции.

Этот тезис подчеркивает задачу внедрения в сотовые сети альтернативной им D2D технологии, которая могла бы в то же время использовать поддержку исторически существующих конструкций в современных сото-

вых сетях. Немного отвлекаясь от академических исследований в дополнение к разработке и оценке предлагаемых решений, целесообразно также поддерживать практические действия по прототипированию и возможные испытания.

4.2 Возможности D2D в сотовых сетях

В 2010 — 2011 годах D2D взаимодействия не рассматривались близко к проблемам сотовых систем. В те годы основная идея базировалась на том, что *покрытие* в сотовых сетях может быть улучшено с введением систем *клиентской ретрансляции* (client relay, CR), рассмотренной достаточно подробно ранее в данной диссертации, а также в фундаментальных работах [117; 143; 144], [125; 145; 146].

Далее, эти идеи были расширены на передачу основных данных [147] между терминалами. Появление такой технологии как FlashLinQ уже предполагало, что сотовые сети могут использовать одноранговые линии, что теперь хорошо известно как внутриволновое взаимодействие D2D. Прототип технологии FlashLinQ рассмотрен в [148; 149] и проанализирован в [150]. Это представляло собой распределенную технологию взаимодействия D2D в лицензируемом спектре для синхронизации и ранжирования в P2P сетях. Однако, предложенный как решение, FlashLinQ вызвал и множество вопросов, в основном связанных с устранением помех и контролем мощности, которые до сих пор еще не решены.

Лицензируемый спектр продолжает оставаться очень ограниченным и дорогим, в то же время есть попытки лицензировать еще больше спектра, поскольку того, который есть уже просто недостаточно для удовлетворения существующего спроса [66]. Поэтому, использование любого типа внутриволнового D2D для целей передачи данных является исключительно сложной задачей, поскольку ресурсы не могут быть окончательно зарезервированы только для D2D, и, следовательно, D2D должен совместно использовать их с существующим сотовым доступом.

Расширенное управление интерференцией с контролем доступа и мощности [151] требуется для поддержки нескольких соединений D2D в одном и том же покрытии. Исследовательские работы (например, [152; 153] или [154]) предлагают способ устранения интерференции, используя информацию о состоянии канала в реальном времени (CSI), местоположении пользователя, буфере состояний и многих других. При этом схема управления интерференцией должна решить, что делать с каждым соединением D2D:

- распределять ресурсы лицензированной полосы со стандартными сотовыми передачами (между пользователями и базовыми станциями) и испытывать интерференцию;
- использовать выделенные ресурсы и стараться обеспечить их использование или
- оставаться в сети сотовой инфраструктуры.

Этот комплексный процесс, известный так же, как *выбор режима*, привлек большое внимание исследователей, сфокусировавших свое внимание на широком круге оптимизационных задач, для таких показателей как отношение сигнал/интерференция плюс шум (SINR) [155; 156], пропускная способность [157], энергетическая эффективность [158], задержки данных [159], справедливое распределение ресурсов и вероятность отказа [160]. Большинство из этих исследований интегрировали D2D технологию в лицензируемом спектре в технологию систем длительной эволюции LTE 3GPP [161; 162]. Однако некоторые статьи были адресованы к существующим сотовым сетям как таковым [163].

В конечном счете, нехватка ресурсов влияет на лицензионный диапазон D2D так же, как и на сами сотовые сети, которые эта технология пытается улучшить. Поэтому, существует мотивация для улучшения производительности каналов D2D. Текущие работы включают в себя множественные антенны MIMO [152; 164], сетевое кодирование [165], успешное подавление интерференции [166] и кеширование реле [167].

Конечно, не только лицензированные диапазоны доступны. Действительно, существенная часть спектра по-прежнему не лицензирована, достаточно посмотреть на диапазоны 2,4 и 5 ГГц для WLAN. Вследствие сказан-

ного исследования в этом направлении не менее важны. Большое будущее использования D2D в нелицензированном спектре подтверждено работами в этом направлении 3GPP [168], которые известны как LTE-у (LTE для нелицензированного спектра) и LAA (поддержка доступа в лицензируемом спектре) [169].

С другой стороны, оператор не может требовать эксклюзивного использования какого-либо нелицензионного спектра. В результате интерференция становится неконтролируемой, что требует надежного контроля доступа к среде, который может справиться с этой проблемой. Например, были разработаны технологии Bluetooth и WiFi с учетом этого явления, которые становятся все более популярными в беспроводных персональных и локальных сетях (WPAN/WLAN). Основанный на стандартах IEEE 802.11, Wi-Fi в настоящее время является преобладающим решением (как с поддержкой, так и без поддержки точек доступа к инфраструктуре) для подключения пользовательских устройств [170].

Однако, поскольку WiFi не имеет встроенного контроля доступа (он позволяет каждому устройству выбрать канал на основе конкурентных решений «более или менее» одинаково), эту технологию часто критикуют за то, что она не удовлетворяет требованиям по качеству обслуживания QoS [171]. С другой стороны, WiFi обеспечивает более высокие скорости передачи данных и энергоэффективность, чем любая из сотовых технологий [170]. В принципе *при наличии соответствующего интеллекта во всех устройствах* для ad-hoc соединения WiFi можно добиться и стабильных результатов работы [172]. Однако, увеличивая количество соединений D2D и персональных сетей, в том числе и таких, как WiFi Direct (WFD) технология [173], надо отдавать себе отчет в том, что в конечном итоге какой-то ресурс управления будет необходим даже для нелицензионных диапазонов D2D как, например, это предлагается в [58].

Сотовые сети могут обеспечить ряд важных услуг для систем не лицензионного спектра, например, интеллектуальное управление радио ресурсами, а также столь необходимую анонимность при обслуживании. Если пользовательские устройства постоянно связаны с сотовой сетью, сотовая сеть может предложить резервное соединение в случае, если для нели-

цензионной полосы QoS оказывается неудовлетворительным. Кроме того, при поддержке со стороны сотовой сети аутентификация устройств и безопасность линии D2D могут быть значительно улучшены, что, например, делает любой вид атаки «человек посередине» практически невозможным. Другими словами, *степень* поддержки сотовой сети для D2D-соединений в нелицензированных полосах спектра представляется очень перспективным подходом.

Потенциальные приложения технологии D2D в сотовых сетях активно обсуждаются вплоть до настоящего времени [60]. Рассматриваются и услуги по передаче речи (вызовы между близлежащими пользователями) и иные приложения для близко находящихся людей и/или машин, например, поиск контента, игры, задачи, относящиеся к инфраструктуре, управление домашними роботами, общественная безопасность [174]. Более широко, в терминах устройств помимо мобильных телефонов и планшетов, это, конечно, приложения для машин (машинно-ориентированные коммуникации или MTC/M2M). В основном, D2D с сетевой поддержкой на равных используются как ориентированных на человека приложений, так и для приложений и услуг MTC/M2M [175].

4.3 Способ решения системы треугольника

4.3.1 Описание проблематики треугольника

Рассмотрим сценарий, представленный на рисунке 4.1. В данном упрощенном сценарии имеется два передающих устройства, Tx_1 и Tx_0 , разнесенные на определенную дистанцию U друг от друга, являющуюся случайной величиной (СВ) с функцией плотности распределения $f_U(u)$. В свою очередь, данное распределение может быть задано природой случайного расположения передатчиков на плоскости. Описанным выше передатчикам соответствуют назначенные им примыные устройства, Rx_0 и Rx_1 ,

которые расположены равномерно внутри кругов фиксированного радиуса с центром в точке расположения их соответствующего передатчика. Далее, соответствие между приемным и передающим устройствами будем называть кластером, при этом случайное расстояние L между парой устройств задается функцией плотности $f_L(l)$. В свою очередь, предполагается, что угол $\gamma = \angle Tx_1Tx_0Rx_0$ распределен согласно равномерному распределению, хотя все приведенные ниже утверждения будут справедливы и для любого другого распределения. Стоит заметить, что все три случайные величины U , L , и γ предполагаются взаимно независимыми.

Таким образом, задачей является нахождение распределения расстояния D между интерферирующим передатчиком Tx_1 и интересующим нас приемником Rx_0 . На основе полученного распределения можно будет определить производные характеристики качества канала связи между интересующими нас устройствами, такие как отношение сигнала к интерференции [31; 176].

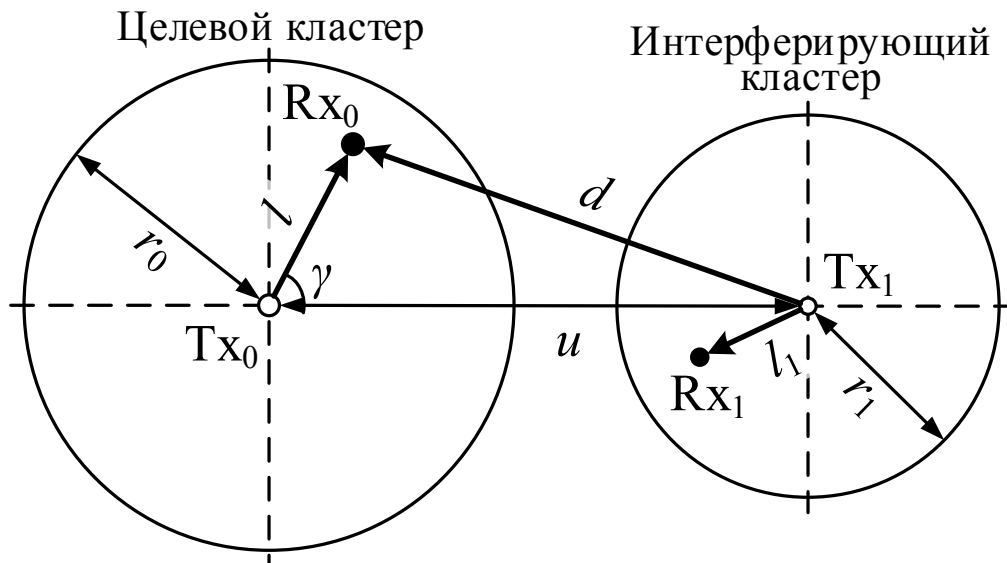


Рисунок 4.1 — Модель рассматриваемой системы

4.3.2 Специальный случай

В работе [177] рассматривается набор особых случаев распределений расстояний для точки, равномерно распределенной внутри окружности. В частности, как показано в [178], расстояние от фиксированной точки вне окружности до случайной точки внутри нее определяется путем геометрического представления вероятности, т.е. как отношение определенной площади ко всей площади. В этом случае решение приведенной выше задачи сводится к нахождению площади пересечения двух дисков, отнормированное на πr^2 .

Как было показано на странице 187 работы [177], для заданного радиуса r плотность распределения расстояния между случайной точкой внутри окружности и фиксированной точкой за ее пределами на расстоянии $r_1 + r$ от ее центра определяется формулой:

$$g(\rho) = \frac{2\rho}{\pi r^2} \arccos \left(\frac{r_1}{\rho} + \frac{\rho^2 - r_1^2}{2\rho(r + r_1)} \right), \quad (4.1)$$

где $r_1 \leq \rho \leq r_1 + 2r$.

Применив этот результат, можно получить распределение расстояния между двумя точками с учетом того, что r_1 является СВ с плотностью распределения $f_{r_1}(x)$, $x \geq r$:

$$g_2(\rho) = \int_r^\infty \frac{2\rho}{\pi r^2} \arccos \left(\frac{x}{\rho} + \frac{\rho^2 - x^2}{2\rho(r + x)} \right) f_{r_1}(x) dx. \quad (4.2)$$

К сожалению, (4.1) крайне тяжело расширить на случай произвольного распределения расстояния между центром окружности и точки распределенной внутри нее. По этой причине далее приводится общий метод нахождения распределения расстояния между точками, частным случаем которого является формула (4.1).

4.3.3 Общий случай

Поставленная задача, заданная в общих предположениях, где рассматриваемых передающие и принимающие устройства фактически сводится к нахождению распределения случайной длины одной из сторон случайного треугольника, заданного распределениями длин двух других сторон и угла между ними [179].

При решении стоит учитывать, что угол γ распределен равномерно в пределах между 0 и π из-за симметричности сценария.

Общий подход к решению этой задачи основан на последовательном применении метода преобразования случайных величин, в подробностях описанного в [180].

Введем новые обозначения:

$$\xi_1 = L, \quad \xi_2 = U, \quad \xi_3 = \gamma, \quad \eta_1 = D, \quad (4.3)$$

с соответствующими им плотностями:

$$w_{\xi_1, \xi_2, \xi_3}(\mathbf{x}) = f_{L, U, \gamma}(\mathbf{x}), \quad W_{\eta_1}(y_1) = f_D(y_1), \quad (4.4)$$

где $f_{L, U, \gamma}(x_1, x_2, x_3)$ — это совместное распределение СВ L , U и γ , а $f_D(y_1)$ — искомая плотность распределения.

В связи с тем, что распределения величин U , D и γ взаимно независимы, получаем:

$$f_{L, U, \gamma}(x_1, x_2, x_3) = f_L(x_1) f_U(x_2) f_\gamma(x_3). \quad (4.5)$$

Применив теорему косинусов, получим следующее функциональное взаимоотношение:

$$\eta_1 = f(\xi_1, \xi_2, \xi_3) = \sqrt{\xi_1^2 + \xi_2^2 - 2\xi_1\xi_2 \cos \xi_3}. \quad (4.6)$$

Введя две новые вспомогательные переменные $\eta_2 = \xi_2$ и $\eta_3 = \xi_3$, запишем:

$$\begin{cases} y_1 = f(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2 - 2x_1x_2 \cos x_3}, \\ y_2 = x_2, \\ y_3 = x_3. \end{cases} \quad (4.7)$$

Тогда, согласно [180; 181], искомое распределение D определяется выражением:

$$f_D(y_1) = \sum_{j=1,2} \int_{\mathbf{Y}_j} w_{\xi_1, \xi_2, \xi_3}(\phi_j(y_1, y_2, y_3), y_2, y_3) \times \left| \frac{\partial \phi_j(y_1, y_2, y_3)}{\partial y_2} \right| dy_3, \quad (4.8)$$

где ϕ_j — это обратная функция правой части выражения для y_1 из (4.7) с двумя ветвями преобразования:

$$\begin{cases} \phi_1(y_1, y_2, y_3) = y_2 \cos y_3 + \sqrt{y_1^2 - y_2^2 + y_2^2 \cos y_3^2}, \\ \phi_2(y_1, y_2, y_3) = y_2 \cos y_3 - \sqrt{y_1^2 - y_2^2 + y_2^2 \cos y_3^2}. \end{cases} \quad (4.9)$$

Область интегрирования \mathbf{Y}_j в (4.8) получается путем решения системы неравенств:

$$\begin{cases} \phi_j(y_1, y_2, y_3) \geq 0, & i = 1, 2, \\ y_1 \geq 0, \\ y_2 \geq 0, \\ 0 \leq y_3 \leq \pi. \end{cases} \quad (4.10)$$

Решением данной системы является набор условий:

$$\begin{aligned} \mathbf{Y}_1^1 &= \begin{cases} 0 \leq y_2 \leq y_1, \\ 0 \leq y_3 \leq \pi, \end{cases} \\ \mathbf{Y}_1^2 &= \begin{cases} y_2 \geq y_1, \\ 0 \leq y_3 \leq \arccos \sqrt{\frac{y_2^2 - y_1^2}{y_2^2}}, \end{cases} \\ \mathbf{Y}_2 &= \begin{cases} y_2 \geq y_1, \\ 0 \leq y_3 \leq \arccos \sqrt{\frac{y_2^2 - y_1^2}{y_2^2}}, \end{cases} \end{aligned} \quad (4.11)$$

где $\mathbf{Y}_1 = \mathbf{Y}_1^1 \cup \mathbf{Y}_1^2$.

Таким образом, финальный вид выражения для плотности распределения длины D записывается в виде:

$$\begin{aligned} f_D(y_1) &= \sum_{j=1,2} \int_{\mathbf{Y}_j} f_L(\phi_j(y_1, y_2, y_3)) f_U(y_2) \frac{1}{\pi} \times \\ &\quad \frac{y_1}{\sqrt{y_1^2 - y_2^2 + y_2^2 \cos^2 y_3}} dy_3 dy_2. \end{aligned} \quad (4.12)$$

Стоит заметить, что уравнение (4.12) далеко не всегда можно представить в явном виде, особенно для неэлементарных изначальных распределений. Но всегда можно применить методы численного интегрирования для получения результата.

Однако, величины D и L взаимозависимы, так что для получения производных характеристик расстояния, например SIR, необходимо получить их совместное распределение. Применяя тот же самый метод, что и для получения функции плотности распределения величины D , совместное распределение имеет вид

$$\begin{aligned} W_{L,D}(y_1, y_2) &= \\ &= \sum_{j=1,2} \int_{\mathbf{Y}_{3,j}} \frac{f_L(y_1) f_U(\phi_j(y_1, y_2, y_3)) f_\gamma(y_3) y_2}{\sqrt{y_2^2 - y_1^2 + y_1^2 \cos^2(y_3)}} dy_3, \end{aligned} \quad (4.13)$$

где ϕ_j — это обратное преобразование функции (4.7), но по отношению к y_2 :

$$\begin{cases} \phi_1(y_1, y_2, y_3) = y_1 \cos y_3 + \sqrt{y_2^2 - y_1^2 + y_1^2 \cos y_3^2}, \\ \phi_2(y_1, y_2, y_3) = y_1 \cos y_3 - \sqrt{y_2^2 - y_1^2 + y_1^2 \cos y_3^2}, \end{cases} \quad (4.14)$$

а $\mathbf{Y}_{3,j}$ — пределы интегрирования для y_3 :

$$\begin{aligned} \mathbf{Y}_{3,1} &= \mathbf{Y}_{3,1}^1 \cup \mathbf{Y}_{3,1}^2 \\ \mathbf{Y}_{3,1}^1 &= \begin{cases} 0 \leq y_2 \leq y_1, \\ 0 \leq y_3 \leq \frac{1}{2} \arccos \frac{y_1^2 - 2y_2^2}{y_1^2}, \end{cases} \\ \mathbf{Y}_{3,1}^2 &= \begin{cases} y_2 \geq y_1, \\ 0 \leq y_3 \leq \pi, \end{cases} \\ \mathbf{Y}_{3,2} &= \begin{cases} 0 \leq y_2 \leq y_1, \\ 0 \leq y_3 \leq \frac{1}{2} \arccos \frac{y_1^2 - 2y_2^2}{y_1^2}. \end{cases} \end{aligned} \quad (4.15)$$

4.3.4 Численный пример

В качестве примера применения предложенной модели приведем типичный случай прямого взаимодействия двух устройств (Device-to-Device, D2D). Так расположение передающих устройств определяются согласно пространственному распределению Пуассона с интенсивностью λ . Соответствующие передающим устройствам приемники равномерно расположены внутри круга единичного радиуса с центром в точке расположения передатчика. Интерферирующий же передатчик расположен за пределами указанного круга, при этом являясь ближайшей точкой пространственного Пуассоновского процесса к рассматриваемому передатчику.

Соответствующие распределения расстояний задаются формулами [182]:

$$\begin{aligned} f_L(l) &= \frac{2l}{r_0^2}, & 0 \leq l \leq r_0, \\ f_U(u) &= 2\pi\lambda u e^{\pi\lambda r_0^2 - \pi\lambda u^2}, & u \geq r_0, \lambda > 0. \end{aligned} \quad (4.16)$$

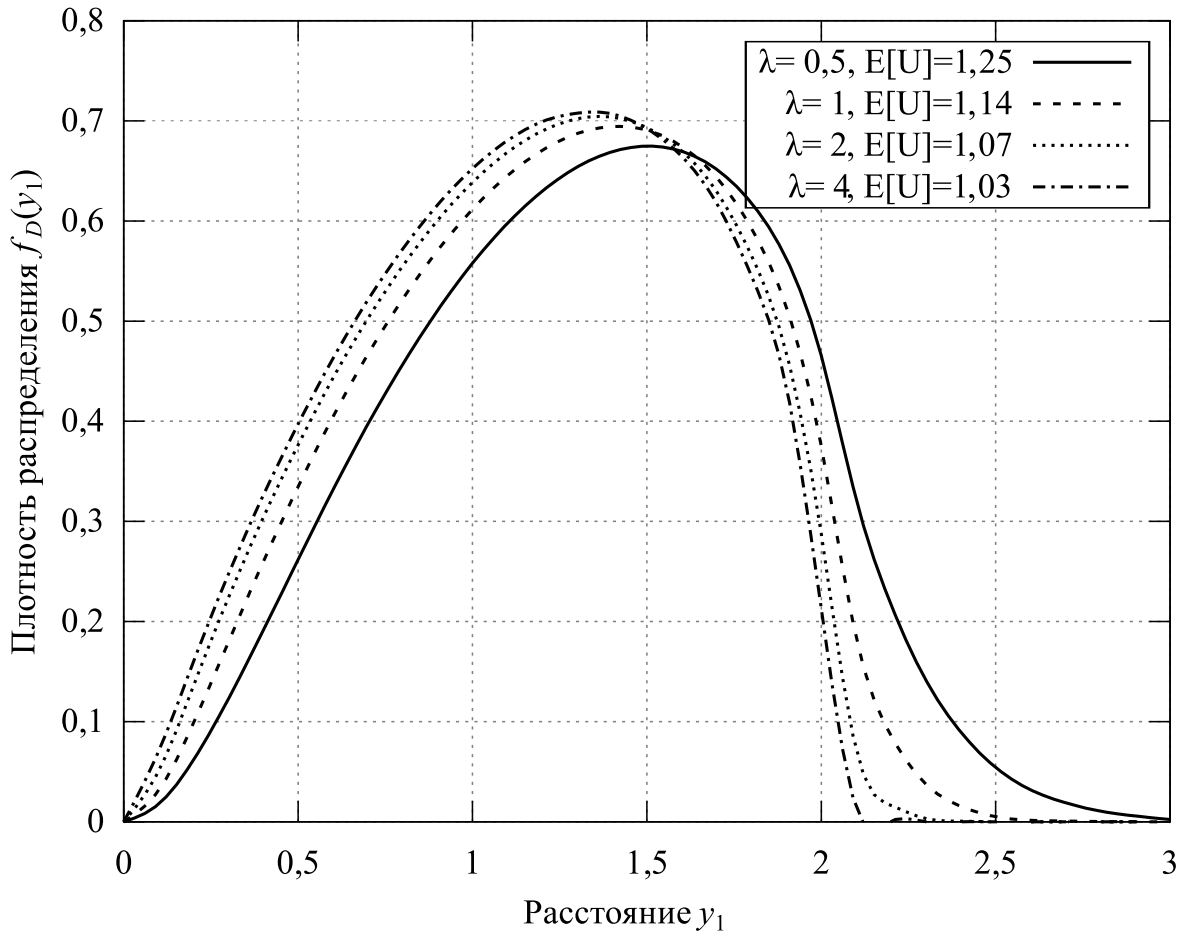


Рисунок 4.2 — Распределение показателей интерференции

Результатирующая форма полученного распределения (4.12) в зависимости от параметра плотности передающих устройств на плоскости λ показана на рисунке 4.2. Расстояние на данном графике приводится в условных единицах, нормированных на радиус круга r_0 , тем самым обобщая полученный результат.

Далее рассмотрим применимость модели для анализа характеристики SIR в применении к сценарию прямого взаимодействия двух устройств в условии ограничения по интерференции. В качестве реалистичных пара-

метров модели были взяты: интересующий нас приемник расположен внутри круга радиуса $r_0 = 30$ метров, интерферирующий передатчик находится за его пределами, параметры модели распространения сигнала заданы соответственно $\alpha = 2,61116$ и $c = 0,000201944$, а мощности передатчиков отнормированы к единице [183].

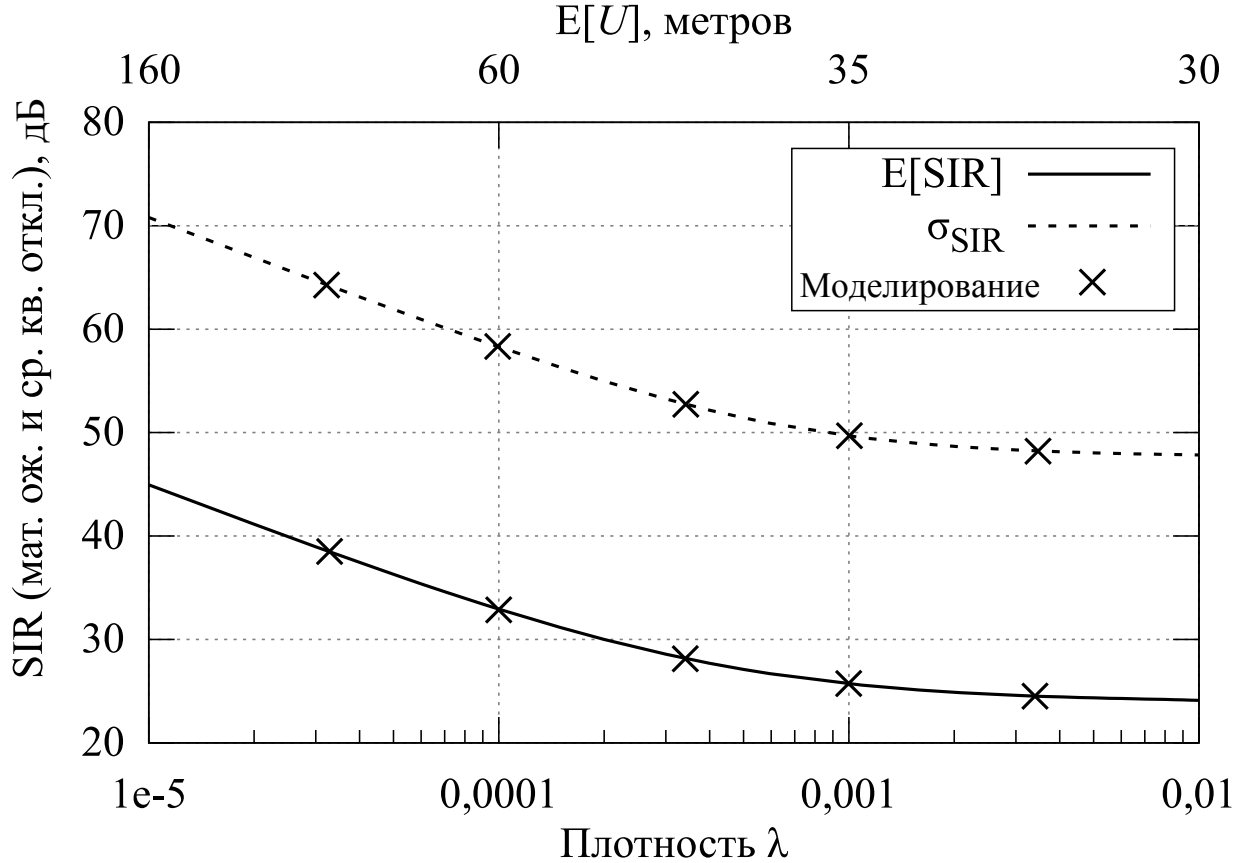


Рисунок 4.3 — Численные показатели SIR

С учетом введенных предположений, итоговая характеристика рассчитывается по формуле $SIR = (D/L)^{2,61116}$. Применив формулу (4.13), n -ый момент SIR можно рассчитать согласно формуле:

$$E[SIR^n] = \int_{0 \leq y_1 \leq r_0} \int_{y_2 \geq 0} \left(\frac{y_2}{y_1}\right)^{n\alpha} W_{L,D}(y_1, y_2) dy_2 dy_1. \quad (4.17)$$

На рисунке 4.3 приведены математическое ожидание и среднеквадратическое отклонение, $\sigma_{SIR} = \sqrt{E[SIR^2] - E[SIR]^2}$, величины SIR от параметра распределения дистанции до интерферирующего передатчи-

ка. С целью проверки полученных результатов, было разработано средство иммитационного моделирования описанного сценария.

4.4 Архитектурные вопросы построения динамической D2D

С концептуальных позиций должно быть исследовано взаимодействие двух независимых сетей: существующей сотовой сети и сети D2D, которая не использует сотовый принцип построения сети, а также функционирует в нелицензируемом диапазоне. При этом некоторые дополнительные услуги станут доступными для сотовой сети такие, как P2P, управление соединениями и т.д. При этом сама сеть D2D исходно отвечает за передачу данных P2P и размещение ресурсов. Такое разделение ответственности достаточно обосновано, как это будет показано далее. Кроме того, будут рассмотрены вопросы выгрузки трафика на уровне приложений, а также необходимые решения для обеспечения безопасности и соединений P2P. В результате исследований, проведенных в данном разделе, в конечном итоге будет предложено завершённое архитектурное решение для сети связи при внедрении технологии D2D.

Далее сосредоточимся на концепции D2D с сетевой поддержкой D2D и превращение этой концепции в полнофункциональное решение, позволяющее обеспечить динамическое управление локальными радио линиями в дополнение к сотовой сети. Стремясь обеспечить совместимость предложения с иными решениями, будем использовать, в основном, стандартные интерфейсы и протоколы, облегчая интеграцию.

Будем избегать разработки или модификации любых технологий канального уровня, так как это может потребовать замены оборудования, которое уже находится в эксплуатации, что экономически нецелесообразно в обозримой перспективе. Аналогично не будем нарушать сотовую архитектуру, а модернизировать ее только путем эволюции.

Кроме того, ограничим область разработки тем, что новая система будет оставаться в *основном сотовой* с добавлением только нескольких ком-

понентов. Помимо соответствия требованиям 3GPP, такой подход облегчает процесс планирования, поскольку в сотовых системах уже существует большой набор специальных функций, таких как хэндовер, аутентификация, безопасность, биллинг, позиционирование и многие другие. Предположим также, что все приложения, которые будут использовать новые функции должны быть легко адаптированы к их текущим версиям и должны быть совместимы с существующей экосистемой.

Для обеспечения совместимости с планируемыми сетями сфокусируем внимание на LTE Release 10 и выше в качестве базовой сотовой сети. Где это возможно, будем рассматривать проблемы с точки зрения *восходящих линий*, так как это как правило то, что больше всего характеризует производительность системы. Кроме того, в предлагаемых решениях будет использоваться только оборудование, соответствующее современным смартфонам.

Далее также рассмотрим *набор технологических средств* для конвергентной сотовой/D2D-сети, где мульти радио устройства имеют возможность использовать все свои интерфейсы для одноранговой связи для таких целей, как обмен файлами, выгрузка сотового трафика и т.п. Разработанный в диссертационной работе набор решений достаточно широк, чтобы охватить сценарии выгрузки, алгоритмы управления и администрирования, пользовательские настройки и любые потенциальные будущие беспроводные технологии (например, IEEE 802.11 ad/ay и LTE-нелицензированный).

4.4.1 Интеграция архитектуры

Как уже было обсуждено ранее, для операторов желательной является выгрузка трафика в сеть D2D нелицензируемого спектра как конкурентное решение с LTE и WiFi линиями для мульти радиоустройств. Однако ни WiFi, ни WiFi Direct не имеют быстрого и эффективного способа обнаружения устройств/услуг, требующих этого, и не видно простого пу-

ти внедрения D2D взаимодействий вследствие очень коротких расстояний между устройствами. Наконец, анонимность для P2P соединений остается открытой для решений D2D в нелицензируемом спектре. С целью преодоления этих ограничений в диссертационной работе предлагается новая архитектура для внедрения поддерживаемых сетью D2D решений.

Архитектура уровня обслуживания. Для начала идентифицируем какие именно услуги должны быть обеспечены для конечного пользователя. Естественно, что очень сложно гарантировать какие-то хорошие условия для D2D линий и качество обслуживания для таких линий может существенно отличаться в различные моменты времени и в связи с перемещением пользователей. Поэтому, услуги, толерантные к задержкам, такие, как распределенное кэширование и кооперативные загрузки (мультикаст) могут рассматриваться как первые кандидаты на выгрузку трафика. Такие услуги, однако, достаточно редко необходимы для мобильных пользователей, наиболее интересными для мобильных пользователей представляются услуги по запросу, требующие стабильных линий высокого качества и, таким образом, никогда не рассматривавшиеся для случая использования D2D технологий. Однако, если устройства находятся близко друг от друга и можно предсказать достаточно стабильный уровень качества связывающей их линии, многие запрашиваемые P2P услуги становятся возможными. На коротком промежутке времени приложения, подобные потоковому видео, играм со многими участниками и много других того же плана могут быть предоставлены посредством D2D линий. Однако, для того, чтобы сделать эти услуги коммерчески привлекательными, определенные гарантии необходимы, а также способы определения того, как выполнены предварительные условия для использования D2D.

Поэтому, критически важные требования разработке архитектуры сети D2D состоят в том, чтобы дать клиентам знание о том, когда можно устанавливать D2D соединения, с кем, какой контент доступен, и что делать, если почему-то нет возможности поддерживать QoS. Для этого, в частности, должны быть предусмотрена поддержка следующих функций:

- Идентификация цели взаимодействия (хранение и поиск файловых и потоковых метаданных, идентификация участника игры и т.д.)
- Поиск ближайших партнеров, проявляющих интерес к контенту, не подверженный риску по доступности контента широкой публике
- Аутентификация и авторизация по отношению к линии для целей занятия и передачи контента
- Поддержка *надежного* возврата к сотовому соединению при отказе D2D линии

Технически нахождение пары для подходящего контента не является сложной задачей вследствие широкого распространения сегодня облаков и социальных сетей. Важнейшая проблема состоит в том, что соединение, которое еще не установлено, не может быть представлено или управляемо существующим образом: нет ассоциированных интерфейсов с ним, нет IP адресов и т.д. Естественным выводом в такой ситуации является необходимость создания нового объекта, специально разработанного для управления линией и принятия решений в реальном времени, основанных на позиционировании и доступности радио ресурсов. Поскольку такая информация собирается только оператором в системе управления сетью, этот объект должен быть частью операторской сети.

С другой стороны, отслеживание контента и нужды систем безопасности масштабируемы и разнообразны, а программные интерфейсы для приложений (API), доступны только провайдеру услуг. Тогда, и управление связностью, и компоненты социальной сети должны вступить в кооперацию между собой для предоставления комплексных услуг пользователю. Это наиболее естественный интерфейс между двумя казалось бы несовместимыми технологиями: специфичными возможностями управления линией (управляемого по запросу сервером D2D как часть транспортной сети) и специфическими возможностями отслеживания контента (управляемого по запросу сервера приложений). Относящиеся к безопасности проблемы в этом случае дублируются для обеих из них. Предложенное в диссертации решение проиллюстрировано на рисунке 4.4 и работает следующим образом:

1. Каждое пользовательское оборудование использует возможности уровня приложений для аутентификации самого себя с сервером приложений (например, Facebook). Это позволяет обеспечить взаимодействие с контентом, что авторизуется третьей стороной доступа. Это устраняет требования по обработке метаданных, а также предоставляет возможности авторизации со стороны социальных сетей.
2. Пользовательское оборудование UE разрешает серверу D2D своего оператора представлять его на сервере приложений при D2D соединениях. При этом сервер D2D никогда не получает доступ к контенту или его метаданным, просто осуществляется идентификация пользователя, чтобы убедиться, что устройство действительно принадлежит владельцу заявленных учетных данных уровня приложения. Сервер D2D, таким образом, позволяет присваивать имена прикладного уровня фактическим физическим устройствам, в том числе и в случае, когда имеется несколько устройств, принадлежащих одному и тому же лицу.
3. Пользовательское устройство UE может публиковать или искать линии для контента в сервере приложений, и эти линии будут ссылаться на конкретный контент конкретного пользователя (но не устройства). Линии могут быть постоянными, ограниченными по времени или одноразовыми в зависимости от приложения.
4. Пользователь UE запрашивает сервер D2D для облегчения создания соединения D2D, таким образом, чтобы обеспечить линию на уровне приложений, а сервер приложений преобразует этот запрос в фактическое соединение на уровне линии и обеспечивает IP-адрес, к которому могут быть привязаны сокеты.
5. Наконец, может начаться обмен данными. Заметим, что сервер приложений не участвует в этом процессе и не отслеживает обмен R2P непосредственно, будучи только уверенным в том, что линии следуют модели безопасности, не принимая участия в их микроуправлении. Сервер D2D в то же время может осуществлять мо-

ниторинг и регулировать свойства D2D линии при необходимости в зависимости его функций.

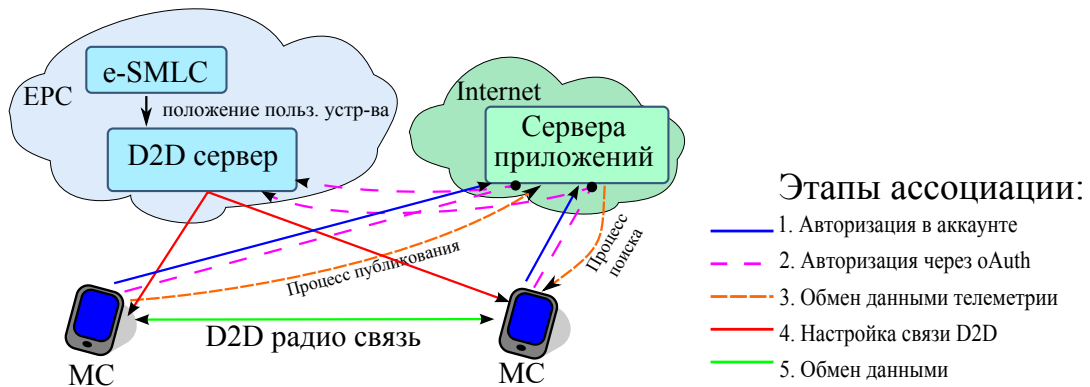


Рисунок 4.4 — Поддержка в создании D2D линии

Предложенная схема позволяет сохранить модели безопасности и разрешений, используемых услугами приложений. Например, если доступ к контенту ограничен для определенной группы пользователей, то только пользователи, которые имеют соответствующие D2D линии, получают доступ к этому контенту. Дополнительной целью является гарантия того, что пользователи не будут отслеживаться после того, как их D2D сессия будет завершена. Эти цели представляют собой отдельные исследовательские направления, и исследования в области безопасности не являются частью исследования представленного в данной диссертации. Поэтому, предложенная схема поддерживает хорошо известные решения для обеспечения целей безопасности, но при этом не утверждается, что предложенная здесь схема является эффективной.

Далее рассмотрим кратко те особенности предложенной схемы, которые делают ее приемлемым решением с точки зрения безопасности в поддерживаемой сотовой сетью D2D.

- Каждое соединение начинается с проверки контента линии. Потенциальным злоумышленникам необходимо будет найти компромисс с сервером приложений для получения такой линии. Поэтому, любая атака на систему по существу блокируется одноразовым паролем, распространяемым через сервер приложений через существующие защищенные каналы.

- Ни одно из пользовательских устройств UE не должно широкополосным образом транслировать никакую информацию об обнаружении или прослушивать запросы на обнаружение. Поскольку устройства поддерживают радио интерфейсы D2D выключенными до тех пор, пока не будет получена соответствующая команда активации, они не смогут быть атакованы или отслежены, прежде чем перейдут в активное состояние.

Подчеркнем еще раз, что предложенный протокол является рабочим решением и не претендует на оптимальность. Он в значительной степени опирается на существующие протоколы, такие как OAuth и SSL, для авторизации на сервере D2D для управления соединениями и передачи команд соответственно. Кроме того, протокол предполагает, что серверу D2D всегда можно доверять только потому, что он находится в сети оператора. Поэтому, необходимо оценить потенциальные риски безопасности в конкретных случаях, когда применяется D2D с сетевой поддержкой со стороны сотовой сети.

Интеграция с архитектурой 3GPP. Теперь, когда ясно, как должна работать система, можно сосредоточиться на вопросах ее интеграции с существующими сетевыми архитектурами. Благодаря гибкости ядра сети 3GPP, относительно легко выполнить интеграцию сервера D2D в существующую архитектуру 3GPP LTE. На рисунке 4.4 показано, как интеграция может быть осуществлена. Действительно, предложенная в диссертационной работе схема может быть сопоставлена с объектами архитектуры 3GPP почти точно, с единственным дополнительным объектом — сервером D2D, находящимся в EPC фрагменте сети. Такое размещение позволяет серверу D2D связываться с центром местоположения (SMLC), чтобы позиционировать пользовательское оборудование UE, а также эффективно взаимодействовать с серверами приложений внешних провайдеров и операторов.

Сигнальные опции для выгрузки трафика в D2D. Фактически протокол выгрузки трафика для D2D сегодня является частью стандартов 3GPP. При этом некоторые детали стандартизованного решения 3GPP

отличаются от первоначально предложенных. В частности, поток сообщений, предложенный в документах 3GPP, не учитывает проблемы взаимодействия с серверами приложений, поскольку это выходит за рамки работ 3GPP. В дальнейших исследованиях, что используется такой поток сообщений, который адекватно реализует функциональность, рассмотренную выше в диссертации.

Что касается состояния исследований в данной предметной области, то технология D2D с поддержкой сети по-прежнему в значительной степени не изучена. Например, если сеть давала бы возможность контролировать, какие D2D-линии используются, возможно было бы избежать выгрузки трафика на те линии D2D, которые могут ухудшить производительность сети и/или пользователя. Аналогичным образом, если бы сеть могла управлять процессом передачи через линии D2D, возможно было бы создать расписание, учитывающее, когда группы неконкурирующих D2D линий могли бы взаимодействовать, что потенциально существенно уменьшило бы конкуренцию и, соответственно, увеличило бы пропускную способность и энергетическую эффективность линий D2D [151;161;184;185]). Стоит отметить, что, конечно же, улучшенные параметры управления мощностью также становятся доступными, когда сеть помогает D2D-линиям [186]. В любом случае, предложенная в диссертационной работе архитектура достаточно масштабируема, учитывает почти все будущие модели сетевой поддержки, которые были предложены по сей день, и, таким образом, обеспечивает отличную основу для дальнейшего развития технологии D2D с поддержкой от сотовых сетей.

Модели взаимодействия пользователей разнообразны для будущих архитектур D2D. В соответствии с эволюцией технологий новые возможности появляются для людей, интегрируясь в их повседневную жизнь, и существенно влияют на нее. Предложенная в диссертации архитектура содержит первый набор функций, необходимых для интеграции D2D с поддержкой от сотовой сети и взаимодействия с остальной частью всемирной сети. Можно достаточно легко вернуться к простому D2D функционированию, когда по какой-либо причине сетевая поддержка становится недоступной или даже вредной, или предоставить соединения с приложениями

иным путем. Возможности здесь очень большие, но важно двигаться быстро, так как при предстоящем развертывании сетей связи пятого поколения будет наблюдаться значительный сдвиг в масштабе внедрения D2D технологии, при этом операторы будут рассматривать ее как «обязательную» технологию, а не только прототип.

Далее сфокусируем внимание на анализе производительности предложенных решений по выгрузке трафика и обсудим их применимость.

D2D в контексте создания 5G. Хотя не всегда существует необходимость отправки всего трафика через Интернет для синхронизации пары телефонов в одной и той же маршрутной таблице, пользователи все еще не привыкли к этому. Важно отметить, что попытки мониторить пользовательскую активность в целях выгрузки сессий по альтернативному пути в D2D, похоже, не дают прагматического решения и, таким образом, не могут использовать весь потенциал сетей D2D.

С другой стороны, разрешение D2D взаимодействия отражает важность приложения или предпочтения пользователя, а не стремление сети к выгрузке сессии, существенно улучшая качество обслуживания для этой сессии в перспективе. В то время как *производительность соединений D2D относительно сотовых каналов делает их потенциально привлекательными для пользователей*, это позволяет еще и *организовать целевое взаимодействие между близлежащими пользователями*, что дает возможность практически ими воспользоваться. Это важное положение не только лежит в основе исследований, проводимых в диссертационной работе, но и подчеркивает также изменения представления о том, как пользователи могут взаимодействовать с сетями.

Ожидается, что мобильные сети пятого поколения позволят принципиально увеличить пиковые скорости передачи данных, спектральную и энергетическую эффективность, а также обеспечить единообразное обслуживание в любое время, в любом месте и для чего-либо [187; 188]. Ориентированные на сети связи пятого поколения научные исследования стали основными в последнее время и при этом возникло понимание того, что одна технология не сможет удовлетворить все требования таких сетей, в

том числе по производительности и характеристикам качества обслуживания и качества восприятия [48]. Например, сверхплотные гетерогенные сети HetNets, скорее всего, станут одним из типовых решений по увеличению емкости в более чем 1000 раз в чисто сотовой топологии [189], хотя они очень дорогостоящие для создания и, следовательно, скорее всего, не будут решать проблемы 5G повсюду.

Технология D2D, скорее всего, станет частью сетевых сервисов по умолчанию, а также DNS и NTP. С точки зрения оператора, взаимодействие D2D, вероятно, может быть бесплатным для обычных приложений при обеспечении качества обслуживания и качества восприятия, а также новой контекстной информацией для сети. D2D можно использовать для достижения недорогих, с ультра малыми задержками или высокоскоростных соединений в приложениях, которые нуждаются в любой из этих особенностей, в то время как обычные услуги предоставляются сотовой сетью.

Поддержка услуг сети D2D, скорее всего, станет частью будущего портфеля оператора вместе со сверхплотными возможностями сети, взаимодействием машина-машина и множеством других услуг, которые станут возможными в сетях 5G. Сегодня сложно предсказать, будет ли D2D завтрашнего дня выглядеть точно как сейчас: технология может мигрировать в лицензированные полосы, несмотря на проблемы управления спектром, альтернативные уровни линий, использующие mmWave, могут также появиться. Остается надеяться, что в сетях 5G более не будет легче общаться на полпути по всему миру, чем в 5 метрах от требуемого пользователя, как это часто бывает сегодня.

4.5 Оценка предложенной архитектуры

Оценка производительности всегда являлась достаточно сложной задачей. В этом разделе обсудим, как именно производится оценка производительности динамической структуры топологии, и почему именно такая методология правильна. Далее представим подходы к оценке систем D2D

путем имитационного моделирования и, наконец, используем результаты некоторых из аналитических работ, которые выполнялись параллельно.

4.5.1 Методология оценки

В этом подразделе вводятся сетевые объекты и соответствующие механизмы, которые требуются для поддерживаемой со стороны сотовой сети технологии WiFi Direct. В том числе рассматривается предложенная методология оценки, которая может быть использована для широкого спектра перспективных технологий D2D и моделей P2P. Эта же методология используется на большинстве системных уровней оценки, используемых во второй части этого раздела.

Сетевые объекты. В диссертационной работе беспроводная сеть представляется как сеть, состоящая из ряда объектов, которые имеют возможности поддерживать различные радио технологии доступа. Это, во-первых, наложенная 3GPP LTE сеть, представленная базовыми станциями E-UTRAN Node B (eNB). Каждая базовая станция связана с ядром сети, обеспечивающим сотовую связность для всех беспроводных клиентов, ассоциированных с сетью.

Каждая базовая станция eNB доступна ряду мульти радио клиентских устройств, каждое из которых имеет возможность взаимодействовать как с LTE, так и с WiFi. Каждому клиенту доступны услуги, которые поддерживаются координатором MAC по сути являющимся коммутатором потоков трафика. В зависимости от рекомендаций координатора MAC клиент может воспользоваться для передачи данных интерфейсами LTE или WiFi. Координатор MAC абстрагируется от всех возможных сценариев выгрузки трафика.

Необходимо также принимать во внимание интерференцию в нелицензируемом спектре от устройств WLAN, которые могут быть расположены в непосредственной близости от точек доступа AP WiFi [190; 191].

Эти устройства имеют равные права с мульти радио клиентами сети по доступу к каналным ресурсам в соответствии с принципами построения сетей WiFi. В то же время они не могут быть ассоциированы с сотовой сетью, их функционирование в нелицензируемом спектре не может быть контролируемо или отслежено сетью LTE, будем далее называть их «изгоями».

Потоки трафика и нагрузка сети В соответствии с методологией исследований, отвечающей рекомендациям, изложенным в [192], предположим, что N клиентов LTE (которые в рассматриваемом случае имеют также возможность использования D2D) расположены равномерно в зоне развертывания сети. Поскольку существует возможность использования D2D взаимосвязей, две различные модели будут рассмотрены далее.

В работах [193;194] и [195] плотность клиентов выбирается достаточно высокой, так что каждый клиент в зоне радиуса D2D взаимодействия имеет возможность D2D соединений, по крайней мере, с еще одним клиентом. Однако, только половина клиентов могут быть «источниками» данных и передавать данные, поскольку P2P взаимодействия являются однонаправленными, при этом только один источник и один получатель данных существуют в каждой P2P сессии. Такой трафик моделируется как полностью заполненный буфер с пакетами длиной 1500 байт каждый. Далее, вместо моделирования распределения контента и запросов клиентов, предполагается, что определенный процент клиентов источников x в радиусе взаимосвязи D2D имеет своих P2P клиентов «получателей» данных [196; 197].

Устройства-изгой, как правило, представляют другим пользователям интернет тяжелый контент и представляются моделью заполненного буфера с пакетами длиной 1500 байт, направляемыми к своей точке доступа AP. Для упрощения оценки не будем использовать модель точки доступа WiFi AP для нисходящего трафика явно, поскольку это по существу просто другой пользователь с точки зрения протоколов. Таким образом, для моделирования нисходящего трафика просто откорректируем число клиентов-изгоев, чтобы имитировать желаемый уровень конкуренции в нелицензируемых диапазонах.

4.5.2 Имитационное моделирование поддержки D2D

Для проведения исследований в рассматриваемом разделе будем использовать симулятор WINTERsim, базирующийся на методологии оценивания LTE, рассмотренной в TR 36.814 [192] и других документах 3GPP, а также спецификациях IEEE для оценки производительности беспроводных сетей [198] и текущих стандартах IEEE 802.11, представленных в [199]. Симулятор может поддерживать различные стратегии построения сетей, модели трафика, характеристики каналов и беспроводные протоколы. Он позволяет моделировать все существующие инфраструктуры систем LTE и все известные методики размещения клиентов (гексагональные и квадратные соты, внешние условия, равномерное и кластерное распределения клиентов и т.д.).

Для каждого клиента предусмотрен собственный выделенный генератор трафика, позволяющий варьировать показатели для моделирования. Каналы моделируются с учетом всех соответствующих характеристик источника, получателя и среды распространения. Каждый узел способен поддерживать несколько радио интерфейсов, которые могут использоваться для восходящих и нисходящих соединений и моделируются при этом в пределах единой временной шкалы.

Далее обсудим основные предположения для имитационного моделирования, которые легли в основу сценариев для оценки эффективности развертывания D2D.

Сотовая сеть. Для оценки характеристик WiFi Direct с поддержкой сотовой сети рассмотрим следующий пример сценария, основанного на рекомендациях 3GPP при развертывании сетей в городских условиях. Сеть LTE включает в себя 19 гексагональных сот с тремя секторами в каждой (рисунок 4.5). Базовые станции eNBs поддерживают технологические решения Release 11 LTE и расстояние между близлежащими базовыми станциями eNBs составляет 200 метров, а радиус соты равен примерно 110 метрам.

Вложенная технология используется для повышения точности моделирования на краях области функционирования сети [198].

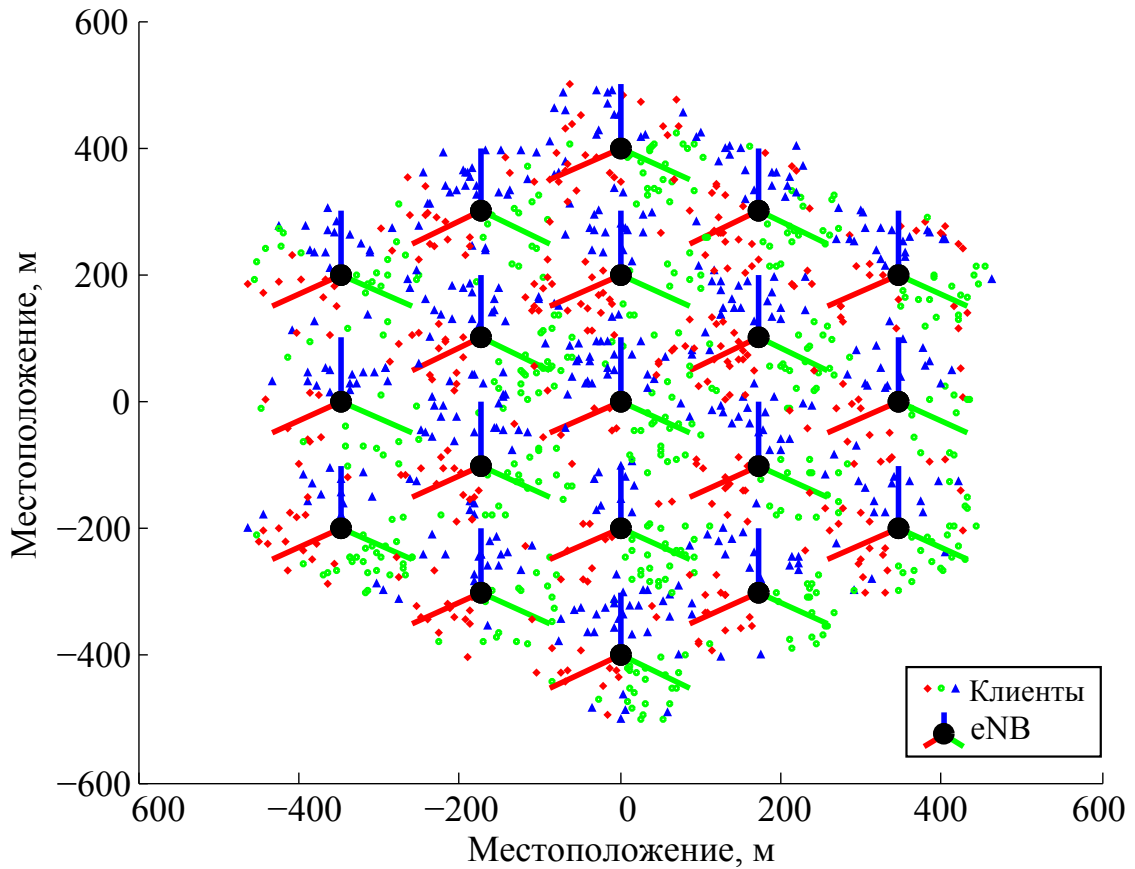


Рисунок 4.5 — Построение сети

Все соты работают в одном и том же 60 МГц лицензионном спектре, который разделен на три пары полос по 10 МГц для функционирования в режиме FDD. Каждая сота, в свою очередь, разделена на три сектора, и каждому сектору выделяется пара полос по 10 МГц, что приводит к шаблону повторного использования частот $1 \times 3 \times 3$. Клиенты 3GPP LTE связываются с базовой станцией eNB на основе лучшего значения DL SINR с порогом хэндовера в 1 дБ. Для получения дополнительной информации о стандартной конфигурации LTE можно обратиться к таблице 4.1 и соответствующим документам по стандартизации (например, 3GPP TR 36.814-900 и МСЭ-R M.2135-1). Для целей проверки производительности реализован сценарий калибровки по 3GPP TR 36.814-900, таблица A-2.1 и соответствующие тесты. Результаты моделирования хорошо вписываются в требуемые ограничения по спектральной эффективности как для центра соты, так и для ее границ (рисунок 4.6).

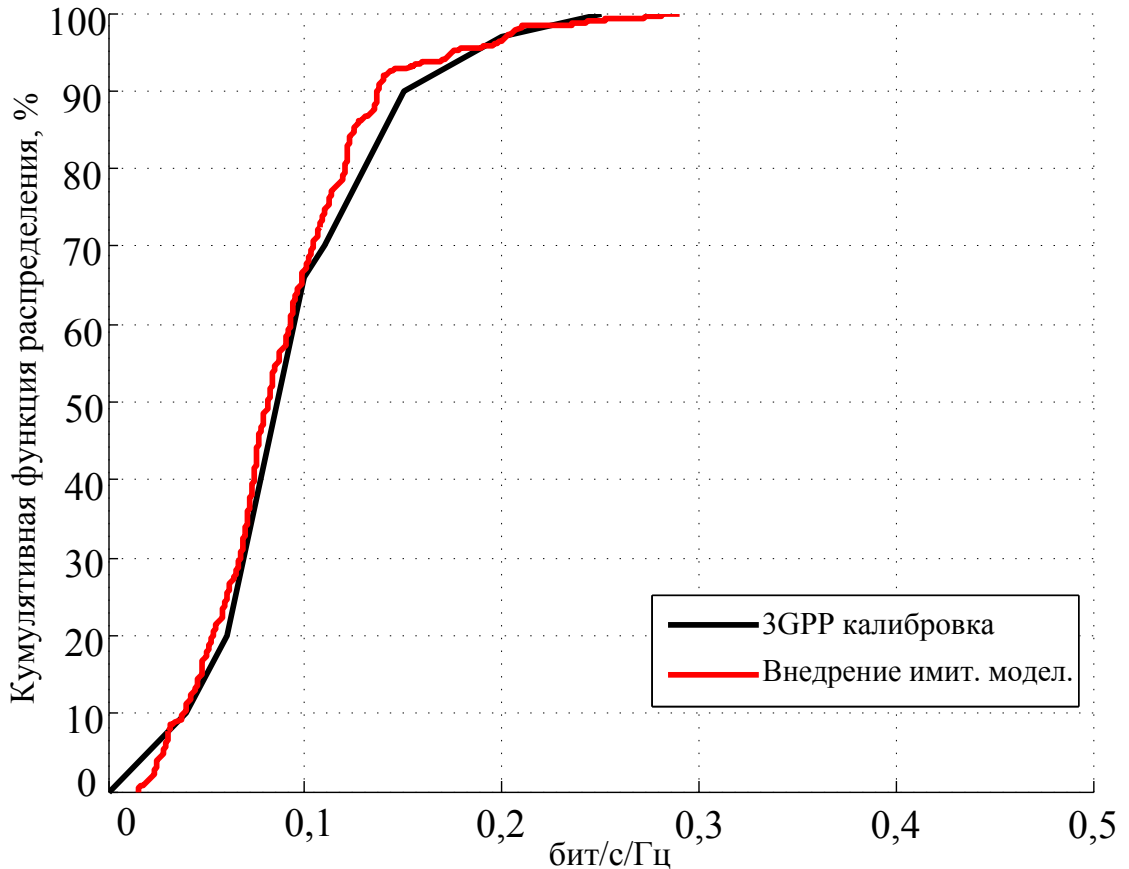


Рисунок 4.6 — Калибровка LTE: спектральная эффективность

Существующие построения сетей для WiFi. Будем полагать, что все точки доступа AP и относящиеся к ним клиенты (так называемые устройства-изгой) выполнены по одной и той же технологии, называемой IEEE 802.11-2012 [199]. Для того, чтобы имитировать реальные сетевые структуры, устройства-изгой должны быть расположены вокруг их соответствующих точек доступа. Точки доступа AP могут быть расположены где угодно внутри исследуемой сети: в кафе, на транспорте, в магазинах и торговых центрах и т.д. Расстояние от устройства-изгоя до его точки доступа ограничено максимально допустимыми потерями. Будем полагать при этом, что точки доступа AP и устройства-изгой не изменяют положения во время моделирования, поэтому процедура хэндовера не рассматривается. Далее будем рассматривать случай, когда число устройств-изгоев равно максимально возможному числу активных D2D-соединений, поэтому эти устройства могут задействовать около половины эффективных системных ресурсов в условиях максимальной выгрузки трафика.

Представляемое исследование предполагает, что все WiFi соединения (и AP, и D2D) используют один и тот же частотный диапазон и должны уступить любой активной передаче данных, для которой мощность превышает установленный порог. Для получения дополнительной информации о конфигурации сетей Wi-Fi можно посмотреть таблицу 4.1 и документацию драйвера с открытым исходным кодом Atheros, доступную онлайн. Для целей калибровки были использованы результаты публикаций по структурам ad-hoc сетей WLAN. Калибровка производительности Wi-Fi приводит к использованию результатов [148], при этом достигается почти идеальное сглаживание (рисунок 4.7) и разумное согласование со следующей опцией технологии D2D в при использовании FlashLinQ.

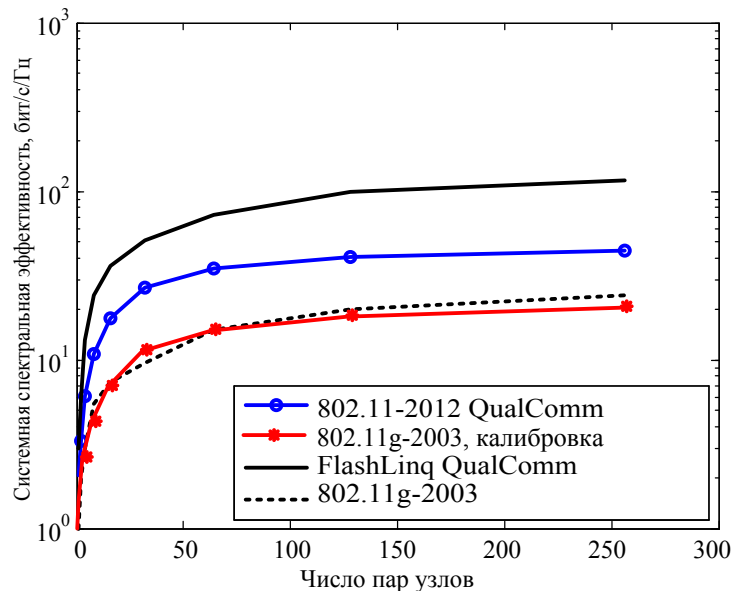


Рисунок 4.7 — Калибровка WiFi: спектральная эффективность 1×1 на квадратный км

В настоящем исследовании будем рассматривать функционирование WiFi в простом 20 МГц канале. Естественно, это можно легко экстраполировать на более общий случай, когда несколько каналов доступны в одно и то же время или канал агрегирован для использования.

Дополнительная функциональность D2D. Помимо перечисленных выше технологий, в диссертации дополнительно рассматривается технология WiFi Direct, которая в большинстве случаев похожа на Wi-Fi, но отличается местом назначения трафика. Специфика протокола WiFi Direct

Основные параметры для имитационного моделирования

Параметр	Значение
Параметры ядра сети	
Предел клиентской мощности для конкуренции Tx	23 дБм IRP на интерфейс
Модель мобильности	случайного направления движения, 3 км/ч скорость
Период наблюдения	10 секунд
LTE	
Модель затухания	ITU-R M.2135-1 [200], Tbl. A.2.2-1, A1-3
Модель замираний	ITU-R M.2135-1 [200], Sect. 1.3.1.1
Доступ к среде	Round-Robin расписание
Мощность и контроль скорости	SINR целевое значение на шлейф 15 dB
Частотные ресурсы	10 + 10 МГц FDD в каждом секторе, короткий CP
Режим сигнализации	2 из 20 специальных подкадров, 10 ms кадр
Радио оборудование	ITU-R M.2135-1 [200], Tbl. 8-4
Конфигурация антенны	1x2 (разнообразный прием на базовой станции eNB)
WiFi	
Модель затухания	Эмпирическая, основанная на [201]
Модель замираний	Только корреляция, основанная на [202]
Доступ к среде	CSMA/CA, -76 дБм мягкий порог
Управление мощность и скорость передачи	Целевое значение SINR для открытого шлейфа в 25 dB
Частотные ресурсы	20 МГц TDMA
Режим сигнализации	Green-field, управляемая скорость 18 Мбит/с, RTS/CTS
Радио оборудование	Чувствительность -95 дБм
Конфигурация антенны	1x1 (простая антенна)

в значительной степени не применяется в данном исследовании, поскольку большинство специфических функций не нужны из-за того, что используется сетевая поддержка. Вместе с тем, такие функции WiFi Direct, как

многопоточность полезны, но моделировать их в явном виде нет необходимости.

После того, как в модели настроено функционирование WiFi Direct, можно разрешить координатору MAC управление им так, как это обычно и делается. В действительности, WiFi Direct линия имеет свой собственный буфер, так что пакеты, которые могли бы быть безвозвратно потеряны, перенаправляются в этот буфер с помощью управляющей логики.

4.5.3 Сценарий применения гетерогенных сетей

Рассматривается типичный сценарий использования гетерогенных сетей (Heterogeneous Network, HetNet), представленный на рисунке 4.8. В основе покрытия беспроводной сети лежит макро базовая станция (БС) с радиусом покрытия r_M . С целью повышения пространственного переиспользования беспроводного ресурса сети также используются M микро БС, расположенных вокруг макро БС, каждая из которых имеет радиус покрытия равный r_m . Также в сценарии присутствуют мобильные пользователи с устройствами, поддерживающими технологию прямого взаимодействия (Device-to-Device, D2D), при этом некоторые из пользователей могут выступать в роли активных D2D партнеров, тем самым предоставляя дополнительные возможности для подключений. Пользователи, желающие получить запрошенные данные, могут быть подключены только к одному уровню рассматриваемой гетерогенной сети, т.е. могут быть подключены только к макро БС, микро БС или же к активному D2D партнеру, если последний находится в непосредственной близости (не далее, чем r_D).

В моменты глобальной оптимизации макро БС следует алгоритму, ориентированному на перераспределение беспроводных радиоресурсов всей рассматриваемой сети с целью обеспечения каждого пользователя максимально возможным качеством обслуживания (т.е. применяется алгоритм нацеленный на максимальную пропускную способность или максимальную энерго эффективность). Так как целью исследования является анализ вре-

менных зависимостей поведения системы, то в случае любых локальных изменений подключений пользователям дается возможность самим решать к какому уровню сети они подключатся, максимизируя свою собственную пропускную способность. При этом могут получаться случаи, когда выбранное подключение не будет оптимальным по причине отсутствия достаточного количества радио ресурсов, повышенного уровня интерференции или же отсутствия прямого соединения. Так как пользователи в первую очередь пытаются получить максимальное качество обслуживания именно для себя, то получается так, что D2D соединения, обладающие наибольшей пропускной способностью, получают больший приоритет, а макро БС — наоборот меньший. Стоит отметить, что есть случаи, когда применение алгоритма гарантирующего максимально возможную пропускную способность для каждого пользователя приведет и к улучшению энерго эффективности сети [190].

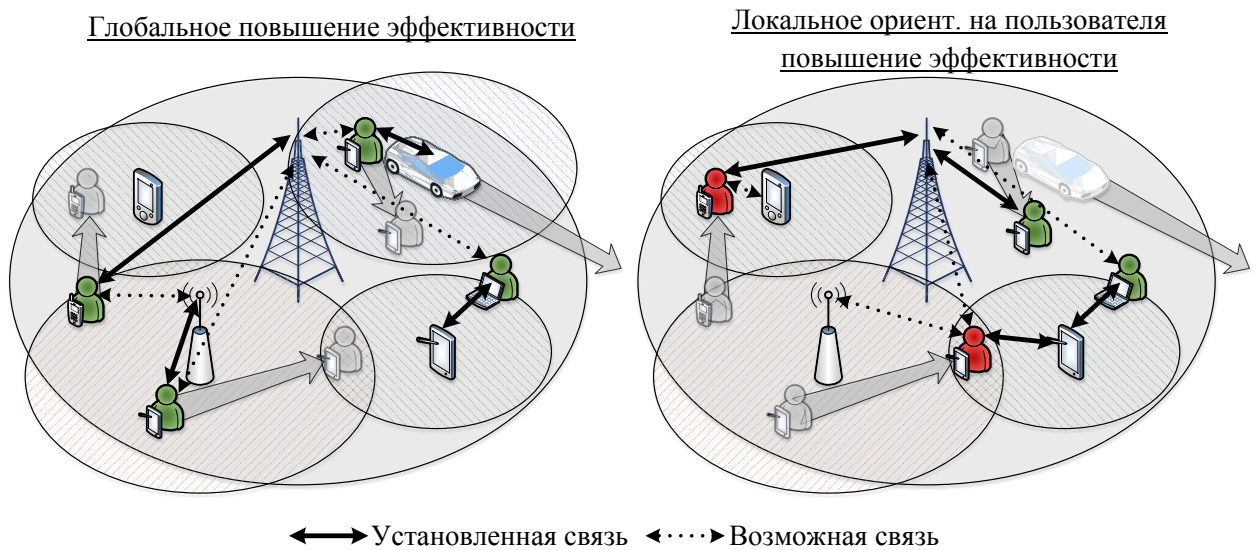


Рисунок 4.8 — Пример D2D сценария

Будем предполагать, что пользователи запрашивают какой-либо популярный медиа-ресурс (к примеру, обмен медиа на массовом мероприятии, таком как футбольный матч или концерт). Подобного рода информация обычно кэшируется макро и микро БС, расположенными неподалеку от точки концентрации запросов на него. Также предполагается, что этот ресурс также доступен и у $N - K$ из N пользователей. таким образом, остается K пользователей, генерирующих запросы на получение этого ресурса.

Все пользователи мобильны и движутся согласно Random Direction Model (RDM). В работе [203] было показано, что данная модель движения дает качественно схожие результаты в сравнении с применением гораздо более сложных и реалистичных моделей движения. Так как макро БС представляет собой центральную сущность, координирующую процесс подключения пользователей, то на ней доступна дополнительная информация такая, как расположение всех пользователей в каждый момент времени, карта интерференции и возможности подключения пользователей. Эта информация используется макро БС для расчета оптимальных подключений и распределения радио ресурса для пользователей, которые сообщаются им с целью управления выбора оптимального подключения пользователем в рассматриваемой сети.

4.5.4 Моделирование D2D системы

Временные зависимости поведения описанной выше HetNet сети могут быть качественно описаны при помощи разработанной модели. Рассмотрим начальный момент времени $t = 0$ сразу после глобальной оптимизации 5G системы. Состояние системы можно описать вектором, содержащим количество подключенных пользователей к определенным сетевым сущностям, включающих в себя макро БС, микро БС и активных D2D партнеров. Изменение состояния сети может быть описано полумарковским процессом, в котором время между изменением состояния соответствует минимальному времени изменения состояния подключения одного из рассматриваемых пользователей, которое в первую очередь зависит от средней скорости мобильного пользователя.

В общем случае, задачу анализа поведения 5G можно представить в виде цепи Маркова в дискретном времени с интервалами дискретизации равными τ , минимальному времени смены состояния подключения одного из пользователей, с последующим расчетом необходимого числа шагов для смены глобально оптимизированного состояния в момент времени t_0

на состояние, к которому пришли пользователи к какому-либо времени t_1 , $t_1 > t_0$, делая исключительно эгоистичные решения. Пусть Δ_g — абсолютная разница времени между глобально оптимальным состоянием и состоянием, определенным пользователями, а Δ_t — разница времени до оптимального состояния, $\Delta_t = t_1 - t_0$. Стоит отметить, что параметризация данной модели является очень сложной задачей, так как расчет оптимальных распределений пользователей по уровням сети, т.е. задание состояний Марковской сети, и последующее получение минимальных времен смены состояния подключения пользователей являются нетривиальными задачами. Также тот факт, что пользователь в один и тот же момент времени находится в различных зонах обслуживания нескольких уровней сети, значительно увеличивает пространство состояний цепи. Но, не смотря на это, предложенный подход позволяет дать качественную оценку поведения беспроводной сети после момента глобальной оптимизации.

Заметим, что в целом данная модель может рассматриваться как стохастическая система, приближающаяся к ее стационарному состоянию, следуя локальным решениям отдельных пользователей. Известно, что время прихода к стационарному состоянию пропорционально сумме $C \sum_{i=1}^{H-1} \nu_i^n$, где n — это номер шага цепи, $\nu_i < 1$, $i = 1, 2, \dots, H - 1$ — собственные значения матрицы переходных вероятностей, H — общее количество состояний цепи, а C — константа. Матрица переходных вероятностей неприводимой цепи Маркова всегда имеет доминантное собственное значение, т.е. собственное значение близкое к единице, тем самым перевешивая остальную сумму. Положив $\Delta t \rightarrow 0$, приближенная функция равна $y(t) = Ce^{-\lambda t}$, где λ — это интенсивность отклонения от оптимального состояния по отношению к выбранной характеристике.

Таким образом, ожидаемо экспоненциальное ухудшение качества после момента глобальной оптимизации системы. С целью подтверждения этих теоретических заключений и с целью получения количественных характеристик поведения беспроводной сети, далее будет описана процедура моделирования системы в целом.

4.6 Результаты имитационного моделирования

4.6.1 Имитационное моделирование системного уровня

Численный анализ системы был проведен при помощи системы имитационного моделирования WINTERsim. Был построен сценарий, в котором пользователи были равномерно распределены внутри покрытия multi-RAT 5G сети (сети с множественным набором технологий доступа) и передвигались согласно RDM модели движения, в то время пока они принимают большой объем данных. Были смоделировано три типа уровней HetNet сети: (i) макро 3GPP LTE eNodeB, (ii) 3GPP LTE малая сота (микро, фемто сота) и (iii) пользовательские устройства с возможностью прямого взаимодействия на основе Wi-Fi. При этом 20% из N пользователей являются D2D кэширующими точками с целью распространения данных другим пользователям. Макро LTE eNodeB с радиусом покрытия в 500 метров предоставляет 100 ресурсных блоков, из которых 25 доступны 10 раскиданным в ее радиусе фемто сотам.

Состояния отдельных беспроводных каналов оцениваются на основе SINR на каждой из поднесущих с учетом затухания и замирания сигнала. Обнаружение и установление D2D соединений управляется eNodeB (т.е. при помощи 5G ProSe функционала), а передача данных осуществляется при помощи протокола Wi-Fi-Direct.

Передаваемый поток данных моделировался на основе сервиса потокового вещания видео Facebook Live в качестве 720p и максимальной скорости передачи данных в 4 Мбит/с. Рассматриваемые системные метрики были: (i) общая пропускная способность сети в зависимости от времени и (ii) общая энерго-эффективность системы от времени. Была дана оценка данным параметрам с учетом (i) их изменения в течении определенного временного окна и (ii) соответственной интенсивности мобильности пользователей. Остальные параметры использованные при моделировании перечислены в таблице 4.2.

Основные параметры моделирования

Системны параметр	Значение
Радиус макро БС	500 м
Радиус D2D канала	30 м
Радиус фемто сот	50 м
Частота несущей LTE	2,6 ГГц
Частота несущей WiFi-Direct	2.5 ГГц
Мощность передачи абонента	23 дБм
Мощность передачи макро БС	46 дБм
Мощность передачи фемто соты	20 дБм
Время установления D2D канала	1 с
Макс. скорость D2D канала	40 Мбит/с
Количество запусков	500
Параметры приложения	Значение
Разрешение видео	720p, 30 fps
Интервал ключевого-кадра	Каждые 2 с
Макс. скорость	4 Мбит/с
Управление скоростью	CBR
Частота аудио	44,1 КГц
Скорость аудио потока	128 кбит/с

С учетом применяемого алгоритма оптимизации, данная модель управляет глобальным выбором наиболее подходящих подключений при помощи специальной функции, которая учитывает пропускную способность отдельного пользователя и мощность передатчика по отношению к текущей загрузке сети. Также разработанный метод выбора уровня сети HetNet.

4.6.2 Результаты численного анализа

На рисунках 4.9 и 4.10 представлены результаты моделирования для трех различных параметров мобильности пользователей: 5.5 км/ч (малая мобильность), 45 км/ч (средняя) и 100 км/ч (высокая). Можно легко за-

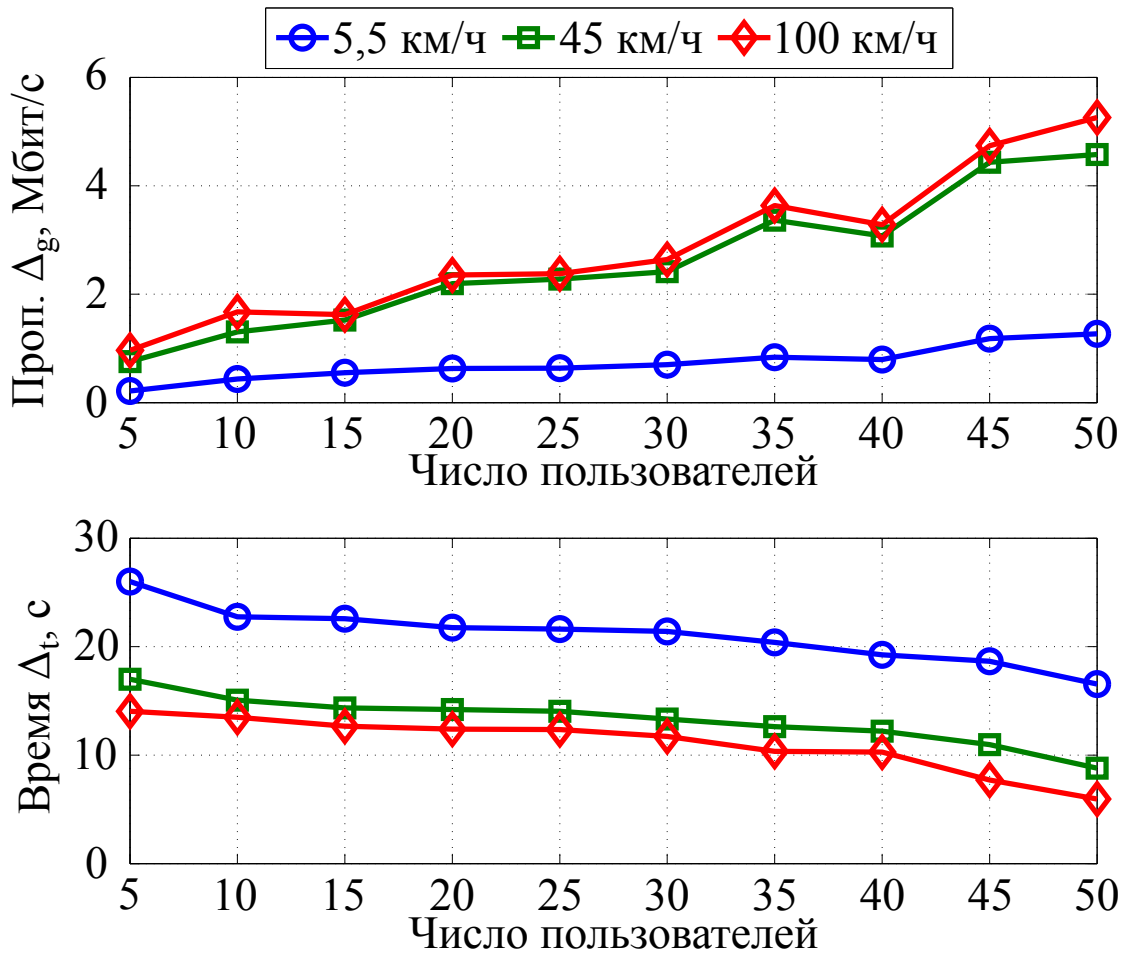


Рисунок 4.9 — Совокупная пропускная способность в зависимости от числа пользователей

метить, что общая пропускная способность сети и ее энерго-эффективность начинает быстро ухудшаться с течением времени с момента глобальной оптимизации системы вне зависимости от параметра мобильности пользователей. Время перехода сети из глобально оптимального состояния в состояние, управляемого пользователями, определяется путем применения экспоненциально-взвешенного скользящего среднего (exponentially-weighted moving average, EWMA) с соответствующим коэффициентом равным $\gamma = 0.1$.

Важно отметить, что понижение интенсивности мобильности пользователей влечет за собой повышение времени оптимальной работы системы, так как сама система в целом начинает гораздо медленнее изменять свое состояние — более быстрые пользователи гораздо быстрее покидают свою зону обслуживания и у них становится гораздо больше шансов встретить

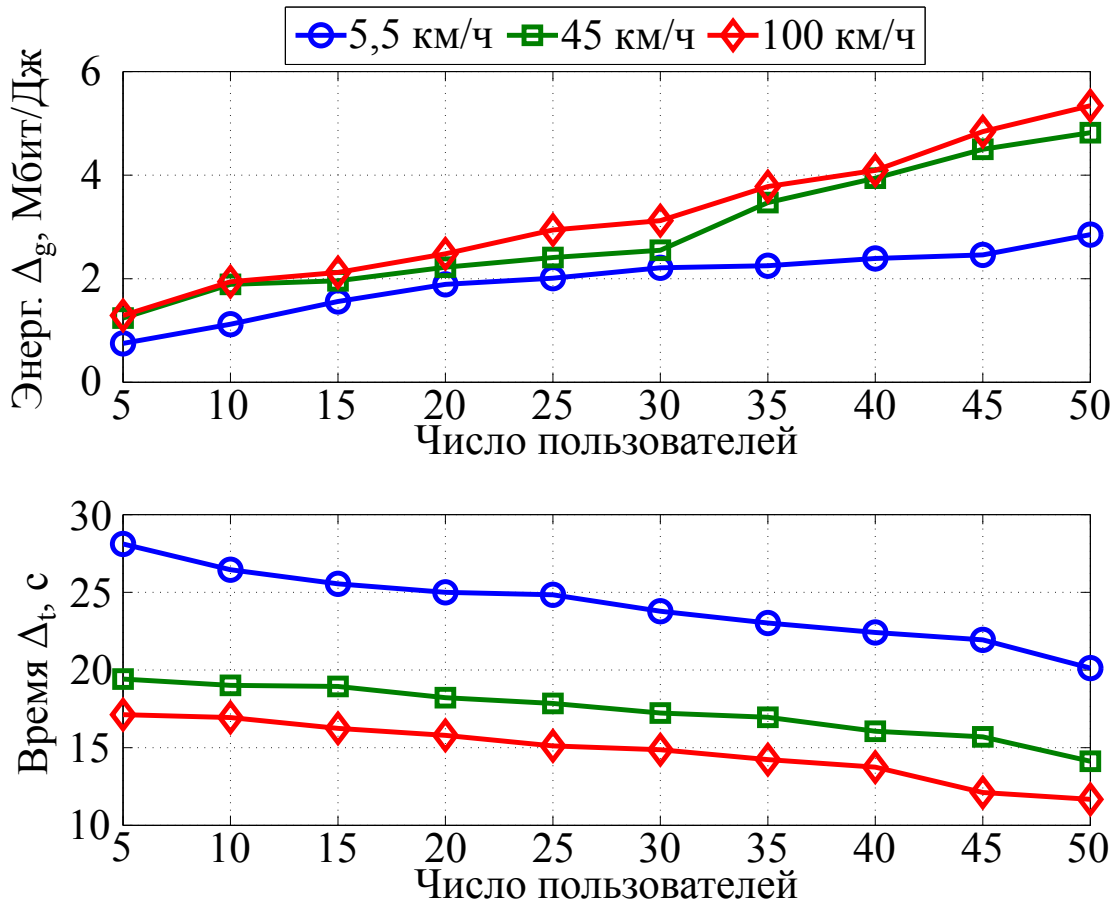


Рисунок 4.10 — Совокупная энергоэффективность в зависимости от числа пользователей

альтернативный (неоптимальный) источник данных. Не менее важно отметить, что разница между оптимальным и неоптимальным состояниями огромна. Так энерго-эффективность сети может ухудшиться на порядок.

На рисунке 4.11 представлены величины Δ_g и Δ_t как для общей пропускной способности сети, так и для ее энерго-эффективности в зависимости от числа пользователей, запрашивающих потоковые данные. В соответствии с предыдущими результатами, более высокая мобильность пользователей влечет за собой более быстрое ухудшение характеристик сети. В этом случае приходится гораздо более часто проводить глобальную оптимизацию сети.

Данное поведение системы хорошо видно из рисунка 4.11, на котором видно, что время Δt отклонения состояния сети от оптимального становится меньше с ростом интенсивности мобильности пользователей. Похо-

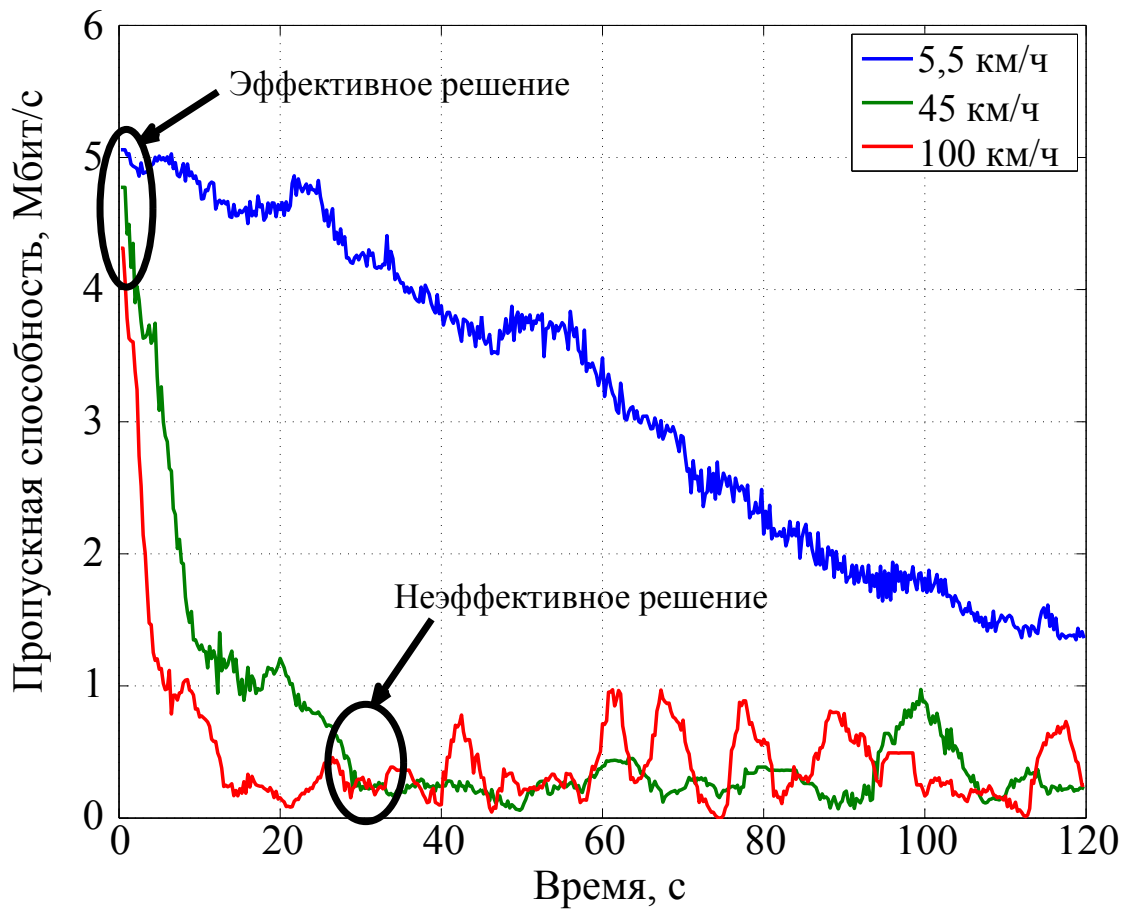


Рисунок 4.11 — Совокупная пропускная способность ($N = 50$)

же поведение наблюдается и для характеристики энерго-эффективности сети. Заметим, что при определенных значениях скорости пользователей увеличение количества самих пользователей не так сильно влияет на характеристики системы. При этом рост и убывание значений Δ_g и Δ_t всегда остаются линейными.

Подводя итог, можно сказать, что мобильность пользователей является ключевым параметром, оказывающим влияние на выбор необходимого периода переоптимизации системы. В то время, с малой мобильностью (около 5 км/ч) время ухудшения состояния сети достаточно велико и остается на уровне нескольких минут, а с высокой мобильностью оно значительно убывает. В результате этого выбор моментов времени, когда необходимо переоптимизировать беспроводную сеть становится основной проблемой для операторов сотовых сетей.

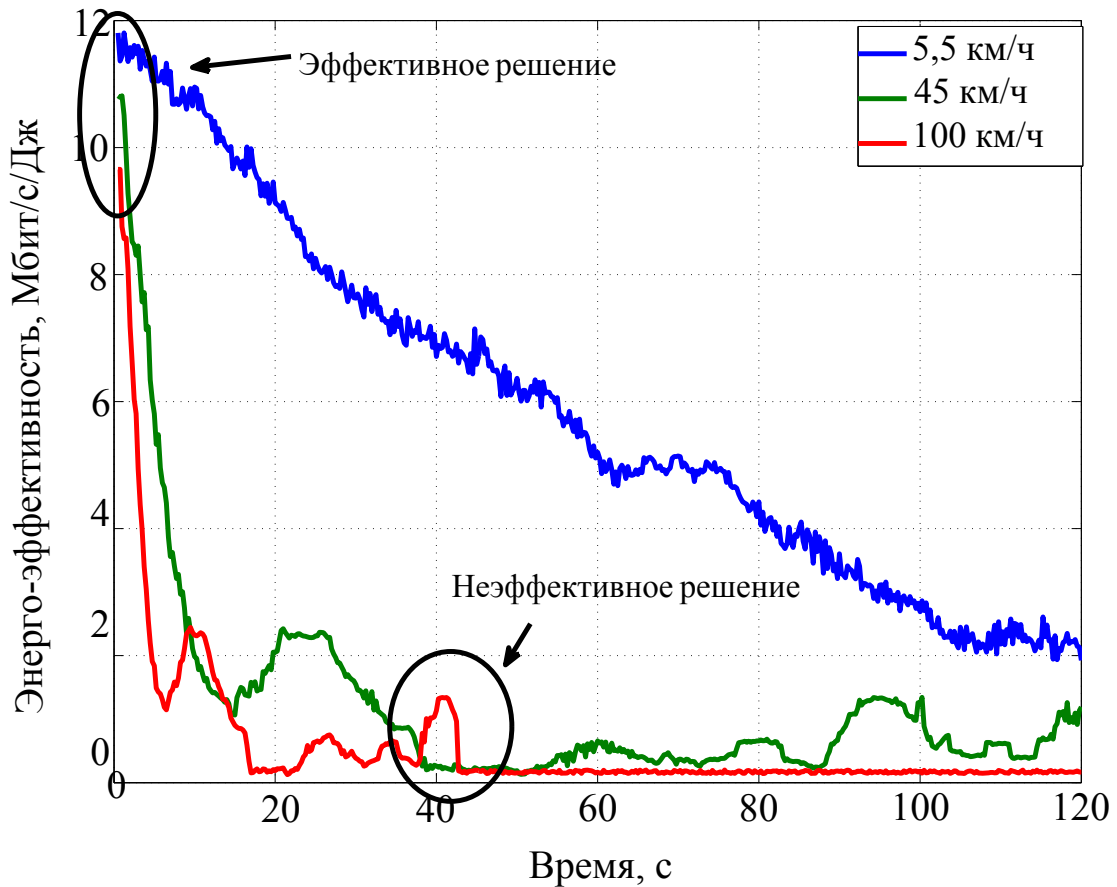


Рисунок 4.12 — Совокупная энергоэффективность ($N = 50$)

Значения интенсивности отклонения состояния сети от оптимального, λ , могут быть получены путем решения уравнения $y(t) = Ce^{-\lambda t}$ по отношению C и λ в моменты времени $t = 0$ и $t = \Delta t$. Таким образом получается, что $C = G_2$, а $\lambda = (1/\Delta t) \ln(G_2/G_1)$, где $\Delta_g = G_2 - G_1$. Эти характеристики приведены в таблице 4.3.

4.7 Практическая реализация

В этом разделе рассмотрим некоторые основные детали разработки, что делает предложенную в диссертации систему выгрузки трафика в D2D с сетевой поддержкой достаточно реальной для внедрения в сотовые сети и Интернет. На сегодняшний день ключевые аспекты протокола, а так-

Интенсивность отклонения λ при различных уровнях мобильности.

Число пользователей	Скорость		
	5.5 км/ч	45 км/ч	100 км/ч
5	0.01022	0.09001	0.17585
10	0.01233	0.08188	0.22426
15	0.01625	0.12330	0.18381
20	0.01235	0.13359	0.23081
25	0.01207	0.13837	0.18455
30	0.01279	0.13927	0.23914
35	0.01172	0.15923	0.28507
40	0.01255	0.16279	0.28836
45	0.01400	0.21258	0.45732
50	0.01600	0.24040	0.59603

же сервисная архитектура, уже являются частью версий Release 12 и 13 в стандартах 3GPP. Однако, как правило, некоторые аспекты реализации не являются частью каких-либо спецификаций или стандартов, поскольку достаточно много требований всегда остается на усмотрение разработчиков. Хотя большинство из них и не очень сложные, необходимо обеспечить их комплексное взаимодействие при реализации предложенной технологии D2D с сетевой поддержкой.

4.7.1 Программная платформа

Хотя сетевая архитектура D2D является довольно специфичной, даже в терминах поддерживаемых уровней линий, программные платформы для реализации ранее не обсуждались. Далее в этом подразделе рассмотрим программные компоненты для реализации технологии D2D сетевой поддержкой, а также необходимые изменения в них.

Андроид и IP сети. Android, как любая Linux-базирующаяся система, позволяет одновременные соединения с несколькими сетями через более,

чем один интерфейс, но даже если оба интерфейса 3GPP LTE и WiFi активны, телефон обычно использует только один из них в качестве шлюза по умолчанию для IP трафика. При этом можно использовать D2D линию только тогда, когда адресом назначения IP-пакета является WiFi Direct адрес партнера. Из-за отсутствия резервных IP-адресов, однако, для линии WiFi Direct необходимо использовать некоторый набор адресов, который означает, что если линия D2D отключается, то партнер становится недоступным, даже если существует альтернативный путь (поскольку частный набор пакетов не переадресовывается). По этой причине желательно, чтобы было возможным использовать общедоступный IP-адрес партнера в системе 3GPP LTE в WiFi Direct.

Таким образом, одна из задач практического применения предложенного в диссертации подхода заключается в создании решения, которое будет прозрачным для уже существующих приложений и таким образом облегчит совместимость разработанной технологии выгрузки трафика. По этой причине изменения физического уровня не может принести желаемого результата, поскольку он в большой степени зависит от поставщика. Аналогичная ситуация связана с канальным уровнем: для изменения логики маршрутизации в существующем оборудовании потребуется поддержка от поставщиков, а создание виртуальных интерфейсов вызывает дополнительные накладные расходы в заголовках. Поскольку приложения существенно зависят от существующих магистральных сетей всевозможные модификации на этом уровне тоже невозможны. На сетевом уровне IP-адреса привязаны к физическим интерфейсам, но решения о переадресации принимаются независимо от интерфейсов. Это позволяет создать независимое от интерфейса решение без необходимости внесения изменений на верхние уровни модели взаимодействия открытых систем с использованием мобильных IP/виртуализации.

Конфигурация по умолчанию в системе Android позволяет иметь мульти маршруты для шлюзов, но шлюзы включаются в таблицу маршрутизации с различными затратами на пути соединений. Поэтому балансировка нагрузки не осуществляется и только один маршрут шлюза используется в конкретный момент времени. В случае LTE (или любой другой

сотовой сети) интерфейса и WiFi, выбор маршрута шлюза LTE предпочтительнее, когда ожидается соединение с Интернет, в то время как для линии D2D не предназначены для Интернет соединений вообще. Следовательно, изменение стоимости маршрута шлюза может привести к тому, что Интернет будет недоступен для всех приложений в мобильном устройстве, поскольку весь трафик будет перенаправлен на одноранговый D2D.

Поэтому в диссертационной работе предлагается решение о введении маршрута, основанное на разрешении мобильному устройству маршрутизировать IP-трафик как обычно, и вводить в таблицу маршрутизации маршрут для связи пользователем для определенного однорангового пути. Благодаря ядру Linux в системе Android, это возможно путем изменения конфигурации системы *net.ipv4.conf.all.forwarding*. После этого изменения мобильное устройство Android имеет возможности обычного маршрутизатора и, таким образом, может передавать пакеты от одного интерфейса к другому. Это позволяет отправлять IP-пакеты с помощью адреса источника, эквивалентному общему IP-адресу интерфейса LTE и месту назначения, и адрес, равный частному IP-адресу одноранговой сети, при этом IP-адрес IP уровня системы Android будет отправлен через Wi-Fi линию.

Поскольку следует выгрузить трафик только для одного партнера, можно вставить маршрут в таблицу маршрутизации таким образом, что общедоступный IP-адрес LTE будет возможно использовать через Wi-Fi Direct частный IP-адрес. Вставка выполняется командой:

```
| ip route add PEER_LTE_IP/32 via PEER_WD_IP
```

После этого весь трафик с IP-адресом назначения PEER_LTE_IP будет перенаправлен напрямую на WiFi Direct интерфейс с кратчайшим путем, проходящий через линию WiFi, как это и предполагалось. Когда такая связь больше не нужна, маршрут можно удалить следующим образом:

```
| ip route del PEER_LTE_IP/32 via PEER_WD_IP
```

Вставка и/или удаление выполняется приложением на стороне клиента, которое работает как системная услуга. Обе операции могут быть выполнены под управлением сервера D2D, но удаление может быть также проведено и на основе качества линий (например, при неудовлетворительном значении RSSI).

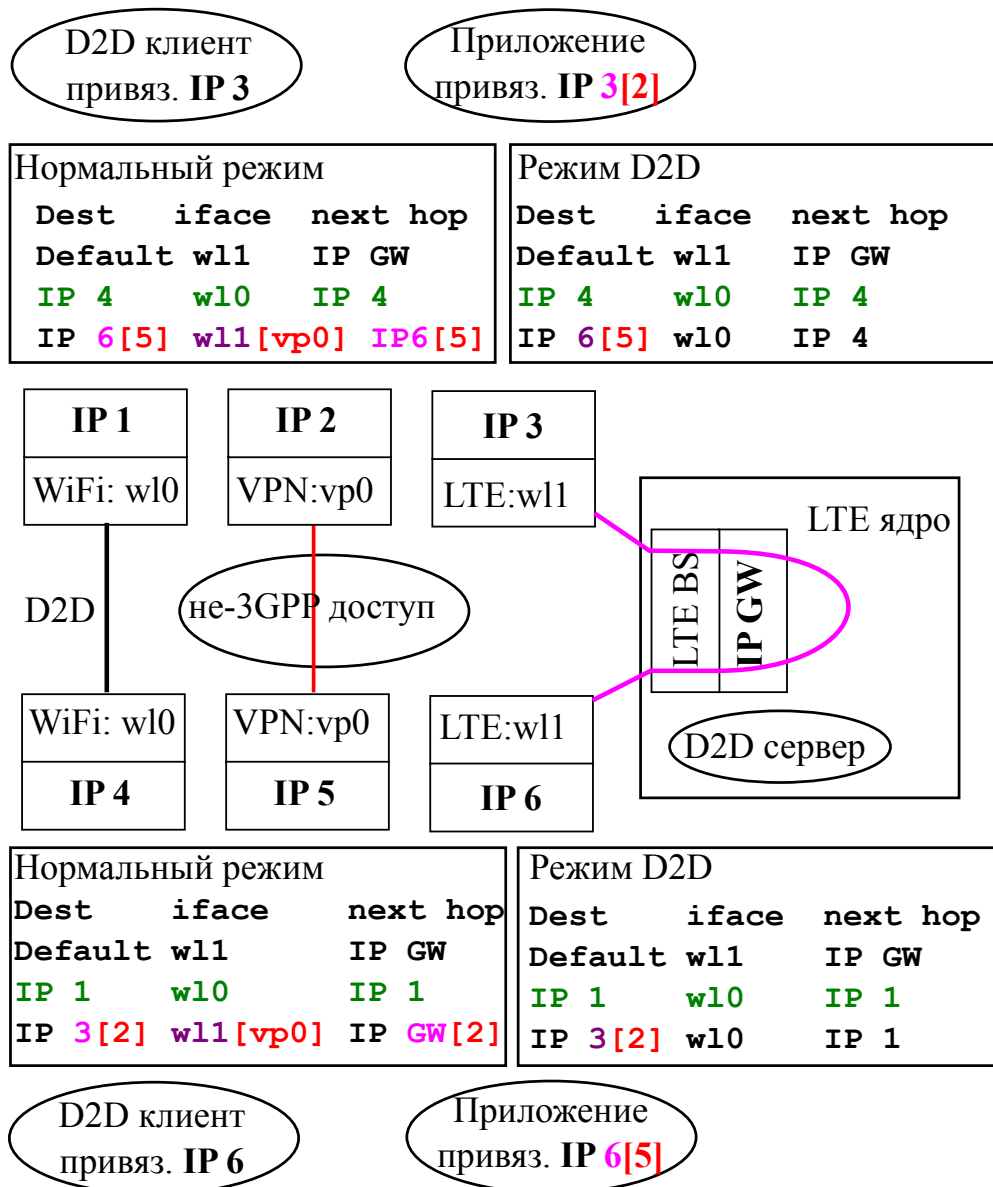


Рисунок 4.13 — Пример введения маршрута

При достижении определенного порога уровня сигнала маршрут можно вставлять или удалять по мере необходимости. Пример таблицы маршрутизации во время введения маршрута в случае WiFi Direct представлен на рисунке 4.13. Можно видеть, что трафиком можно управлять не только через WiFi Direct, но и через другие формы не-3GPP доступа, например, кампусные сети.

Рассмотренное введение маршрута в таблицу маршрутизации не требуется на обоих устройствах, участвующих в выгрузке трафика в D2D; на самом деле обратный трафик может быть обслужен через сотовый интерфейс, что может быть полезно в некоторых сценариях. Также введе-

ние маршрута не ограничено единственным одноранговым узлом: множество партнеров могут иметь свои собственные пользовательские маршруты при условии, что им выделены различные частные IP-адреса (сервер D2D вполне может это обеспечить).

Инфраструктура и поддержка платформы. Был разработан прототип прототип технологического решения для выгрузки трафика при поддержке сотовой сети, включающий в себя три основных компонента:

1. Обслуживание на клиентской стороне, работает в фоновом режиме на пользовательских устройствах UE.
2. База данных контента, которая поддерживает линии контента и управление доступом.
3. Сервер D2D, который также выполняет функции прокси сервера (proximity-based services, ProSe), управляет соединениями и координирует их [204].

База данных контента — это веб-платформа, предоставляющая услуги обмена данными для его пользователей, хорошими примерами являются социальные сети, такие как Facebook, Google+ и YouTube. Сервер ProSe предназначен для оператора сотовой сети и/или поставщика услуг Интернета (ISP). Следует отметить, что альтернативно предложенной архитектуре аналогичная услуга может быть предоставлена и когда оба объекта управляются оператором. При этом перенос пользователей из известных социальных сетей на новую платформу может быть более сложной задачей. Примеры реализации приведены на рисунке 4.14.

В качестве доказательства [205] предложенной концепции приведем то, что все функции на стороне сервера были развернуты внутри разработанной облачной инфраструктуры. Две виртуальные машины были настроены так, чтобы выполнять функции регистра контента и сервера D2D соответственно. Контентный регистр был реализован как PHP-скрипт, обслуживаемый веб-сервером Apache, используя при этом API Google+. В предложенной реализации приложение является регулярным веб-сайтом, что дает зарегистрированному пользователю возможность опубликовать свое намерение обменяться некоторым контентом или искать общие за-

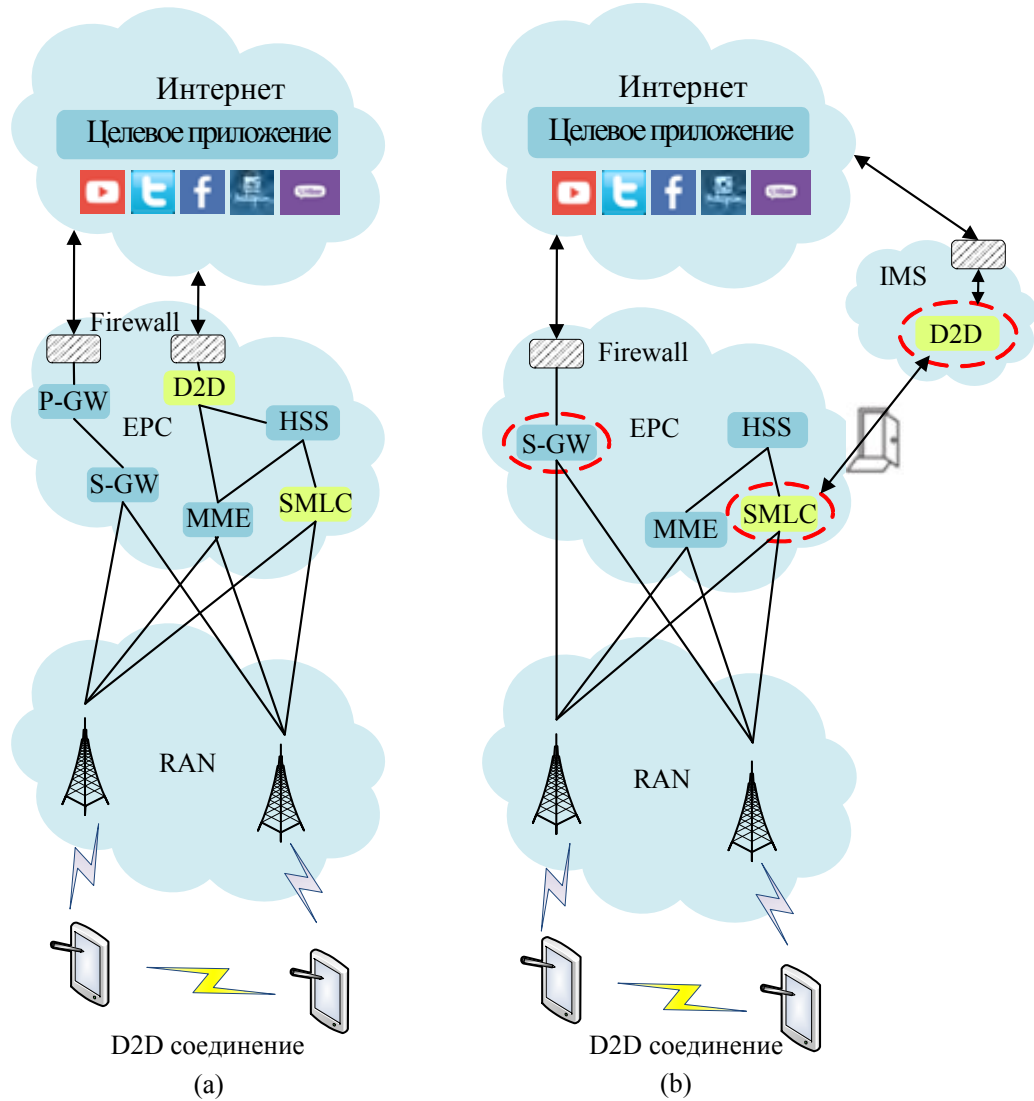


Рисунок 4.14 — Сравнение стандартизированной и исследуемой архитектур

писи данных (например, торрент-трекер). Пользователь вводит только ту информацию, которая необходима для доступа к данным, а не к их местоположению, то есть протокол совместного доступа и номер порта, в то время как настройке IP-соединения помогает сервер D2D ProSe. При этом внедрена новая схема адресации для D2D-соединений, а также новый идентификатор протокола в URI: *d2d://*. Устройства конечного пользователя могут быть легко сконфигурированы для того, чтобы интерпретировать этот тип протокола как запрос на обслуживание клиента, который, в свою очередь, способен связываться с сервером ProSe для того, чтобы разрешить исполь-

зывать имя пользователя на уровне приложения для совместного доступа и соединения к одному из его устройств, на котором размещается контент.

Сервер ProSe был реализован на Python как отдельное приложение с использованием HTTPS в качестве транспорта для управляющих сообщений. Система предполагает, что мобильная линия передачи данных пользовательского оборудования UE должна быть включена во время использования услуги. Предложенное решение показывает, что регистр контента и сервер(ы) ProSe легко интегрируются в бесшовной форме в существующую инфраструктуру веб-обслуживания и не требуют каких-либо решений, которые не вписывались бы в строго определенную веб-парадигму. Соответствующий протокол приведен на рисунке 4.15.

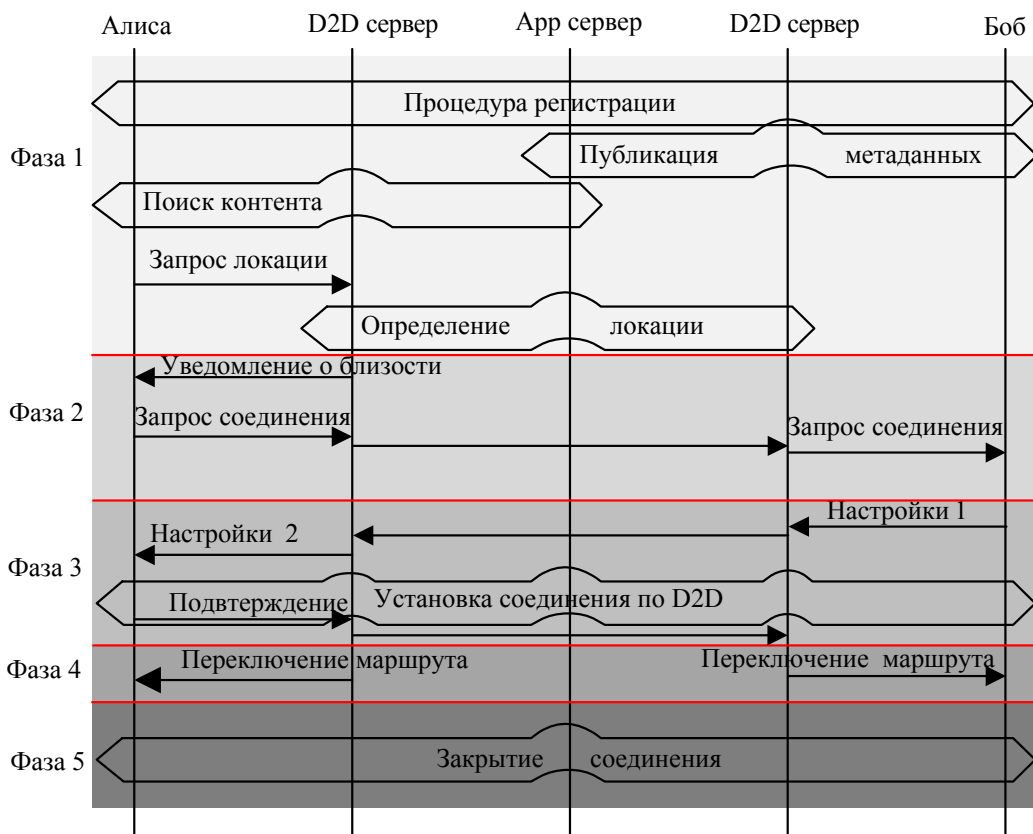


Рисунок 4.15 — Протокол управления соединением D2D

В демо версии использовались телефоны Sony Xperia ZL с Android OS и Samsung Galaxy S4. Одной из основных функций, необходимых для пользовательского оборудования UE, была поддержка сотовых и WiFi соедине-

ний одновременно. Из-за ограничения потребления энергии большинство систем избегают такой функции, поэтому необходимо было обойти систему управления услугами на базе Android WiFi и напрямую взаимодействовать с драйвером WiFi. Аналогично GNU Linux Android предоставляет необходимые инструменты: *wpa_supplicant* — интерфейс, управляемый с помощью утилиты *wpa_cli*. К сожалению, никакая прошивка не позволяет получить доступ к этим утилитам, даже для разработчиков, и, поэтому, использовалась послепродажная прошивка Cyanogenmod, которая может быть применена любым пользователем, кто владеет соответствующим неблокированным устройством.

Одним из ключевых требований для мобильной платформы конечного пользователя является способность принимать входящие P2P-соединения через сотовый канал передачи данных. В связи с тем, что большинство операторов используют частные пулы адресов IPv4 для назначения пользовательским устройствам и предоставления Интернет подключения через NAT и брандмауэры, обеспечить доступ к услугам на устройстве пользователя вне его местной линии не является тривиальной задачей. Одним из возможных решений этой проблемы будет использование IPv6, но мобильные операторы еще не полностью перешли на IPv6. И даже тогда маловероятно, что операторы удалят брандмауэры, поскольку это — эффективное средство борьбы с нелегальным P2P трафиком. С учетом изложенного, простейшим протестированным вариантом для демонстрации технологии явилась инкапсуляция мобильных линий данных обоих коммуникационных устройств внутрь туннеля VPN с общим VPN сервером, тем самым располагая оба устройства в одной и той же IP-подсети. Однако, из-за чрезмерного усложнения решения и больших накладных расходов на заголовки, этот вариант мало подходит для практических решений.

При фактическом развертывании очень важно разработать способ динамического конфигурирования файервола оператора, поскольку нет достаточного числа адресов IPv4 для размещения на каждом мобильном устройстве. Наиболее разумное решение, похоже, заключается в том, чтобы позволить серверу D2D выбирать политику межсетевое экранирования непосредственно перед фактическим подключением с целью обхода функ-

ций NAT. Такое решение позволит устройствам обеспечивать прямое соединение, не прибегая к возможностям виртуальной частной сети VPN, и передавать при этом данные без заголовка с очень небольшим влиянием на безопасность.

4.7.2 Испытания

Рассмотренная реализация общей концепции системы D2D, обсуждавшаяся в стандартах, столкнулась с рядом практических проблем, не учтенных в стандартизирующих документах. Текущие архитектурные представления о сетях LTE не позволяют создавать функции сетевой поддержки так, как было бы самым естественным с технической точки зрения, и вместо этого приходится искать обходные пути для адаптации существующих систем технологий. Далее рассмотрим ключевые решения, принятые на каждом этапе процесса создания и испытаний, и опишем процесс создания экспериментальной сотовой сети на базе департамента телекоммуникаций Технологического Университета Брно (BUT) в Чехии. BUT LTE зона — полная коммерчески пригодная версия для внедрения всех важнейших подсистем, входящих в современные сети 4G (рисункок 4.16). Тестовая версия была сконфигурирована для обеспечения необходимых услуг при доступе к сети с коммутацией пакетов, таких как VoIP поверх конвергентной инфраструктуры радиодоступа LTE и WiFi. Ядро сети EPC обеспечивает высокую скорость передачи данных с соответствующим уровнем QoS, имея возможность поддерживать при этом до 100 тыс. одновременно обслуживаемых пользователей. Для голосовых и видео услуг используется высокопроизводительная IP-мультимедийная подсистема (IMS), поддерживающая как мобильный, так и фиксированный доступ, а также соединения с существующими сетями телефонной связи и системами видеоконференции. Упрощенный набор фаз работы D2D соединений приведен на рисунок 4.17.

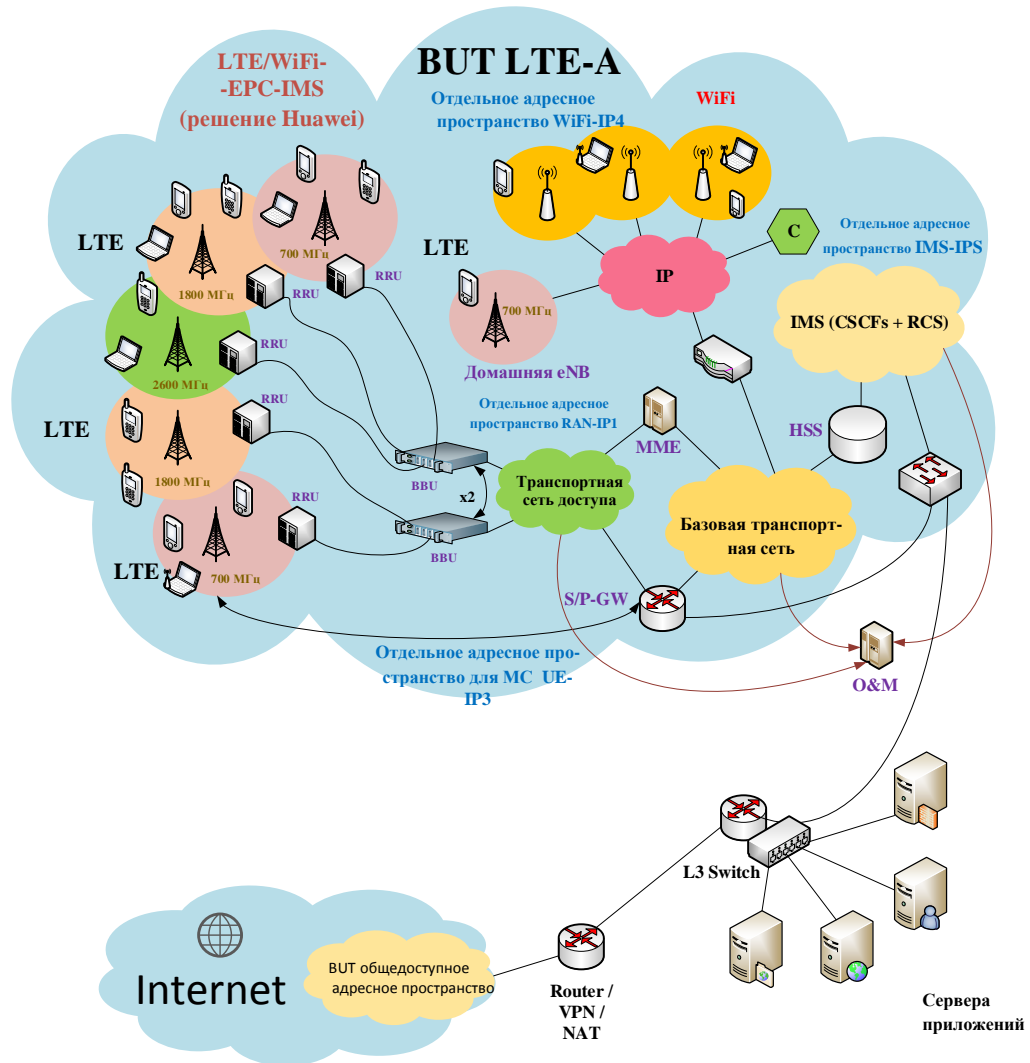


Рисунок 4.16 — Сеть LTE Технологического Университета Брно

Глобальная цель построения тестовой системы заключается в доказательстве существования синергетического эффекта при совместном использовании мобильной сети и сети D2D подробно исследуется в данном разделе. К сожалению не все механизмы поддержки со стороны LTE для D2D были доступны при поставке оборудования для тестовой зоны. Отсутствовала возможность использования сервера сервисного центра мобильной связи (e-SMLC), поэтому в качестве замены был разработан альтернативный интерфейс для получения информации о местоположении устройства с использованием Mobile Management Entity (MME). Кроме того, сервер D2D был реализован как виртуальное устройство, тогда как оконча-

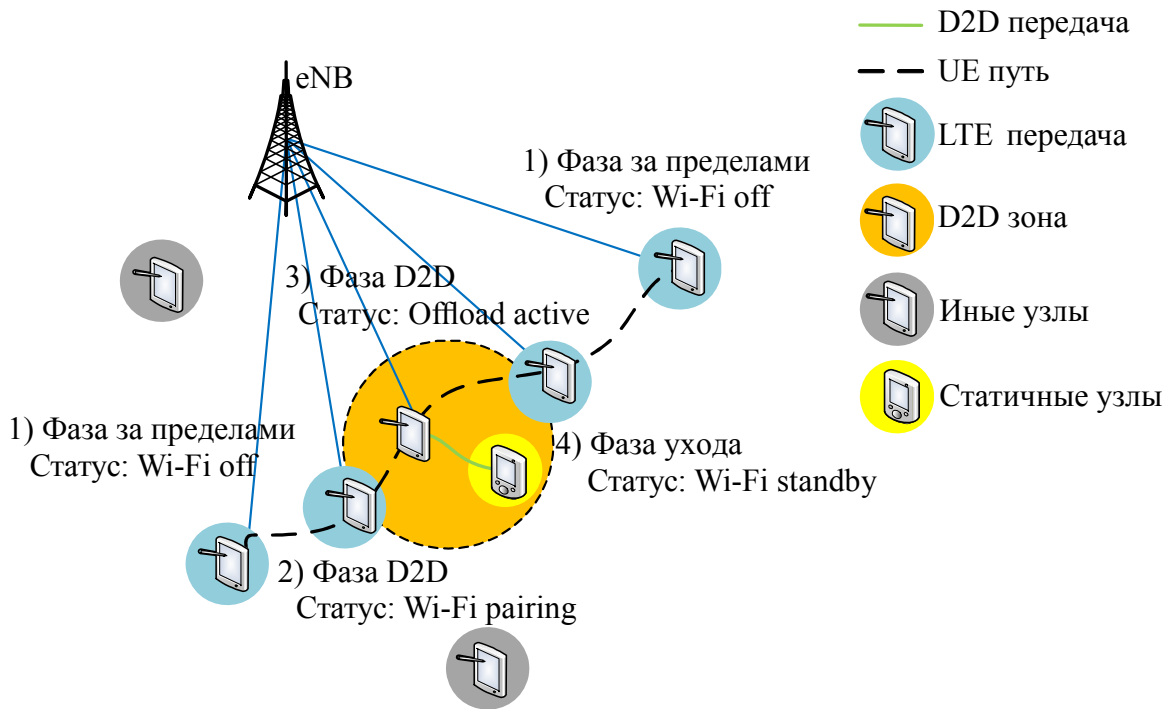


Рисунок 4.17 — Переключение между интерфейсами в D2D

тельная реализация для коммерческих сетей связи оператора должна использовать более надежную технологию.

Методология оценки эффективности. На самом деле достаточно сложно определить методологию измерений на тестовой сети так, чтобы все хорошо функционировало в реально создаваемых сетях D2D. Это, в первую очередь, требует организовать управление нагрузкой сотен пар соединений D2D, когда сотни клиентов используют в работающих телефонах пользовательское программное обеспечение. Поэтому, основные цели оценки были ограничены так, чтобы дополнить те результаты, которые были получены ранее при аналитическом и имитационном моделировании. Поскольку пропускная способность радио линий легко тестируется, наибольший интерес вызывали те ситуации, когда возникали локальные перегрузки, что потенциально могло бы уменьшить связность для D2D-технологий в будущих реальных сетях. Установление соответствующих пределов производительности и ограничений для технологии D2D и определение того, какие

услуги могут быть наиболее подходящими для такого взаимодействия в обозримом будущем, является одной из важнейших целей создания тестовой зоны. Для достижения этих целей следующие процедуры измерений были разработаны в данном подразделе.

Одним из важнейших параметров при установлении соединений D2D является длительность интервала времени занятия линии. Этот параметр имеет решающее значение, так как более низкое значение существенно снижает нагрузку в инфраструктуре LTE RAN, перемещая поток на альтернативную линию, а также улучшает другие параметры, такие как энергетическая эффективность и пользовательский опыт. Время ответа при этом может быть декомпозировано на несколько компонентов, как это описано ниже. Из-за небольшого размера сообщения, всего несколько десятков байтов, другие требования к QoS для сигнализации D2D тривиальны и являются такими же, как и для любой сети доступа и, следовательно, не рассматриваются как ключевые показатели эффективности для этой технологии. Иные аспекты производительности линии D2D, которые напрямую не связаны с инфраструктурой мобильной сети, выходят за рамки рассматриваемых оценок.

Изучая протокол для взаимодействия D2D, можно установить, что ключевым показателем является время отклика сервера, которое существенно зависит от нагрузки на соту. Численные результаты представлены на рисунок 4.18. Можно ясно видеть, что задержка на самом деле не достигает предельных значений, если сота не загружена на 100%, но когда это происходит, вариация задержки растет до недопустимых значений. На практике система работает неудовлетворительно, если задержка превышает приблизительно 100 мс, что означает, что сигнальный трафик D2D должен получить приоритет в планировщике LTE, если это возможно. Аналогично, было произведено измерение задержек при различных нагрузках на соту, а результаты приведены в таб. 4.4.

Результаты оценки качества восприятия QoE. С учетом приведенных выше показателей эффективности можно сделать следующие выводы по проблемным вопросам реализации WiFi Direct с сетевой поддержкой

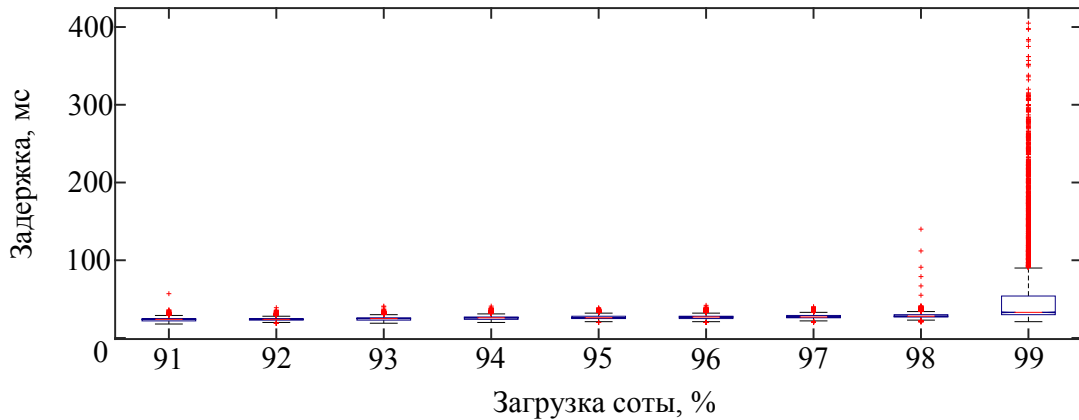


Рисунок 4.18 — Измерение задержки сервера

Таблица 4.4

Основные параметры для имитационного моделирования

Нагрузка	10%	50%	90%	99%
RTT, мс	18	25	27	60
Фаза 2, мс	36	50	54	120
Фаза 3, мс	750+36	850+50	1,000+120	1,150+120
Фаза 4, мс	9	13	13	30
Полное время, мс	849	988	1,148	1,480

LTE. Согласование соединения D2D как такового является основным узким местом в системе, поскольку это может включать в себя множество RTT циклов. В загруженной сети любая связь с сервером D2D может потребовать значительного времени, когда сигнальный трафик выбирает такой же канал, как канал с трафиком данных при отсутствии поддержки в мобильном OS API. Если пользователи высоко мобильны, отслеживание их местоположения становится все более неточным, что приводит к ложному уведомлению о близости пользователя или, наоборот, позднему установлению соединения из-за чего оно может быть потеряно. Несмотря на это, можно с уверенностью утверждать, что предложенный механизм D2D с поддержкой со стороны сети вполне пригоден для мобильных устройств, движущихся со скоростью пешехода.

Принимая во внимание, что может существовать и некоторая дополнительная задержка, своевременное установление соединения даст пользователям уверенность в их контакте, в том числе и в случае, даже когда они

перемещаются относительно друг друга. Можно сделать вывод о том, что в зависимости от качества сотовой связи полезность соединений D2D может отличаться. Однако даже с использованием современных технологий LTE, задержки передачи информации могут быть до 10 раз больше, чем при WiFi, что означает, что технология D2D поверх WiFi Direct является чрезвычайно конкурентоспособным решением.

4.7.3 Перспективы внедрения

Для операторов мобильной сети взаимодействие D2D может стать весьма эффективным механизмом для выгрузки трафика из базовой сети, а также для обеспечения поддержки социальных сетей посредством локализации предоставления услуг. Выше была рассмотрена реализация прототипа технологии D2D с поддержкой сети, определены основные проблемы и потенциальные выгоды от прямого взаимодействия между мобильными устройствами. Далее рассмотрим наиболее важные выводы и уроки из проведенных экспериментальных исследований.

1. Успешная интеграция D2D и обеспечение совместимости с существующей архитектурой 3GPP LTE доказывает, что нет технических проблем, которые препятствовали бы провайдерам предлагать приложения D2D для своих клиентов. Более того, некоторые из них могли существенно улучшить существующую инфраструктуру и предоставлять такие новые услуги, которые были бы невозможны без необходимости непрерывного отслеживания местоположения с помощью глобального позиционирования.
2. Успешное создание системы D2D с поддержкой от сотовой сети с использованием платформы Android доказывает, что системы управления, эксплуатации и обслуживания можно будет достаточно просто реализовать для требуемых протоколов управления. Только некоторые платформы, которые не реализуют стандартизованные стеки, могут столкнуться с определенными трудностями с

процедурами введения дополнительных маршрутов, необходимых для управления трафиком.

3. Качество восприятия для D2D соединений будет превосходить лучшие сотовые соединения в пределах разумных расстояний между пользователями D2D (менее 50 метров в городских условиях). При этом WiFi Direct позволяет передавать потоковое видео высокой четкости, а также приложения реального времени с высокой энергетической эффективностью.
4. Отсутствие возможностей внешнего подключения и политик экранирования, имеющихся в распоряжении операторов мобильной связи, в существенной степени ограничивает доступность соединений D2D там, где еще пока есть IPv4, поскольку число мобильных устройств больше, чем доступных IPv4 адресов. Однако эти ограничения могут быть достаточно эффективно обойдены умным конфигурированием сервера D2D и экранов.
5. Успех всей концепции D2D во многом зависит от поддержки оператором позиционирования в сети, а также от готовности к межоператорскому взаимодействию. Это, как обычно, зависит от как соответствующих стандартов 3GPP.

В целом можно сказать, что проблемы, выявленные в ходе реализации прототипа выгрузки трафика в D2D при сетевой поддержке можно решить оперативно при внедрении на сетях у операторов связи. Поскольку требования пользователей по новым услугам все возрастают, можно утверждать, что разработанная в диссертации система D2D с сетевой поддержкой найдет в той или иной форме широкое применение на сетях связи при внедрении сетей четвертого и, в особенности, пятого поколения.

4.8 Дальнейшее развитие D2D

Одним из ключевых элементов в диверсификации топологии в сотовой сети при использовании технологии D2D — это возможность непосред-

ственного (прямого) взаимодействия пользователей друг с другом, когда в то же время существует управление со стороны сотовой сети. Сочетание высокой производительности, энергетической эффективности и жесткой системной интеграции в качестве целей разработки системы D2D представляет собой исключительно большую область исследований и разработок, где пока, к сожалению, достаточно мало успешных работ.

Это исследование подчеркивает необходимость разработки *нового набора технологий и протоколов, которые сделают сеть D2D неотъемлемой частью современной инфраструктуры сотовых сетей.*

Во введении были рассмотрены несколько аспектов проблемы, что вместе с исследованием особенностей такой системы продемонстрировало, что система сотовая сеть — D2D WiFi имеет существенный потенциал для обеспечения увеличения производительности существующих и будущих беспроводных сотовых сетей.

Предложения по архитектуре для решения D2D на базе 3GPP LTE и технологии IEEE 802.11 следуют лучшим известным практикам. При этом разработанная архитектура предоставляет различные возможности для интеграции с существующими социальными сетями и платформами распределения контента, обеспечивая требуемое масштабирование.

Характеристики предложенной архитектуры были подтверждены при комплексном использовании аналитического и имитационного моделирования для оценки эффективности на основе моделирования. Это дало нам глубокое понимание пределов функционирования сети D2D для нелицензионного диапазона.

С практической точки зрения, представленные в диссертационной работе исследования, сыграли ключевую роль в нескольких демонстрационных представлениях коммерческого уровня, а также в создании экспериментальной сети LTE. Были продемонстрированы возможности совместимости с существующим программным и аппаратным обеспечением, открыв тем самым перспективы для будущих коммерческих проектов.

Подводя итог, можно сказать, что набор решений, разработанных в данной диссертационной работе, показал важность технических исследо-

ваний в области технологии D2D с поддержкой со стороны сотовой сети в нелицензируемом спектре

4.9 Выводы

1. Анализ возможностей технологии взаимодействия устройство-устройство D2D показал, что функциональное расширение возможностей технологии D2D может быть достигнуто за счет использования поддержки со стороны сети, т.е. при совместном использовании ресурсов D2D и сотовой сети.
2. С учетом функционирования устройств D2D, как правило, в условиях сетей сверхвысокой плотности и в нелицензируемом спектре, определены основополагающие теоретические зависимости между плотностью передающих устройств, размещенных на некоторой плоскости и интерференционным расстоянием для взаимодействующих устройств D2D. При этом также получена зависимость математического ожидания и среднеквадратического отклонения величины отношения сигнала к интерференции SIR от параметра распределения дистанции до интерферирующего передатчика.
3. Предложена новая архитектура сетей связи следующего поколения с широким использованием технологии D2D, отличающаяся от типовой архитектуры 3GPP единственным дополнительным объектом — сервером D2D, находящимся в EPC фрагменте сети. Такое размещение позволяет серверу D2D связываться с центром местоположения (SMLC), чтобы позиционировать пользовательское оборудование, а также эффективно взаимодействовать с серверами приложений внешних провайдеров и операторов.
4. Разработаны модели и методы выгрузки трафика в гетерогенных сетях, использующих для выгрузки трафика технологию D2D на базе WiFi Direct с сетевой поддержкой, обеспечивающие удвоение

пропускной способности соты и повышение энергоэффективности до 6 раз.

5. Для натуральных экспериментов была использована модельная сотовая сеть на базе департамента телекоммуникаций Технологического Университета Брно (BUT) в Чехии. BUT LTE зона полностью коммерчески пригодная версия для внедрения всех важнейших подсистем, входящих в современные сети 4G. Сервер D2D был реализован как виртуальное устройство. При натуральных экспериментах наблюдался значимый синергетический эффект при совместном использовании мобильной сети и сети D2D.

Раздел 5. Разработка моделей и методов распределения ресурсов для доступа данных малого размера в гетерогенной сети LTE для приложений Интернета вещей

5.1 Введение и предпосылки

Межмашинные коммуникации (анг. МТС) (или межмашинное взаимодействие (machine-to-machine)) является высокоэффективной технологией Интернета Вещей (IoT), которая обеспечивает, наряду с интеллектуальными машинами, полностью автоматическую генерацию, обмен и обработку данных, а также способна к самоприведению в действие без человеческого вмешательства полностью либо с минимальным уровнем вмешательства [206; 207]. В связи с быстрым ростом по типу и по числу встроенных систем, межмашинное взаимодействие имеет потенциал для того чтобы в корне изменить автоматическое беспроводное подключение в вертикальных приложениях индустриальной и аграрной автоматизации, транспортировки, в области наблюдения, здравоохранения, контроля скота, систем интеллектуального учета и многих других сферах [208; 209]. Таким образом, ожидается, что к 2020 году, M2M, уже наблюдающий ежегодный рост в 20%, перерастет в беспрецедентно большой рынок с доходами более триллиона долларов [210].

С учетом вышеизложенного, становится более важным гарантирование обеспечения M2M эффективной поддержки различных рыночных сценариев, и предоставление доступного беспроводного подключения при низком потреблении энергии и с приложением умеренных усилий по крупномасштабному развертыванию. Однако, как показывает практика, подключение огромного числа машин с различными функциями, оставленных без присмотра, может оказаться затруднительным главным образом благодаря уникальным характеристикам и требованиям межмашинного взаимодействия. Действительно, в то время как обычные беспроводные системы связи исторически оптимизировались для ориентированных на человека комму-

никаций, с соответствующим акцентом на сеансы передачи голоса/видео и данных, передачи M2M, как правило, предназначены для нечастых и небольших данных, зачастую терпимым к задержке образом. По этой причине беспроводная индустрия должна была принять решительные меры для адаптации существующих технологий радио доступа к коммуникациям M2M путем улучшения их эффективности производительности и надежности. В частности, в результате недавней деятельности 3GPP по стандартизации и связанных научно-исследовательских работ, в мобильные сотовые системы дальнего действия (например, HSPA и LTE) были внесены многочисленные усовершенствования. Связанные с этим изменения варьируются от схем защит от перегрузки и легких процедур сигнализации до эффективных механизмов передачи небольших данных и механизмов расширения покрытия, а также передовых подходов к распределению радиоресурсов с учетом M2M. Кроме того, эволюционировали многочисленные маломасштабные радио технологии для беспроводных сенсорных сетей, теперь уже способные к конкуренции за M2M нишу, однако, ярко-выраженный лидер по-прежнему отсутствует. К таким технологиям относятся ZigBee, 6LoWPAN, IEEE 802.15.4, WiFi, BLE и множество проприетарных беспроводных решений.

Учитывая, что связь в лицензированных сотовых диапазонах является дорогой и то, что текущие маломасштабные решения для нелицензированного диапазона потерпели неудачу в предоставлении эффективной поддержки широкого разнообразия вариантов использования межмашинного взаимодействия с разнообразными операционными требованиями, беспроводное общество искало новую эффективную M2M-ориентированную технологию радио доступа. В этой связи вскоре было выяснено, что благоприятные свойства распространения спектра ниже 1 ГГц могут предоставить улучшенную дальность связи в сравнении с, например, традиционными протоколами WiFi, работающими на частоте 2,4 и 5 ГГц [211; 212], как представлено на рисунке 5.1. Однако, доступный спектр ниже 1 ГГц в нелицензированных ISM полосах чрезвычайно скуден и, следовательно, требует тщательного проектирования системы [213]. Анализ поведения каналов при более высоких частотах доступен в [214]

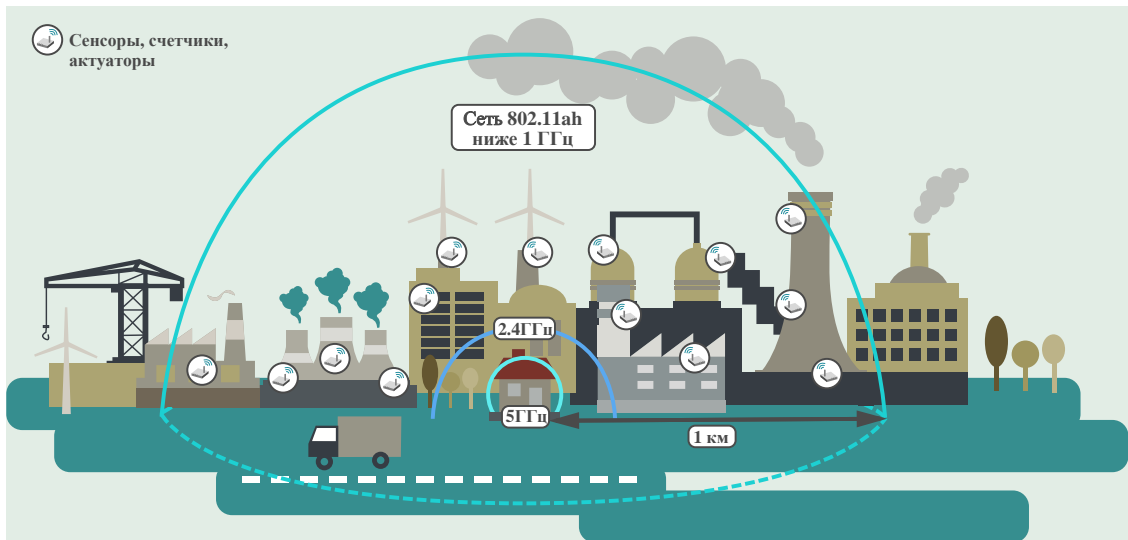


Рисунок 5.1 — Сценарий функционирования МТС

Учитывая все выше сказанное, после изложения главных целей и технических задач нового проекта IEEE 802.11ah [215], работа по стандартизации соответствующей исследовательской группы (TGah) началась в ноябре 2010. Запланированная технология, которая в основном основана на версии стандарта IEEE 802.11ac с пониженной тактовой частотой, в настоящее время находится на завершающей стадии подготовки к развертыванию недорогого и крупномасштабного массового M2M подключения с высокой спектральной и энергетической эффективностью. В настоящее время тысячи M2M устройств, уже могут быть найдены в густонаселенных городских районах, интеллектуальных электросетях [216], и аэропортах [217], что потребовало соответствующих протокольных изменений.

К счастью, технология IEEE 802.11ah не нуждается в поддержке обратной совместимости с другими представителями семейства IEEE 802.11 (как обычными 802.11 a/b/g/n, так и более поздними решениями 802.11ac). Поскольку 802.11ah работает на частотах, отличных от существующих систем 802.11, TGah может позволить себе определять новые компактные форматы кадров, а также предложить более эффективные механизмы поддержки огромного числа устройств, улучшенных схем канального доступа наряду с важнейшими усовершенствованиями в области пропускной способности и сохранения энергии [218]. В результате, 802.11ah, как полагают, значительно обогатит семейство протоколов 802.11, которые уже получают все большее внимание от сотовых операторов связи, желающих внедрить

недорогое подключение в нелицензированных диапазонах, тем самым увеличивая сотовое развертывание [217].

Поскольку новая технология 802.11ah ожидается к выходу на рынок в 2018, она уже привлекла внимание как индустрии, так и научной сферы, судя по количеству исследовательской литературы по данной теме. В соответствующих публикациях проводится обзор характеристик физического уровня 802.11ah при сельском [219] и городском [220] вариантах использования, предоставлен расчет энергетического бюджета канала и определение вероятности отказа в работе. По мимо этого, публикации охватывают важнейшие проблемы, которые возникли в результате деятельности по стандартизации рабочей группы TGah [220] и предлагают соответствующие улучшения по проектированию MIMO-OFDM [221], а также описывают работу прототипов систем [217]. Системная производительность технологии 802.11ah также была сравнена с производительностью Zigbee, для выявления важных областей для улучшений [222; 223].

Кроме того, в [224] и [225] был приведен анализ емкости системного уровня, в последующих работах представлен исчерпывающий обзор технологии 802.11ah [226–228], а также в [229; 230] предложены новые протоколы канального доступа с избеганием/ разрешением коллизий, и для плотных развертываний M2M [231]. Особый интерес представляет работа по улучшению доступа на основе конкуренции с группированием M2M устройств [232] и последующее управление производительностью этих групп [233]. В дополнение, были разработаны эффективные коды проверки на избыточность для 802.11ah [234].

Гораздо важнее то, что многие недавние публикаций были сосредоточены на повышении энергетической эффективности относительно технологии 802.11ah ближнего диапазона, от схем аккумуляирования [235] и сохранения [236] энергии до усиленных режимов [237; 238] и механизмов [239; 240] сохранения энергии. Как следует из нашего анализа, существующие исследовательские работы до сих пор были направлены на физический уровень и другие конкретные улучшения, а также оценки систем полной буферизации.

Тем не менее, исчерпывающий анализ системного уровня 802.11ah, принимая во внимание M2M-специфический трафик, динамическую и реалистичную конкуренцию между огромным числом устройств, на данный момент отсутствует в существующей литературе. Традиционные методы оценки семейства протоколов 802.11 в основном сфокусированы на насыщенном анализе благодаря математической простоте. Например, хорошо известная работа Bianchi и Tinnirello [241] также придерживается аналогичного подхода, основанному на Марковских цепях, но в основном сфокусирована на случае для насыщенного трафика. Существуют также другие работы по анализу полной буферизации [242; 243], но академическая сфера все еще сталкивается с отсутствием результатов исследований по режиму низко-частотного поступления. Фактически, эти результаты весьма ограничены и главным образом сфокусированы на анализе пропускной способности [244].

После обширного обзора литературы мы обнаружили только несколько работ, которые адресованы анализу работы IEEE 802.11ah в ненасыщенном режиме. В [245], авторы подробно останавливаются на модификации механизмов канального доступа для сценариев неполной буферизации, и предлагают простую модель, основанную на методе ограничения рабочего цикла. Полная аналитическая модель для ненасыщенного сценария не представлена в данной работе, но постановка проблемы сформулирована ясно. Еще одна работа, которая принимает внимание поведение в ненасыщенном сценарии раннее оговоренного протокола [246], в основном сфокусирована на проблеме скрытых узлов и предлагает группированное решение, основанное на генерации пакетной скорости.

В этой работе мы намерены преодолеть обозначенный пробел путем предложения новой математической платформы для анализа систем, основанных на 802.11ah, с высокой плотностью узлов, принимая во внимание ненасыщенный тип трафика. Остальная часть статьи организована следующим образом. Во-первых, в следующей секции будет предоставлен обзор возможных приложений 802.11ah. Далее, основные функции протокола 802.11ah описаны в разделе III. После этого обсуждается работа протокола в характерном M2M сценарии. Далее представлена наша системная модель

и подчеркнуты главные аналитические предположения. Модель сопровождается строгими аналитическими расчетами, в то время как в заключительном разделе описана используемая среда моделирования и сравнение полученных результатов с аналитическими. Заключение сконцентрировано на анализе проведенного опыта.

5.2 Перспективы использования стандарта IEEE 802.11ah

Данный раздел представляет вводное видение потенциальных реальных приложений, основанных на IEEE 802.11ah, а также соответствующие преимущества их использования относительно поведения сетевого трафика. Обсуждаемый протокол разрабатывался с целью удовлетворения требований большого числа узлов, и, следовательно, одной из его главных моделей внедрения являются беспроводные сенсорные сети. Другими приложениями являются транспортные сегменты сенсорных и "измерительных" сетей связи, а также расширенное покрытие для протоколов семейства WiFi [226]. Далее, мы подробно остановимся на насыщенных и ненасыщенных сценариях.

Насыщенный сценарий может рассматриваться для приложений, требующих большой емкости сети, т.е. приложений потоковой передачи аудио/видео [247], приложения разгрузки сетей [248], приложения для чрезвычайных ситуаций [249], и т.д. В этом случае, предполагается, что каждая станция, вовлеченная в коммуникационный процесс, имеет полный буфер, таким образом, что пакет готов для передачи в любой момент [250]. Как только пакеты находятся в конфликте, будучи переданными по беспроводной среде, канал становится насыщенным и задержка передачи стремительно возрастает. Эта работа рассматривает обсужденный случай отдельно от непосредственного представления целей проектирования протокола, как это объяснено ниже.

С другой стороны, практические приложения IEEE 802.11ah связаны наиболее всего с ненасыщенными сценариями, когда беспроводной узел

передает малые пакеты в течение относительно долгого периода времени, что в основном отражают сегодняшние операции M2M [211; 239]. В этом случае ожидается, что пакетный буфер будет пустым в течение большей части времени наблюдения - такой случай относится к ненасыщенному сценарию. Основные соответствующие приложения режима неполного буфера приведены ниже.

- *Интеллектуальное измерение:* ожидается, что IEEE 802.11ah улучшит производительность традиционных беспроводных технологий ближнего действия [223], таких как ZigBee, NFC, Bluetooth и т.д. Главные области применения: контроль окружающей среды, умные сети электропитания, промышленная автоматизация и здравоохранение. Учитывая необходимость обслуживания большого количества устройств связи, рассмотренная технология предлагает новые способы уменьшения числа коллизий и увеличению энергетической эффективности [213].

- *Сотовая разгрузка и расширенное покрытие:* Одним из потенциальных преимуществ рассматриваемой технологии является дальность действия до 1 км благодаря использованию низко-частотного диапазона [216]. Эта функция позволяет предоставить дополнительные возможности обычным пользователям сотовой связи, предлагая им определенный уровень обслуживания в экстремальных случаях. В отличие от высокой пропускной способности протоколов IEEE 802.11ac и 802.11n, более низкая скорость передачи данных в 802.11ah может быть применима в случаях национальной и общественной безопасности [251].

- *Транзитные сети:* Простой системный дизайн IEEE 802.11ah позволяет связать низко-частотные пункты агрегирования с узлами, подключенными к традиционной беспроводной сети [252]. Например, IEEE 802.11g способен обеспечить подключение между точкой доступа и удаленным сервером, в то время как IEEE 802.11ah может предложить беспроводную транзитную связь для размещения агрегированного трафика, генерируемого конечными датчиками [253].

Ожидается, что приведенный выше список приложений, использующих технологию IEEE 802.11 ah, будет расти с течением времени. Согласно

имеющейся в настоящее время литературе, перспективные приложения в основном представляют случаи ненасыщенного поведения трафика.

5.3 Обзор основных технических характеристик IEEE 802.11ah

В этом разделе описываются основные операции уровней MAC и PHY, введенные спецификациями IEEE 802.11ah, основанной на последнем проекте стандарта в [254]. В частности, здесь представлены новейшие характеристики протокола IEEE 802.11ah.

5.3.1 Характеристики физического уровня IEEE 802.11ah

Важно отметить, что многие из предложенных модификаций физического уровня находят свои корни в предыдущих версиях IEEE 802.11. Более того, ожидается, что новая технология IEEE 802.11ah будет работать в ISM диапазоне ниже 1 ГГц. Учитывая, что доступность спектра в этом диапазоне варьируется от одной страны к другой, деление каналов для IEEE 802.11ah в настоящее время определяется как 863-868 МГц в Европе, 916,5-927,5 МГц в Японии 755-787 МГц в Китае, 917.5-923,5 МГц в Южной Корее, 866-869 МГц и 920-925 МГц в Сингапуре, а также 902-928 МГц в США. Для более подробной информации см. текущую версию стандарта [254].

Технология IEEE 802.11ah может быть рассмотрена как низкочастотная версия спецификации IEEE 802.11ac, которая является в свою очередь расширением стандарта IEEE 802.11n. В основном, IEEE 802.11ah определен как стандарт с пониженной в 10 раз тактовой частотой от существующей пропускной способности канала IEEE 802.11ac, т.е. в настоящее время предусмотрены каналы 2 МГц, 4 МГц, 8 МГц, и 16 МГц. Для уве-

личения дальности передачи IEEE 802.11ah дополнительно вводит новый режим связи, используя ширину полосы частоты канала в 1 МГц.

5.3.2 IEEE 802.11ah характеристики уровня контроля доступа к среде

Если выделить главное, IEEE 802.11ah предоставляет ряд функций уровня MAC для полного удовлетворения требований сегодняшнего высокоплотного развертывания M2M коммуникаций. Основные обновленные аспекты MAC-уровня могут быть сгруппированы в четыре различных класса, обсуждаемые далее.

Улучшенный режим сохранения энергии

IEEE 802.11ah документация определяет две стратегии энергосбережения, называемые рабочими режимами TIM и не-TIM. Соответственно, устройства M2M можно назвать неTIM и TIM станциями (STAs).

В случае режима не-TIM, устройство IEEE 802.11ah не получает информационный элемент (IE) карты индикатора трафика (TIM) от навигационных/бикон кадров, которые регулярно широковещательно транслируются точкой доступа (AP). Обычно IE TIM используется точкой доступа, чтобы указать, имеются ли буферизованные нисходящие (DL) данные для определенной STA. Не-TIM STA должен отправить PS-Poll кадр к точке доступа, таким образом запрашивая доставку любых ранее-буферизованных нисходящих данных. TIM STA, однако, получают периодически информационные элементы TIM IE от навигационного/бикон кадра, чтобы обнаружить, имеет ли точка доступа какие-либо буферизованные данные для них. Когда такой STA обнаруживает, что точка доступа имеет данные, на-

ходящиеся в процессе обработки, STA передает кадр PS-poll для запроса доставки нисходящего (DL) трафика.

IEEE 802.11ah представляет соответствующие улучшения механизма PS-poll, способных извлечь выгоду из данных схем экономии электроэнергии. Кроме того, это увеличивает время, в течение которого AP не разъединяет STA в случае, если AP не принимает никакого кадра от него, а точнее, увеличивается время максимального периода простоя. Кроме того, схема сигнализации Null data packet (NDP) разработана для уменьшения времени, которое определенный STA может потратить в режиме ожидания, чтобы проверить, есть ли входящие данные. Данное свойство разрешает любому устройству IEEE 802.11ah, которое ожидает входящие кадры, минимизировать продолжительность пробуждения при помощи кадра сигнализации, который значительно короче по сравнению с навигационным/бикон кадром.

Второй метод экономии электроэнергии, введенный IEEE 802.11ah, относится к целевому времени пробуждения (TWT), позволяющее определенному устройству M2M спать в течение особых промежутков времени и затем просыпаться в пред-запланированные моменты. Это сделано для того, чтобы разрешить обмен информацией со связанной точкой доступа в определенное время. Эти моменты согласованы заранее между двумя объектами связи.

Поддержка большого числа устройств

Текущая версия спецификации определяет TИМ как битовый массив, имеющий функцию уникального отображения с Идентификатором Ассоциации STA-станции (AID). Мы предполагаем, что максимальная полезная нагрузка пакета IE равняется 256 байтам [222; 255], следовательно, в общей сложности приблизительно могут быть поддержаны две тысячи AID. Однако, как предусмотрено требованиями IEEE 802.11ah, в некоторых сценариях должны быть размещены до 6 000 устройств M2M.

Иерархические функции битового массива TIM и сегментации TIM были приняты в спецификациях IEEE 802.11ah для поддержки большего количества сетевых узлов. Иерархия TIM состоит из 4 страниц, каждая из которых содержит 32 блока. Каждый блок далее разделен на 8 подблоков по 8 битов каждый. Любой STA определен индексом страницы, индексом блока на странице, индексом под-блока в пределах блока и битовой позиции в пределах подблока. Далее, адрес STA закодирован в его AID. Чтобы эффективно закодировать TIM, IEEE 802.11ah определяет три процедуры кодирования, а именно, Block Bitmap, Single AID и Offset Length Bitmap Mode. Для дальнейшего повышения производительности системы для сценариев, где большое количество устройств MTC передают небольшие пакеты данных, стандарт IEEE 802.11ah позволяет уменьшить количество служебных данных в заголовках MAC посредством сжатия заголовка. С этой целью был определен сокращенный формат заголовка MAC; размер заголовка MAC для типичных 802.11 кадров данных составляет 28 байтов, который был уменьшен до 18 байтов в IEEE 802.11ah. Наконец, IEEE 802.11ah вводит несколько кадров NDP, также известных как сокращенные кадры ACK, сокращенный блок ACK, сокращенный CTS и сокращенный PS - poll, и т.д. Цель их состоит в том, чтобы уменьшить связанную потребляемую мощность и загрузку среды.

Расширенные механизмы канального доступа

В современных системах WLAN, где доступ к среде в обычно основан на конкуренции, эффективность механизмов канального доступа зависит в значительной мере от пользовательского конкурентно-способного уровня. В технологии IEEE 802.11 основной механизм канального доступа использует схему расширенного распределенного канального доступа (EDCA), которая эффективно работает с относительно небольшим количеством STA и обеспечивает стохастическое дифференцирование качества обслуживания (QoS). Однако, новые M2M усложняют ситуацию для сетей следующе-

го поколения IEEE 802.11 благодаря огромному числу устройств, которые нуждаются в поддержке, и, следовательно, высоких уровнях конкуренций. В IEEE 802.11ah, схема EDCA была оптимизирована, для возможности предоставления сенсорным- и несенсорным- STA дифференцированного качества обслуживания (QoS). В результате, EDCA является механизмом доступа по умолчанию для спектра частот ниже 1 ГГц для STA в 802.11ah. Для того чтобы избежать массовых коллизий IEEE 802.11ah предлагает новый важный механизм канального доступа, называемый окном ограничивающего доступа (Restricted Access Window (RAW)) поверх традиционной схемы EDCA с возможностью передачи (Transmission Opportunity)(EDCA TXOP). С новым механизмом RAW, точка доступа распределяет выделенный период доступа к среде в навигационном/бикон интервале, который потом делится на один или несколько временных слотов. Присвоение слотов для определенного STA в пределах определенного RAW основано на элементе TIM в навигационных/бикон кадрах. AP может также присвоить особые слоты к отдельному STA, или к подмножеству STA, для восходящего и нисходящего трафика.

Благодаря тому, что в результате коммуникаций может появиться дополнительная служебная информация, технология IEEE 802.11ah, вводит новый протокол кадрового обмена, называемый реверсивной схемой TXOP (BDT), призванный уменьшить служебные данные протокола. По сравнению с обычным доступом, BDT использует текущий кадр данных в качестве АСК по отношению к предыдущему кадру данных. Поэтому, BDT позволяет AP и не-AP STA обмениваться последовательностью восходящих и нисходящих кадров в течение зарезервированного времени.

Эффективные схемы снижения интерференции

Одним из ключевых механизмов решения проблемы скрытых терминалов является секторизация [156;256]. Данный механизм делит естественным образом зону покрытия базового набора услуг (Basic Service Set, BSS)

на секторы, каждый из которых содержит подмножество STA. Такое разделение достигается благодаря набору антенн, или набору синтезируемых антенных лучей, для покрытия различных секторов BSS. Главной целью этого механизма является уменьшение интерференций между BSS. В спецификации IEEE 802.11ah введены два типа схем секторизаций: групповая секторизация и секторизация, основанная на TXOP.

Другой разработанный механизм — подканальная выборочная передача (Sub-channel Selective Transmission, SST) и Динамическая Операция BW, целью которого является смягчение интерференций, полученных в результате чрезмерного использования канала или как результат наложения друг на друга развертываний BSS. Основным каналом обычно используется сеть заданного BSS. Он является статичным и может не измениться на протяжении всей работы BSS. Этот общий основной канал (обычно 1 или 2 МГц) используется STA для общения с точкой доступа. Однако из-за изменения состояния канала и местоположений STA, уровни помех, испытываемых на основном канале, могут быть значительно выше по сравнению с другими доступными каналами. Следовательно, тогда SST позволит STA выбрать лучший временный канал среди множественных поддерживаемых SST рабочих каналов для связи с пиром STA.

5.4 Функционирование протокола и предложенная модель

Данный раздел основан на предыдущем с целью предоставления рассмотренных характеристик работы протокола и выявлений связанных упрощений.

5.4.1 Протокол доступа к каналу IEEE 802.11ah

Говоря конкретно о доступе к среде, мы ожидаем, что схема канального доступа BASIC (см. рисунок 5.2) будет более эффективной для большинства сценариев M2M с целью дальнейшего снижения служебной информации [257]. Кроме того, STA ниже частоты 1 ГГц использует доступ EDCA, но в нашей работе мы рассматриваем случай использования, когда существует только один тип STA и, следовательно, один тип потока трафика. Следовательно, QoS не является главной целью нашего внимания и ниже вместо EDCA ради простоты мы рассматриваем более простой механизм - распределенную координирующую функцию (DCF).

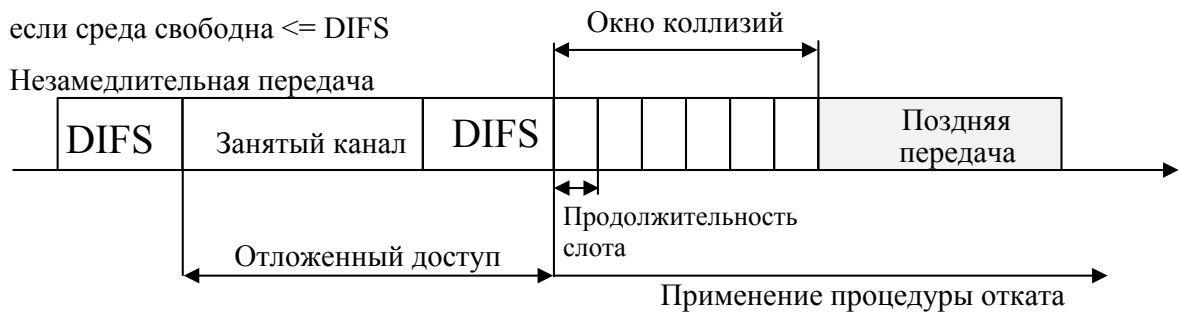
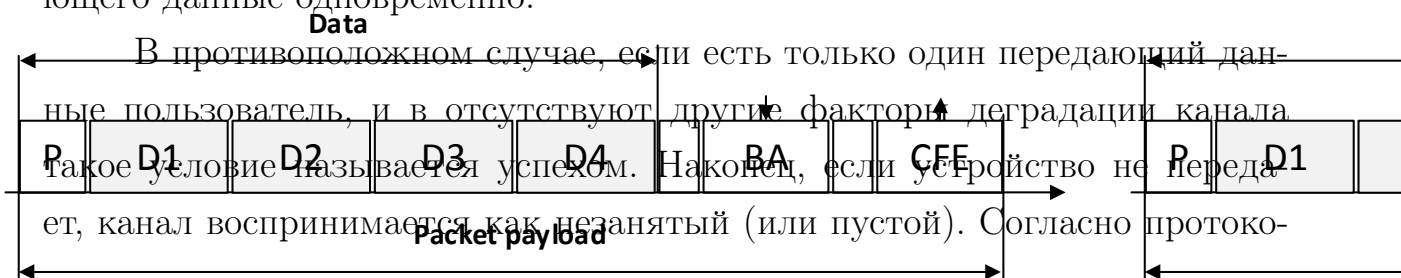


Рисунок 5.2 — Пример работы алгоритма доступа BASIC

В соответствии со стандартом для каждого устройства в системе, если в буфере появляется новый пакет данных (или агрегированный блок пакетов), и нет других передач, обнаруженных в промежутке между кадрами DCF (DIFS), передача данных будет выполнена немедленно. В противном случае доступ к среде будет отложен в соответствии с алгоритмом двойной экспоненциальной выдержки (БЕВ) [135], который используется для разрешения конфликтов, когда в сети более одного пользователя, передающего данные одновременно.



лу, для механизма BASIC в случае коллизии или повреждения каналом сначала применяется операция ВЕВ. Равномерно распределенное значение счетчика отклика (ВС) выбирается случайным образом между нулем и начальным окном коллизии W_0 .

Пока как канал бездействует, ВС уменьшается и, в конце концов, когда он достигает нуля, начинается процедура повторной передачи. Если коллизия происходит снова, текущее окно коллизий (CW) удваивается, и процедура повторной передачи повторно инициализируется столько раз, сколько разрешено счетчиком повторной передачи (РС).

Поскольку в рассматриваемой системе существует необходимость в явном учете ненасыщенного поведения трафика большого количества устройств МТС, мы разрабатываем соответствующую аналитическую модель и вводим ее далее.

5.4.2 Модель системы и основные предположения

В данном разделе описывается модель системы и ее основные предположения. Для ясности изложения мы фокусируемся на анализе одного кластера МТС и не предполагаем наличие скрытых терминалов, а также считаем канал свободным от ошибок. Важные практические расширения этой базовой модели обсуждаются в последующих разделах и [258] и [259].

Для простоты предположим, что механизм разрешения конфликтов одинаков для всех устройств МТС. Мы также предполагаем, что время начала каждого слота воспринимается всеми пользователями синхронно а общее системное время делится на очень маленькие слоты с равным размером и длительностью t_g . Эти слоты можно рассматривать как атомные интервалы времени, когда пользователь может организовать новую передачу пакетов.

Затем мы определяем продолжительность события. Типы событий $s/c/e$ и определяются следующим образом:

- Успех (s): если в системе есть только один передающий пользователь. Продолжительность такого события определяется как t_s , которая рассчитывается в соответствии со спецификациями как сумма: времени между кадрами ($DIFS$); длительности передачи данных; малого времени между кадрами ($SIFS$) и подтверждения (ACK).

- Простой (c): передача не наблюдалась, соответствующая продолжительность определяется как t_e , длительность интервала.

- Коллизия (e). Если имеется более одного одновременно передающего устройства, мы рассматриваем временной интервал t_c между началом передачи и началом следующей возможности для передачи. Он включает в себя: $DIFS$, $SIFS$, PHY длительность ожидания, длительность передачи данных и одну продолжительность простоя.

Важно отметить, что наша модель использует точное количество возможностей создания пакета для пользователя в течение каждого из возможных событий ($L_{s/c/e}$). Другими словами, мы разбиваем время работы системы на более мелкие промежутки возможности генерации пакетов t_g , которые также делят каждую из возможных продолжительностей событий описанных выше. Например, для слота успех можно выразить $L_s = t_s/t_g$. Для удобства, t_g выбирается таким образом, что $L_{s/c/e}$ – целые числа.

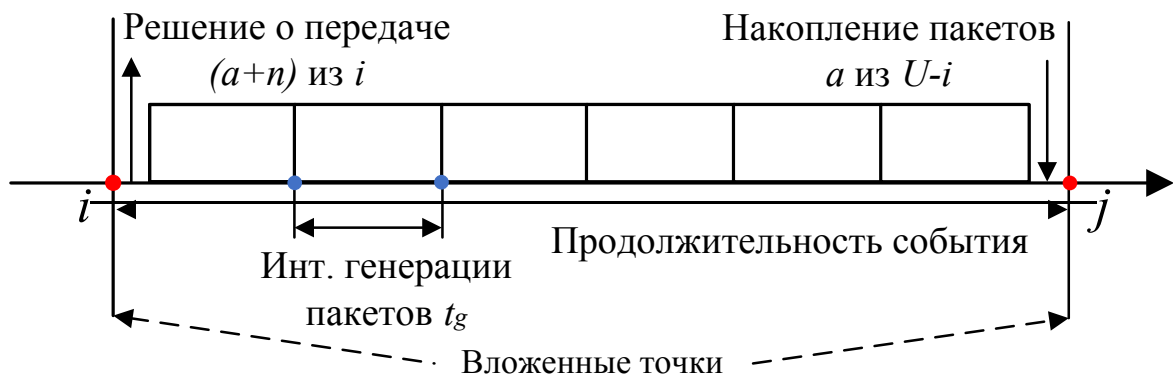


Рисунок 5.3 — Модель генерации входного потока

Рассмотрим определенное количество пользователей U , связанных с одной точкой доступа в течение всего времени работы системы. Что касается предложенной нагрузки, мы предполагаем, что время между прибы-

тием заявок имеет распределение Бернулли [260] с интервалом генерации пакетов t_g в качестве аппроксимации для сценария сети IoT уровня, рассмотренного в [261]. Этот тип потока прибывающих заявок был выбран как простой случайный процесс с дискретным временем, при котором устройство может сгенерировать пакет во время интервала t_g с вероятностью σ . После генерации пакета пользователь будет пытаться выполнить немедленную передачу, если пакет попал в пустой буфер. В противном случае, событие генерации пакета может быть ассоциировано с i ожидающими передачи пользователями. *Ожидающие* пользователи пытаются выполнить повторную передачу с вероятностью p , из-за предыдущей неудачной передачи.

Рассмотрим пример работы такой системы для одной конкретной продолжительности события и заданных количеств *ожидающих* пользователей i и j до и после соответственно (показано на рисунке 5.3). Естественно, $n = \overline{0, i}$ пользователей могут попытаться отправить свои пакеты с вероятностью доступа к каналу p . Пользователи без данных в буфере (a из $U - i$) могут генерировать новые пакеты в течение каждого события. Важно отметить, что переход от i до j забытых пользователей определяется по-разному для событий успех и коллизия/пустое событие.

Во время успешного слота $a = j - i - n - 1$, $0 \leq a \leq U - i - n$. Для коллизии или пустого события $a = j - i - n$, $0 \leq a \leq U - i - n$. Мы предполагаем, что если пользователь успешно осуществил передачу, новый пакет не генерируется немедленно. Процедура генерации применяется начиная со следующего слота. Кроме того, если пакет уже создан во время текущего события, мы предполагаем, что не будет никаких новых прибытий, пока пакет данных не будет успешно передан этим пользователем – наша система работает в режиме без потерь. Это позволяет представить пользователей в виде буферов единичной длины для хранения одного пакета данных.

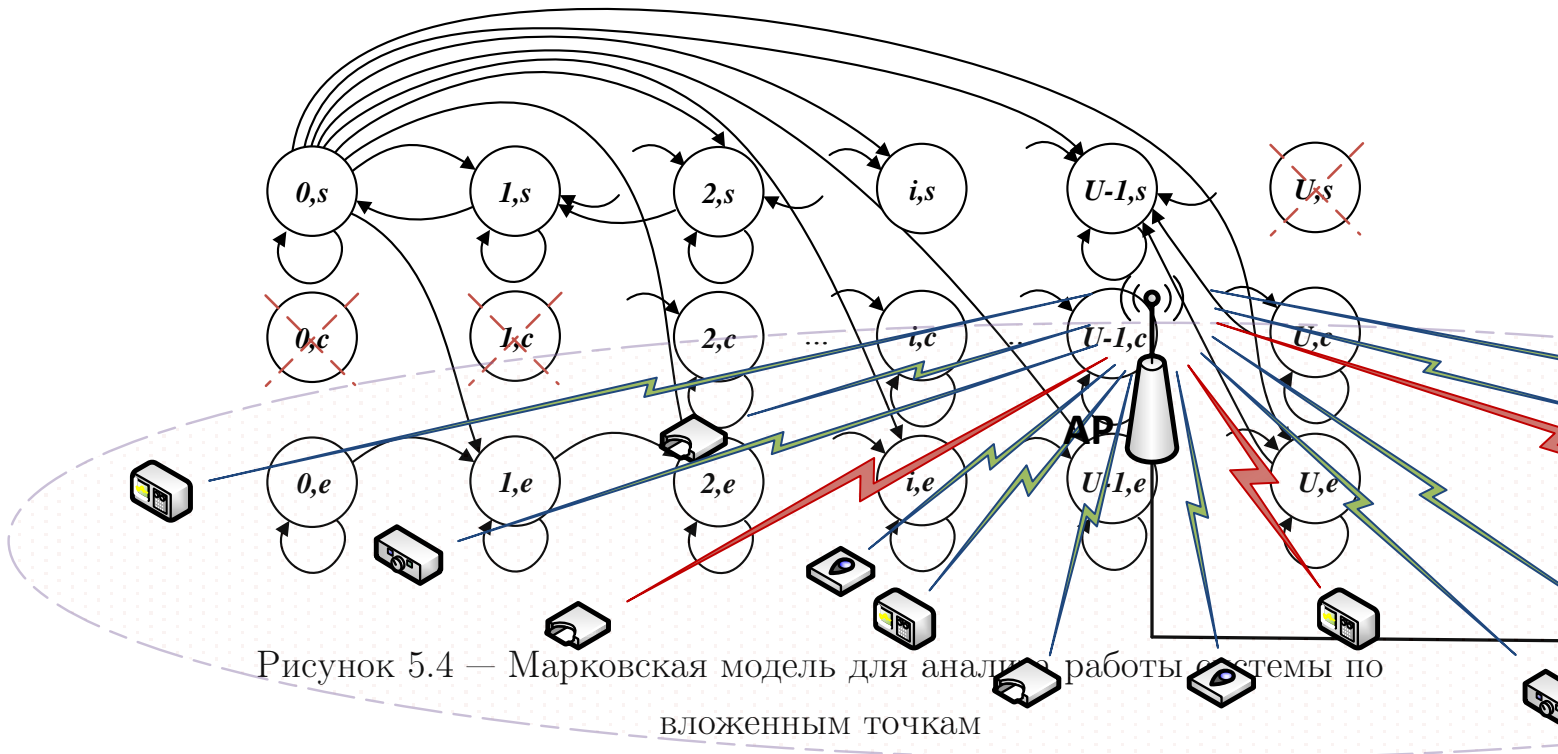
5.5 Математический анализ

В этом разделе мы подробно изложим наш аналитический подход, основанный на рассмотрении цепи Маркова в соответствующих *вложенных точках*. Далее будет выведена матрица вероятностей переходов состояний и соответствующий ей вектор стационарного распределения. Основываясь на котором, в свою очередь могут быть получены ключевые параметры стационарного режима системы, такие как, например, количество ожидающих обслуживания пользователей в системе. В качестве наших основных показателей эффективности мы выбираем пропускную способность данных (с учетом только пакетов данных здесь, за исключением накладных расходов) и задержки доступа к каналу. Важно отметить, что в то время как сама задержка может представлять меньший интерес для устройств с устойчивостью к задержкам МТС, ее можно легко преобразовать в энергоэффективность, что является одним из полезных системных параметров для любого сценария МТС.

5.5.1 Детальный системный анализ

С точки зрения передачи данных наша система интереса работает в двух альтернативных режимах: когда только что появился новый пакет – немедленная передача; или когда пакет повторно передается с использованием вероятностного доступа к каналу. Непосредственная передача и соответствующие системные события предполагают анализ с помощью двумерной цепи Маркова, которая способна учитывать прошлое количество забытых пользователей (состояние системы), как показано на рисунке 5.4.

Для точности, мы рассмотрим процесс $z(t)$, внедренный в начале каждого события (с заданной продолжительностью). Здесь состояние процесса задается как (i, x) , где i – количество забытых пользователей, а x – тип предыдущего события. Важно отметить, что в нашей системе могут воз-



никать три альтернативных события: столкновение, успешная передача и пустой слот.

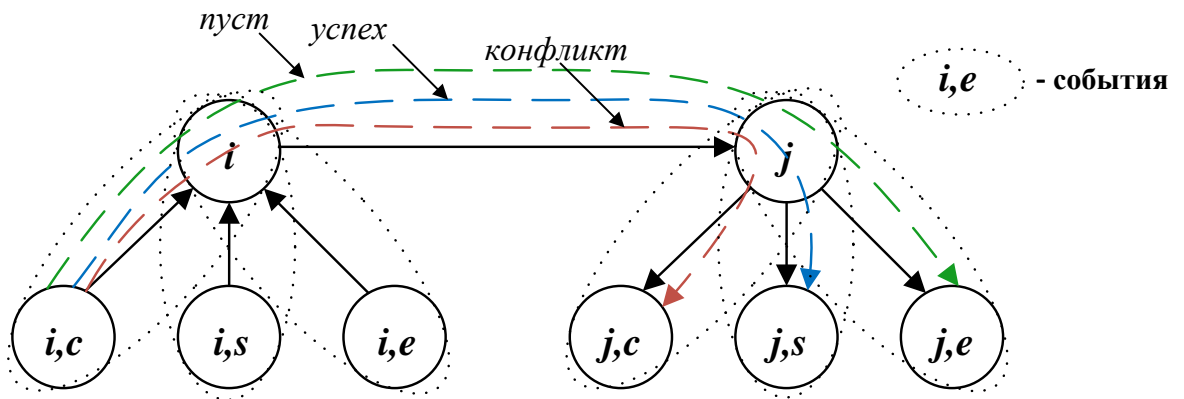
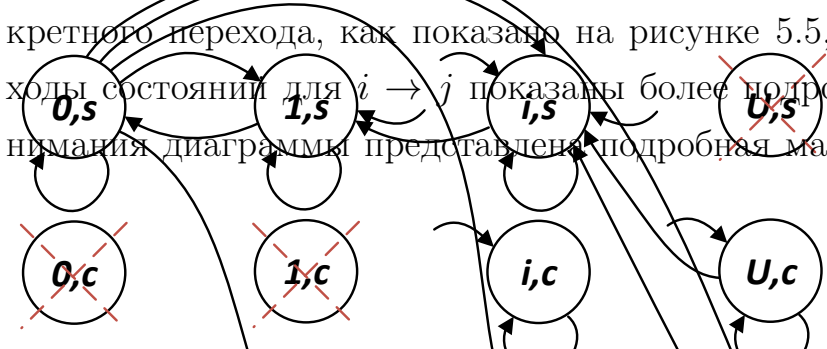


Рисунок 5.5 — Упрощенная Марковская модель для основных переходов

Для вычисления стационарного распределения $z(t)$ нам нужно установить матрицу вероятности перехода P (представленную в таблице 5.1). Здесь $P_{(i,x) \rightarrow (j,y)}$ является элементом, соответствующим вероятности конкретного перехода, как показано на рисунке 5.5, где все доступные переходы состояний для $i \rightarrow j$ показаны более подробно. Для облегчения понимания диаграммы представлена подробная матрица, в которой каждая



из ее ячеек состоит из конкретных условий между двумя конкретными состояниями.

Таблица 5.1

Переходные состояния системы

Состояние (i/j)	1	2	j	U+1
1	$a_{1,1}$	$a_{1,2}$	$a_{1,j}$	$a_{1,U+1}$
2	$a_{2,1}$	$a_{2,3}$	$a_{2,j}$	$a_{2,U+1}$
..				
i	$a_{i,1}$	$a_{i,2}$	$a_{i,j}$	$a_{i,U+1}$
..				
U+1	$a_{U+1,1}$	$a_{U+1,2}$	$a_{U+1,j}$	$a_{U+1,U+1}$

Состояние $a_{i,j}$	j,s	j,c	j,e
i,s	s_{11}	s_{12}	s_{13}
i,c	s_{21}	s_{23}	s_{23}
i,e	s_{31}	s_{33}	s_{33}

В дальнейшем мы изучаем отдельные системные переходы.

Переход $(i,s) \rightarrow (j,s)$: Например, для перехода от предыдущего успешного слота к текущему успешному слоту для хранения должна быть ровно одна передача пакетов

$$P_{(i,s) \rightarrow (j,s)} = \begin{cases} (1-p)^{j-1} p B(0, L_s | U-i), & j \leq i, \\ (1-p)^j B(1, L_s | U-i), & i > j, \end{cases} \quad (5.1)$$

где L_s – количество возможностей для создания нового пакета для пользователя во время успешного события. Здесь $B_{s/c/e}(a, x | n)$ таков, что в течение x число слотов будет равно a , а n – возможное количество пользователей для сгенерировать пакет, который можно вычислить как

$$B_{s/c/e}(a, L_{s/c/e} | n) = \binom{n}{a} \rho_{s/c/e}^a (1 - \rho_{s/c/e})^{n-a}, \quad (5.2)$$

где $\rho_{s/c/e}$ – вероятность создания пакета в течение точно выбранной продолжительности события, т. е. $\rho_{s/c/e} = 1 - (1 - \sigma)^{L_{s/c/e}}$. Для такого перехода

ясно, что $j = i - 1$, если пакет не был сгенерирован во время текущей передачи для этого пользователя, иначе $j = i$ в противном случае.

Переход $(i, s) \rightarrow (j, e)$ из успешного состояния в пустое состояние можно получить как

$$P_{(i,s) \rightarrow (j,e)} = B(0, L_s | U - i)(1 - p)^i, \quad (5.3)$$

когда ни один из пользователей не решит передать в текущем слоте.

Переход $(i, s) \rightarrow (j, c)$: последний оставшийся переход можно вычислить так:

$$P_{(i,s) \rightarrow (j,c)} = \begin{cases} B_0(1 - (1 - p)^i - ip(1 - p)^{i-1}), & k = 0, \\ B_1(1 - (1 - p)^i), & k = 1, \\ B(k, L_s | U - i), & k > 1, \end{cases} \quad (5.4)$$

где $k = j - i$, $B_1 = B(1, L_s | U - i)$, $B_0 = B(0, L_s | U - i)$.

Все следующие значения в подматрице $P_{(i,x) \rightarrow (j,x)}$ могут вычисляться аналогично, но продолжительность предыдущего слота ($L_{s/c/e}$) должна соответственно. Далее, мы вычисляем распределение стационарной системы во вложенных точках путем решения следующей системы линейных уравнений

$$\boldsymbol{\pi}^{(0)} = P^T \boldsymbol{\pi}^{(0)}, \quad \boldsymbol{\pi}^{(0)} \mathbf{e} = 1, \quad (5.5)$$

где P – матрица вероятности перехода, $\mathbf{e} = (1, 1, \dots, 1)$ – вектор единиц, а $\boldsymbol{\pi}^{(0)} = \left\{ \pi_{(i,x)}^{(0)} : x = s/c/e, i = \overline{0, U} \right\}$.

Затем мы выводим среднюю продолжительность следующего системного события для состояния (i, x) как

$$\bar{V}_{i,x} = \sum_{j=0}^U \sum_{y=c,s,e} P_{(i,x) \rightarrow (j,y)} L_y. \quad (5.6)$$

Это, в свою очередь, приводит нас к вычислению *агрегированного стационарного распределения в виде*

$$\Pi_i = \frac{\pi_{(i,s)}^{(0)} \bar{V}_{(i,s)} + \pi_{(i,c)}^{(0)} \bar{V}_{(i,c)} + \pi_{(i,e)}^{(0)} \bar{V}_{(i,e)}}{\sum_{j=0}^U \sum_{x=s,c,e} \pi_{(j,x)}^{(0)} \bar{V}_{(j,x)}}, \quad (5.7)$$

где $\bar{V}_{i,x}$ получено из (5.6).

5.5.2 Основные показатели производительности

Учитывая приведенные выше выводы, мы можем установить

$$E[N] = \sum_{i=0}^U i \Pi_i. \quad (5.8)$$

Кроме того, мы можем получить среднее число пользователей

$$E[\dot{N}] = \sum_{x=s,c,e} \sum_{i=0}^U i \pi_{i,x}^{(0)}. \quad (5.9)$$

Кроме того, мы можем получить $S(i, x)$, который является числом пакетов, отправленных во время события i^{th}

$$S(i, x) = ip(1 - p)^{i-1} B(0, L_x | U - i) + (1 - p)^i B(1, L_x | U - i). \quad (5.10)$$

Наконец, чтобы рассчитать пропускную способность системы, мы должны найти количество пакетов данных, которые могут быть отправлены за интервал генерации пакета

$$E[S] = \sum_{i=0}^U \sum_{x=s,c,e} S(i, x) \pi_{(i,x)}^0 \frac{1}{\bar{V}_{i,x}}. \quad (5.11)$$

Кроме того, мы должны охарактеризовать количество пакетов данных во вложенных точках или на каждое событие

$$E[\acute{S}] = \sum_{i=0}^U \sum_{x=s,c,e} S(i,x) \pi_{(i,x)}^0. \quad (5.12)$$

Следовательно, в это время мы можем рассчитать количество бит, отправленных на интервал генерации пакета, что переводится в пропускную способность системных данных

$$R = \frac{E[S]l}{t_g}, \quad (5.13)$$

где l – фактическая длина пакета в байтах, а t_g – длительность генерации пакета в μs .

Важно отметить, что мы оцениваем задержку пакета (μs) как временной интервал между событием генерации и моментом успешной передачи пакета. Пусть f – количество попыток передачи. Следовательно, задержка может быть получена как сумма одного успеха, $f - 1$ коллизий и $f - 1$ наборов простаивающих слотов после каждой встречной передачи

В конечном счете, задержка, выраженная в количестве событий, может быть получена с использованием закона Литтла как

$$E[\acute{T}] = 1 + \frac{E[\acute{N}]}{E[\acute{S}]} = (f - 1) \left(\frac{1}{p} - 1 \right) + (f - 1) + 1, \quad (5.14)$$

где $(1/p - 1)$ является геометрическим средним простаивающих слотов после каждого столкновения. Следовательно, $f - 1 = p \frac{E[\acute{N}]}{E[\acute{S}]}$. Это, в свою очередь, приводит к фактическому вычислению задержки данных (в μs)

$$E[T] = t_s + p \frac{E[\acute{N}]}{E[\acute{S}]} (t_c + \left(\frac{1}{p} - 1 \right) t_e), \quad (5.15)$$

где $E[\acute{S}]$ получается через (5.12) и $E[\acute{N}]$ задается (5.9).

Наконец, для эффективной оценки энергопотребления мы используем следующее уравнение

$$E = \int P^{(t)} dt = P_s t_s + (f - 1) P_c t_c + (f - 1) \left(\frac{1}{p} - 1 \right) t_e P_b, \quad (5.16)$$

где P_s , P_c и P_e – средние уровни энергопотребления для успешных, столкновений и пустых слотов в mW соответственно. Результаты приведены на рисунке 5.6, показывающем уровни потребления энергии, изменяя высоту колонок, соответствующих состояниям радиоизлучения.

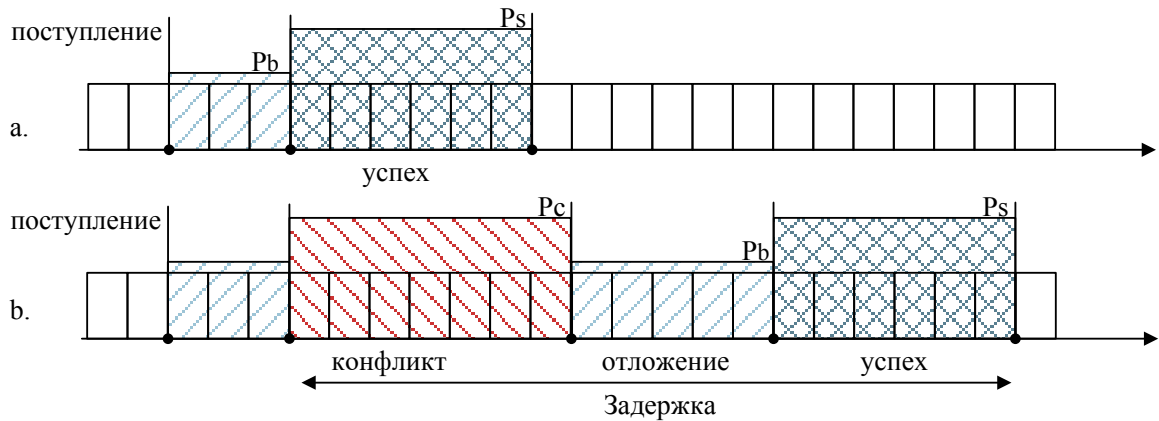
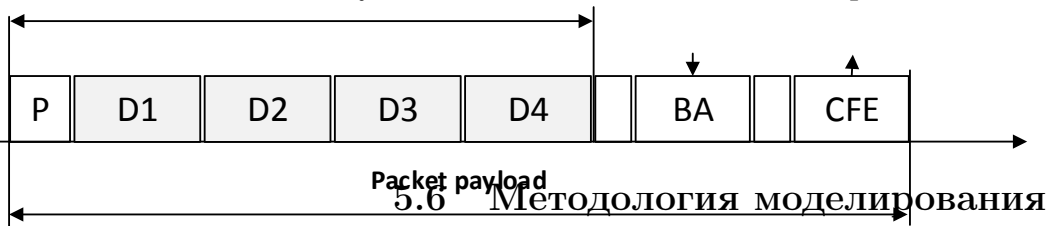


Рисунок 5.6 – Метод оценки потребления энергии



Начальные параметры, связанные с нашим имитационным моделированием, взяты из документа системных требований в [261]. В частности, настройки уровня MAC и PHY суммируются в таблице 5.2 и заимствованы из [254], тогда как большинство связанных с PHY функций были абстрагированы от их включения в соответствующие временные значения для снижения сложности оценки производительности.

Основные параметры для имитационного моделирования

Параметр	Значение
Размер пакета l	256 bytes
Скорость РНУ	0.65 Mbps
Количество пользователей U	5 to 6000
Начальный размер окна backoff(W_0)	16
Максим. размер окна (CW_{max})	1024
Количество переспросов (RC)	4
Время моделирования	200 sec
Длительность DIFS	264 μs
Длительность SIFS	160 μs
Таймаут на РНУ	240 μs
Длительность АСК	240 μs
Длительность слота усп. передачи t_s	4264 μs
Длительность слота коллизии t_c	4316 μs
Длительность слота ожидания t_e	52 μs
Мин. время генерации пакета t_g	52 μs

5.6.1 Описание и проверка моделирования

Используемая здесь среда моделирования на уровне системы представляет собой инструмент с открытым исходным кодом с именем OMNeT++ [262]. Эта среда имитационного моделирования общего назначения основана на языке программирования C++ предоставляет богатыми возможностями моделирования протокола. В частности, мы использовали ее для расширения существующей базовой линии IEEE 802.11 для поддержки необходимых функций IEEE 802.11ah. Оценки были выполнены с использованием наших двух вычислительных кластеров – оба имеют 96 ГБ ОЗУ, 10 ГБ Swap, два Intel (R) Xeon (R) CPU E5-2630 @ 2.30 ГГц с шестью ядрами каждый.

Как результат, различные методы доступа к MAC-уровню, включая EDCA, DCF, PCF и RAW, с их соответствующими параметрами могут быть легко смоделированы и оценены. Могут быть дополнительно протестированы различные модели потерь каналов и потерь для внутренних и наружных сценариев, как описано в спецификации IEEE 802.11ah. Кроме того, можно гибко оценить множество схем модуляции и кодирования (MCS). Как правило, платформа OMNeT++ использовалась в прошлом для реализации многочисленных беспроводных протоколов [263].

Широкая проверка обсуждаемого инструмента в контексте технологии IEEE 802.11ah уже была выполнена в [264]. Там проверка была сосредоточена на оценке пропускной способности системы и потребляемой мощности для основного механизма доступа (например, DCF) с RAW и без него, где предполагался насыщенный трафик. Результаты моделирования были сопоставлены с хорошо установленными аналитическими выражениями для насыщенной пропускной способности от [135], и авторы нашли идеальное соответствие между обоими типами результатов.

5.6.2 Текущие числовые результаты

Как следует из предыдущих разделов, наш подход представляет собой мощную структуру, которая в равной степени применима для насыщенных и ненасыщенных сценариев с большим количеством пользователей для режима BASIC asses. Далее, мы сначала оцениваем общее количество успешно отправленных блоков данных (передач) на каждый слот, который затем преобразуется в фактическую пропускную способность системы (согласно (5.13)).

Обратим внимание на рисунок 5.7 — в настоящее время, при помощи моделирования оказывается трудным получение результатов для очень большого числа пользователей (из-за исключительно высокой сложности моделирования), но, в то же время, полученные данные демонстрируют отличную конвергенцию с нашим анализом для плотности нижних узлов.

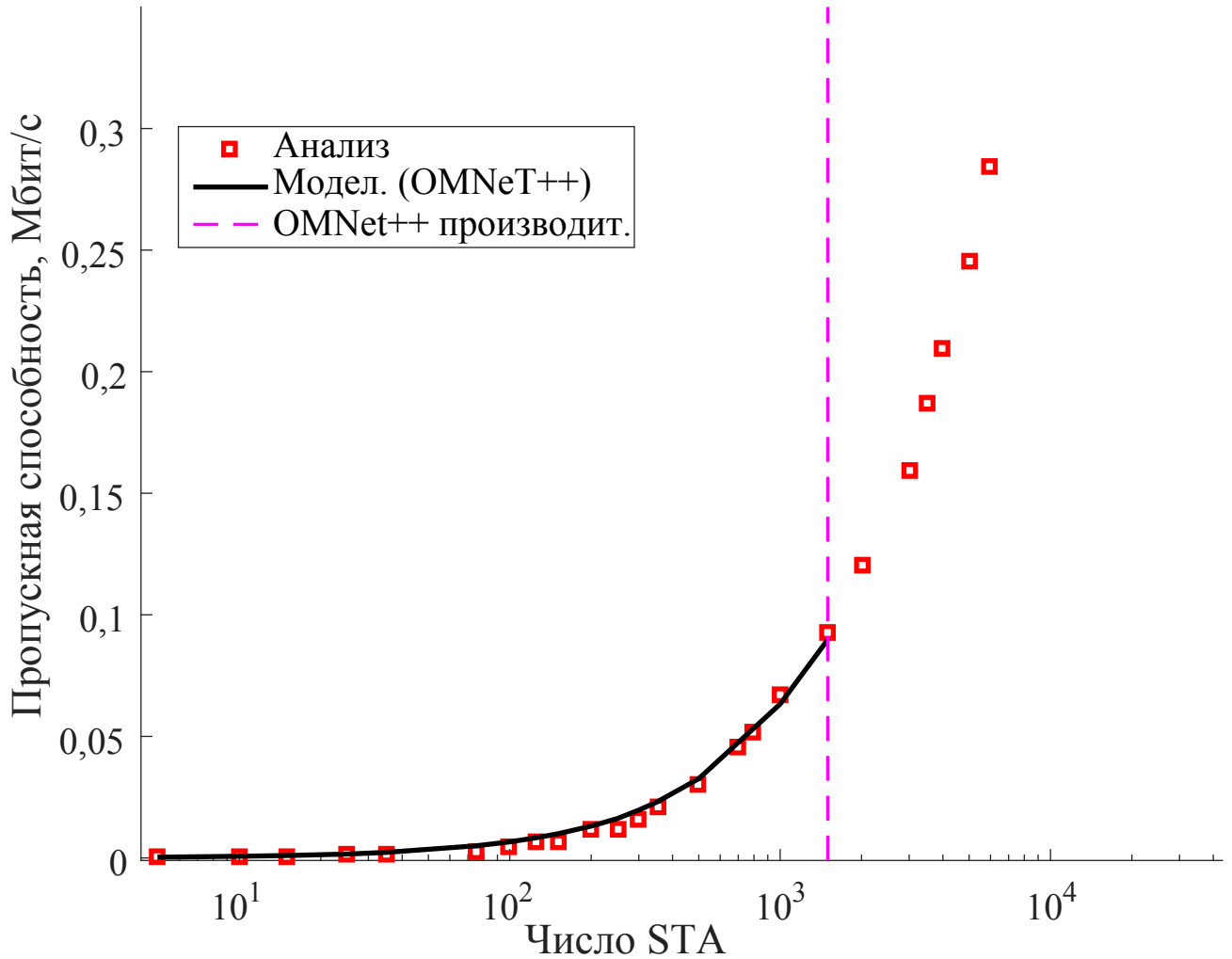


Рисунок 5.7 — Пропускная способность системы при среднем времени генерации пакета 10 сек.)

Отсюда, полученная аналитическая модель используется здесь для прогнозирования производительности до 6000 устройств МТС. Здесь вероятность генерации пакета статична — пакет данных может генерироваться один раз в 1 с. Естественно, что точный период генерации можно вычислить как $1/\sigma \cdot t_g$, а соответствующие выходные результаты как (5.13).

Кроме того, принимая во внимание предлагаемый механизм RAW в рамках рассматриваемой методики оценки, мы можем выделить некоторые из самых новых особенностей технологии 802.11ah. Мы сравниваем нашу модель с результатами моделирования, полученными для различного количества RAW, см. рисунок 5.8. Очевидно, что предлагаемая аналитическая модель и используемая низкая скорость поступления пакетов демонстрируют сходное поведение пропускной способности системы для небольшого

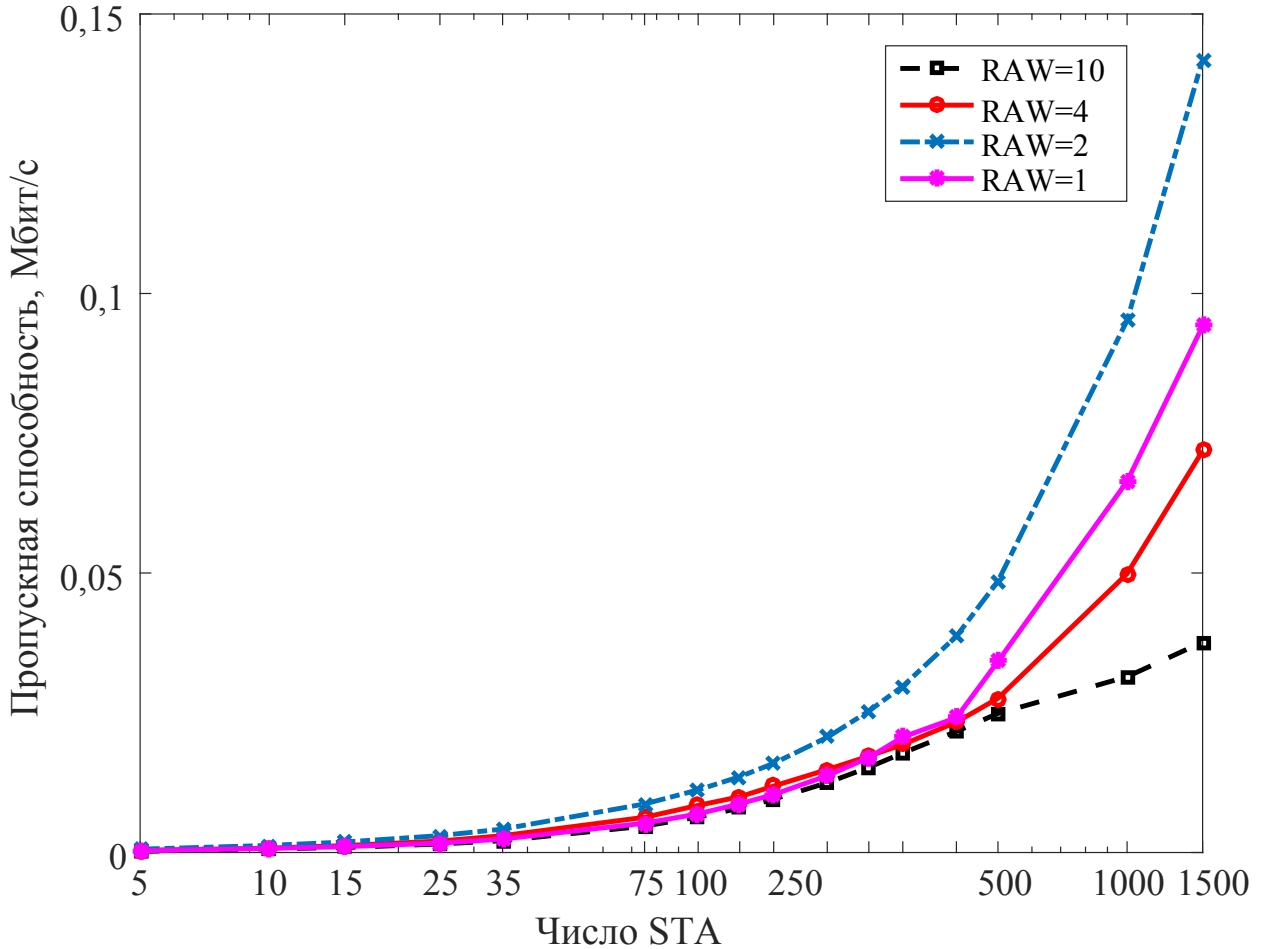


Рисунок 5.8 — Пропускная способность для разного количества окон RAW

количества периодов RAW [265]. Однако оно будет меняться по мере увеличения количества окон. Этот эффект обусловлен тем, что устройство, скорее всего, находится в спящем режиме, когда пакет генерируется и отправляется в буфер. Важно отметить, что для большого количества RAW и пользователей первая передача происходит более редко, и, следовательно, система быстрее сходится к насыщению. Поэтому предлагаемая аналитическая модель может иметь некоторые ограничения из-за возможной несправедливости системы, поскольку она рассчитана на достаточно низкие скорости поступления заявок.

Учитывая, что точная работа протокола RAW по большей части зависит от вендора, мы используем независимую от реализации метрику оценки в виде условной вероятности коллизий. Рассматриваемый сценарий, ориентированный на RAW, следует принципам в [261] с (5.7) и приводит к рисунку 5.9, что подтверждает предположения о том, что группировка RAW

может уменьшить вероятность коллизий в системе. Здесь мы предполагаем, что устройство не генерирует новые данные (находится в спящем режиме) во время интервалов передачи RAW, отличных от собственных. Поэтому мы увеличиваем скорость потока прихода и анализируем альтернативные настройки RAW для данного общего количества устройств МТС. На рисунке показано, как конкретная конфигурация влияет на вероятность коллизии и помогает выбирать соответствующие параметры RAW, поскольку она четко характеризует тенденции роста коллизий данных и, таким образом, насыщения. Данный график также подчеркивает преимущества предложенной аналитической модели, поскольку он может моделировать системы с разными временами взаимного прихода.

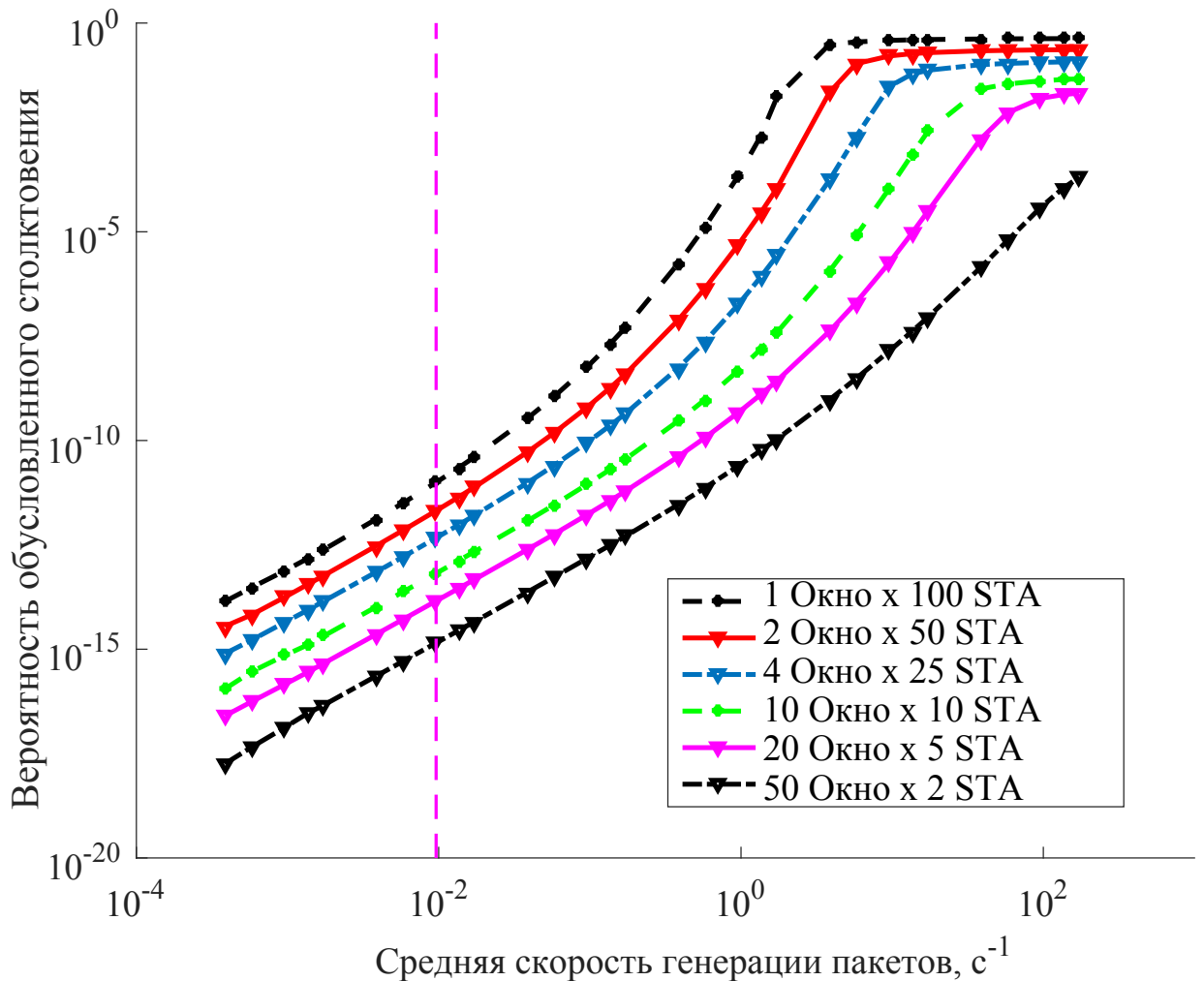


Рисунок 5.9 — Зависимость вероятности конфликта от входного потока

Кроме того, мы оцениваем эффективное потребление энергии для успешно передаваемого пакета с (5.16) для случая, когда RAW не исполь-

зуются. Он рассчитывается как общее потребление энергии с конца события генерации пакета и до тех пор, пока данные не будут успешно переданы. Значения, связанные с мощностью, берутся из [264] — 255 мВт для передачи, 135 мВт для приема и прослушивания канала и 1,5 мВт для режима ожидания. Результаты вывода суммируются в таблице 5.3. Важно отметить, что фактическая интерсивность коллизии напрямую влияет на относительную разницу между имитационной и аналитической моделями. С большим количеством устройств М2М в системе результаты моделирования демонстрируют более высокую нестабильность, тогда как наш анализ предполагает усредненную работу стационарной системы.

Таблица 5.3

Энергоэффективность системы

Количество узлов	Моделирования	Аналитика	Разница
100	1.01	1.008	0.24%
500	1.023	1.017	0.56%
1000	1.055	1.039	1.59%
1500	1.203	1.104	8.92%

Далее на рисунке 5.10 мы увеличиваем скорость прибытия трафика и анализируем условия близкого к насыщению для выбранных сценариев МТС. Хотя аналитическая модель, предложенная в этой работе, не предназначена для решения проблемы с полным буфером, ясно видно, что для периодов генерации пакетов более 10 секунд при обычных настройках насыщение не происходит. Например, обратим внимание на синюю кривую ($t_g = 1s$), из виднеющихся двух экстремумов на графике. Первый из них составляет около 150 узлов, и наиболее виден, при исполнении процедуры ВЕВ (двоичная экспоненциальная отсрочка), то есть когда устройства начинают сталкиваться и, таким образом, увеличивают время повторной передачи. Второй – около 400 узлов и представляет наихудший случай, когда все устройства находятся в глубокой фазе повторной передачи в течение большей части времени и количество коллизий увеличивается. Область после 500 узлов отражает использование неэффективных параметров ВЕВ для определенного количества STA, то есть когда скорость разрешения столкновений быстрее, чем появление новых коллизий.

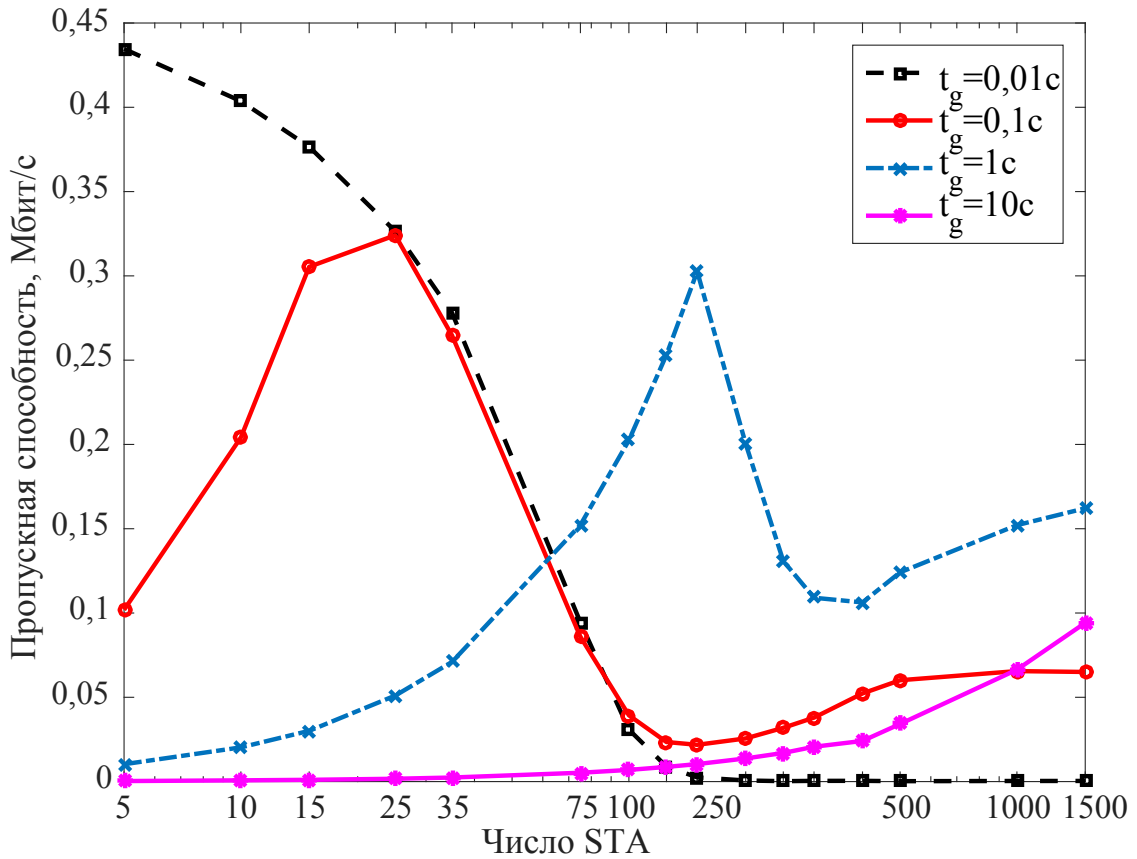


Рисунок 5.10 — Пропускная способность для случаев близких к насыщению

Далее на рисунке 5.11 приведена оценка целесообразности времени моделирования на платформе OMNet ++ для используемого вычислительного кластера. Для 1500 M2M устройств может потребоваться до 14 часов вычислительного времени всего на один этап моделирования, можно ожидать, что время будет экспоненциально расти по мере добавления большего количества пользователей в систему.

В данном подразделе мы обобщили потенциал предлагаемых новых спецификаций IEEE 802.11ah для поддержки массового развертывания МТС. Также, мы предложили новую аналитическую модель, которая характеризует пропускную способность системы, задержку и энергопотребление для большого количества M2M устройств, обменивающихся данными по технологии IEEE 802.11ah в ненасыщенном режиме. Наша математическая модель была проверена инструментом имитационного моделирования OMNeT++ и, как это было показано, предложила адекватные прогнозы

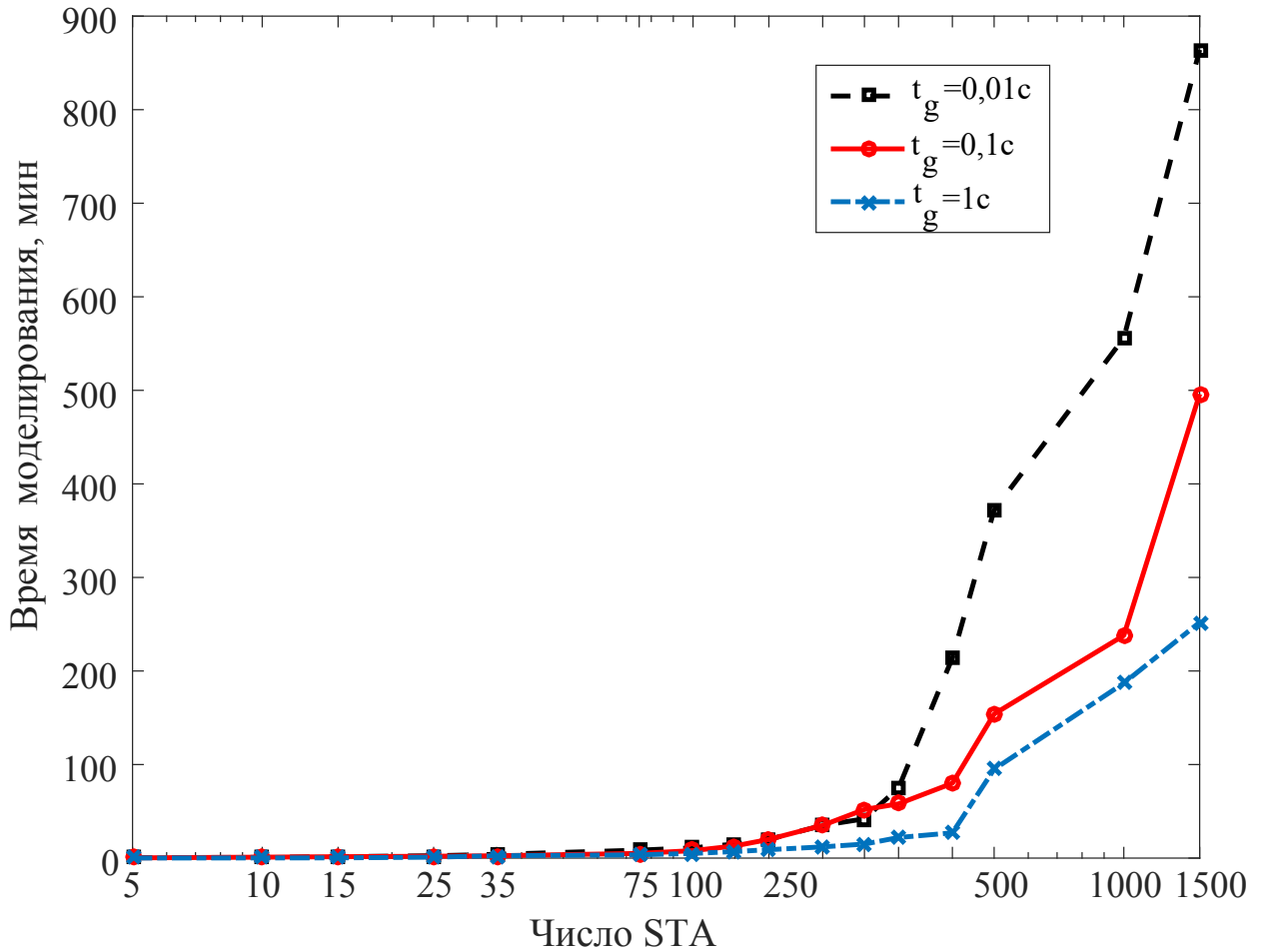


Рисунок 5.11 — Среднее время моделирования

производительности даже в сценариях, где имитационное моделирование имеет ограниченные возможности из-за увеличения числовой сложности.

Кроме того, мы проанализировали схему RAW (ограниченного доступа) IEEE 802.11ah, которая позволяет снизить влияние коллизий и тем самым повысить производительность системы. В частности, мы рассмотрели возможные группировки из 100 устройств и соответствующие значения вероятности столкновения. Также было получено значение эффективного энергопотребления по переданному пакету. Важно отметить, что относительная разница, наблюдаемая между имитационным моделированием (OmNet++) и анализом (цепи Маркова), находится в пределах 10% от эффективного потребления энергии.

5.7 Предложение по эффективной процедуре доступа для малых данных для машинно-ориентированных коммуникаций LTE

Принимая во внимание тот факт, что создаваемый машинно-ориентированными коммуникациями трафик может быть изредка очень большим, в этом подразделе основное внимание уделим эффективному доступу для малых данных к 3GPP LTE системам. Подчеркнем также, что устройства МТС должны потреблять очень мало энергии в течение длительного времени и, поэтому, вопросы энергоэффективности также будут учтены при исследовании. Заметим, что эти важнейшие исследовательские проблемы пока не нашли удовлетворительного освещения в существующей научной литературе, где, в основном, внимание сфокусировано на вопросах перегрузки [266–268].

Далее продолжим исследования с учетом обеспечения доступа для малых данных, когда на сети не наблюдается перегрузки от МТС. Предложим и детально исследуем эффективный механизм передачи малых данных, который может быть использован как альтернатива традиционной сигнализации, существенно увеличивая при этом производительность МТС. Более конкретно, в подразделе будут рассмотрены предложенные в работах автора: (i)- новая интегральная аналитико-моделирующая структура для оценки механизмов доступа данных для МТС; (ii) эффективная схема с учетом специфики данных, названная ранее в работах автора **contention-based LTE transmission (COBALT)**.

5.7.1 Введение в проблематику доступа данных к каналам в LTE

3GPP LTE является относительно новой беспроводной технологией, которая в тоже время достаточно зрелая для обеспечения повсеместной

связности для сотовых сетей. В настоящее время LTE система состоит из малых физических элементов и, в зависимости от конфигурации, комбинация из 72 или 84 из них образует простой ресурсный блок (resource block, RB). Для восходящей линии один RB включает 12 поднесущих в частотном спектре и 6 или 7 SC-FDMA символов во временной области.

В диссертационной работе ограничимся исследованием наиболее используемой на практике конфигурации в полосе 5 МГц с 25 ресурсными блоками RB (рисунок 5.12). Во временной области один блок RB занимает 0.5 мс, при этом пара RB (2 смежных блока RB) формируют подкадр длительностью в 1 мс и являются наименьшей единицей планировщика. Десять подкадров образуют радио кадр длительностью 10 мс. На рисунке 5.12 ресурсы кадров распределяются между тремя каналами как это представлено ниже.

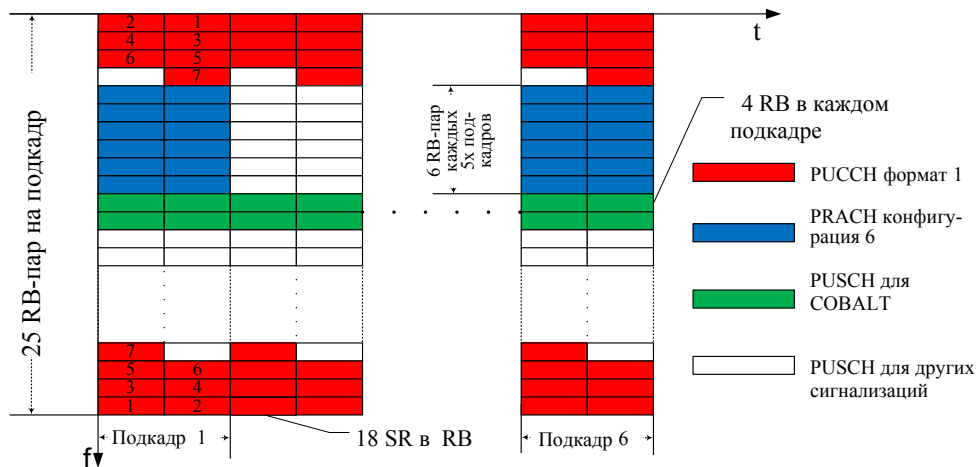


Рисунок 5.12 — Распределение ресурсных блоков для доступа данных к каналам

Физический управляющий канал для восходящей линии (Physical Uplink Control Channel, PUCCH) выделяется для переноса управляющей информации по восходящей линии, включая: (i) информацию о качестве канала для схемы адаптивного управления модуляцией и кодированием, а также мощностью, (ii) запросы планировщика (SRs), требующие системных ресурсов, (iii) индикаторы для управления множественными антеннами MIMO, (iv) ответы на гибридный автоматический повтор запросов (Hybrid Automatic Repeat requests, HARQ). Абстрагируясь от разнообра-

ных PUSCH конфигураций, рассмотрим далее только тип 1, используемый для передачи SR (рисунок 5.13, слева).

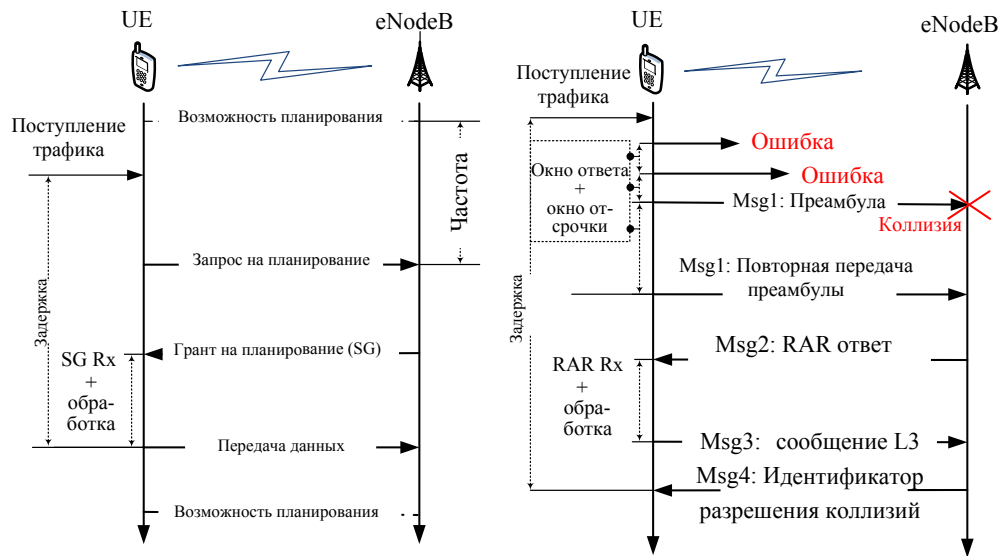


Рисунок 5.13 — Примеры сигнализации для каналов PUSCH (слева) и PRACH (справа)

Случайный доступ к каналу на физическом уровне PRACH используется устройствами при передаче SR вместо PUSCH в случае инициализации вхождения в сеть, точно так же, как PUSCH не используется устройствами при распределении ресурсов. Как правило, устройство выбирает псевдо случайную преамбулу и передает ее в режиме конкуренции для последующего доступа в сеть со своими данными для восходящей линии (рисунок 5.13, справа).

Физический совместный канал восходящей линии (Physical Uplink Shared CHannel, PUSCH) занимается актуальными данными для передачи и частично управляющими сообщениями. Как правило, PUSCH инкорпорирует оставшиеся неиспользуемыми PUSCH или PRACH ресурсные блоки. В зависимости от условий канала, различные схемы модуляции и кодирования могут быть использованы для передачи данных. Вследствие этого, число бит, переносимых простой парой PUSCH RB варьируется от 16 до 712 [269].

5.7.2 Традиционный доступ для данных в LTE

Механизм распределения ресурсов по умолчанию для PUSCH для восходящей линии UL при передаче данных базируется на приоритетной передаче SR. В случае, когда устройство уже имеет временные ресурсы восходящей линии и выделенные ресурсы PUSCH, оно может использовать эти ресурсы для передачи своего SR (рисунок 5.13, слева), в противном случае будет использоваться PRACH (рисунок 5.13, справа). Периодически ресурсный блок PUSCH RB доступен для передачи SR, в зависимости от индекса конфигурации SR может изменяться от 1 до 80 мс (подкадры). Более того, несколько SR от различных устройств могут быть агрегированы в простой ресурсный блок PUSCH RB и в диссертационной работе предполагаем, что до 18 запросов могут быть мультиплексированы [270].

После передачи своего SR устройство нуждается в ожидании до тех пор, пока базовая станция eNodeB не ответит о том, что есть соответствующее разрешение на планирование передачи (рисунок 5.13, слева и рисунок 5.14). Основное преимущество передачи SR через PUSCH состоит в высокой надежности и близкому к детерминированному распределению длительности задержки. Однако, когда число устройств велико, ресурсы PUSCH могут быть быстро исчерпаны.

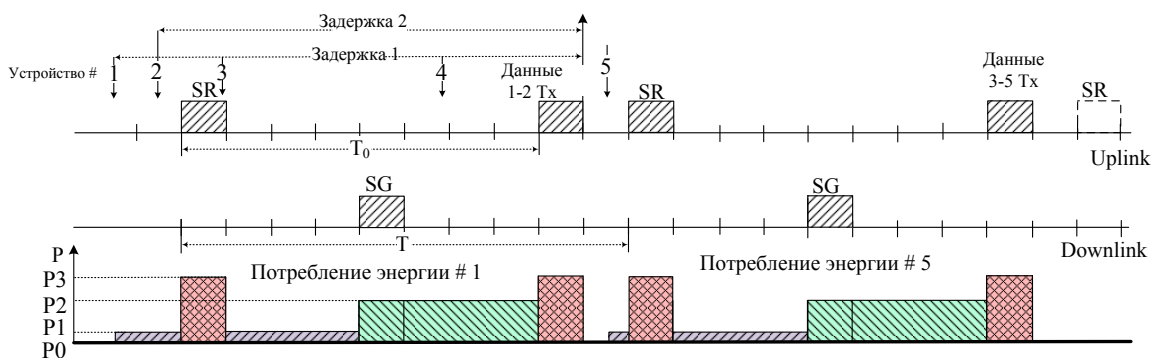


Рисунок 5.14 — Пример временной диаграммы канала PUSCH

Когда использование ресурсов PUSCH становится неэффективным для поддержки каждого активного устройства в конкретной соте сети, другая пара ресурсных блоков RB может быть использована или перио-

дичность передачи SR может быть увеличена. В любом случае, передача SR через PUSCH ожидаемо потребует много системных ресурсов при увеличении популяции устройств. В предельном случае ресурсы PUSCH могут быть истощены даже при большой периодичности и большом числе заявок на мультиплексированные ресурсные блоки RB. Как отмечалось выше, альтернативным методом для передачи SR является случайный доступ с процедурой PRACH.

В этом случае устройство начинает процедуру посредством передачи одной из 54 преамбул (Msg1). Далее, если преамбула передана успешно, базовая станция eNodeB отвечает сообщением RAR (Msg2), а ресурсы для передачи индицируются в сообщении устройства Msg3. Наконец, устройство ожидает ответа на свое сообщение Msg3 в виде Msg4 [271].

В отличие PUSCH передача PRACH не так надежна и может быть неуспешной не только когда не проходит сообщение Msg1 (вследствие коллизий или недостаточной мощности передачи), но также и в связи с проблемами в передаче или в приеме последующих сообщений, требуемых для завершения процедуры. Все отказы во время процедуры случайного доступа приводят к повторной попытке в случайное время. После рестарта время повторной попытки выбирается исходя из равномерного распределения с размером окна повторной попытки в 20мс [271].

Несмотря на ограниченную надежность, основное преимущество процедуры PRACH заключается в том, что требуется фиксированное число ресурсных блоков RB. Например, один PRACH требует занятия точно 6 пар ресурсных блоков RB на подкадр (рисунок 5.12 и [271]) и устройства могут осуществлять попытки передачи своих преамбул только в подкадрах 1 и 6. Однако с увеличением числа устройств или частоты передачи ими трафика данных вероятность коллизий может стать большой, так же, как и задержка на доступе и потребление энергии вследствие повторных передач. Кроме того, экстенсивное использование PRACH для доступа данных может блокировать инициализацию доступа в сеть для других МТС или пользователей H2H. Исходя из сказанного, в диссертационной работе ниже будет предложена альтернативная схема доступа для данных от МТС устройств.

5.7.3 Предлагаемый конкурентный доступ для МТС

Основная идея предлагаемого конкурентного доступа COBALT для систем LTE состоит в передаче пакетов малых данных непосредственно через PUSCH вместо использования ресурсов времени и мощности сигнализации PUCCH/PRACH. Идея сама по себе в оригинальном виде была предложена в [272] в контексте уменьшения задержек в LTE. В диссертационной же работе предлагается использовать механизм COBALT для сценария МТС, когда число устройств очень велико, а трафик нерегулярен и мал. При этом ожидается, что в столь экстремальных условиях предложенная схема будет допускать малые задержки и, что даже более важно, уменьшит потребление энергии на сигнализацию для небольших МТС устройств, питание которых осуществляется от батареи.

Для справедливого сравнения было решено повторно использовать расписание, относящиеся к процедуре PRACH, при необходимости, в порядке сравнения с предложенной схемой в подобных условиях. В общем и целом процедура COBALT представлена на рисунок 5.15. Изначально базовая станция eNodeB рассылает широковещательное сообщение всем ассоциированным устройствам, отражая в нем где и как много ресурсных блоков RB доступны для предложенной конкурентной передачи. В целом, такие сообщения могут быть периодическими или направляться по запросу в зависимости от динамического состояния системы, но для упрощения исследования в направлении статического сценария будем полагать, что число ресурсных блоков, доступных для COBALT в каждом подкадре, постоянно.

Когда ресурсные блоки распределены для процедуры COBALT, два или более устройств могут послать свои данные на один и то же ресурсный блок. Вследствие этого происходят коллизии и подтверждение от базовой станции не приходит. В этом случае устройства, для сообщений которых произошла коллизия, инициируют процедуру backoff и снова передают свои данные. Если подтверждение получено успешно, после некоторого промежутка времени процедура COBALT завершается (рисунок 5.16).

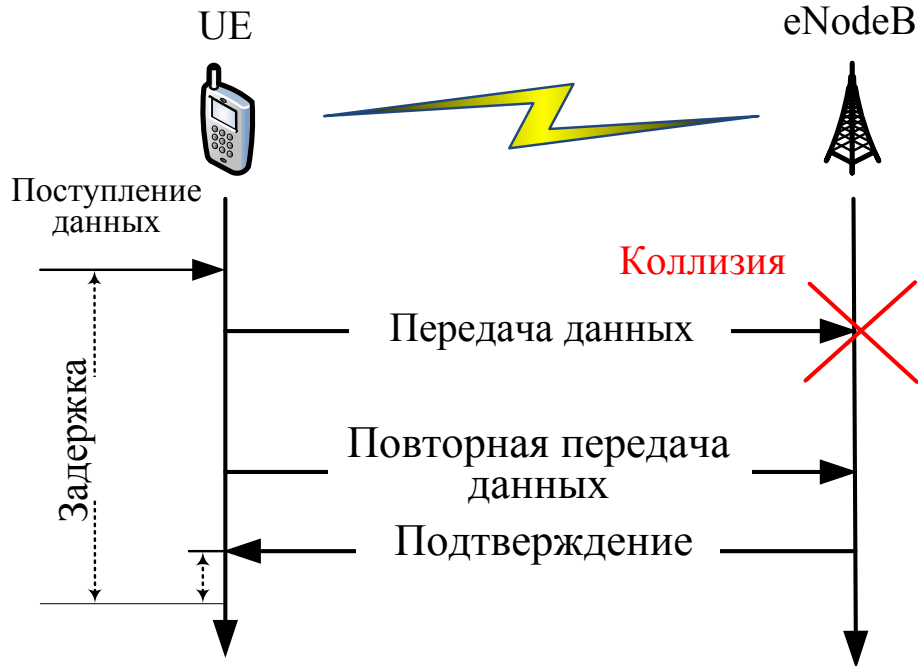


Рисунок 5.15 — Предложенная сигнализация для LTE (COBALT), основанная на конкурентном доступе

Подчеркнем снова, что предложенный конкурентный доступ для данных использует только ресурсы PUSCH. Вследствие этого основными параметрами для процедуры COBALT являются: (i) число доступных PUSCH ресурсных блоков RB в подкадре (предполагается, что это, как минимум, 4 ресурсных блока или две пары блоков), (ii) периодичность такой доступности (предполагаем, что наименьший период составляет 1 подкадр). Размер окна backoff, так же, как и ассоциированные таймеры, принимается подобным процедуре PRACH.

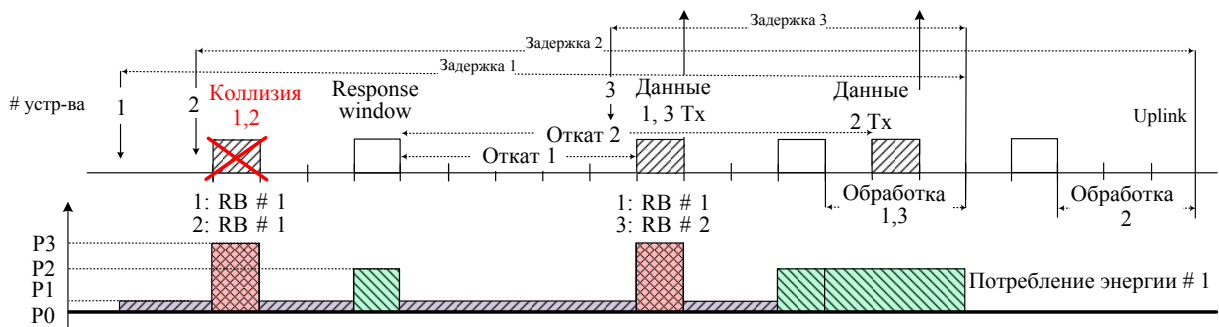


Рисунок 5.16 — Пример временной диаграммы COBALT

5.7.4 Системная модель и предположения

В отношении сравнения эффективности механизмов доступа для данных, рассмотрим M одинаковых устройств МТС, расположенных в сети 3GPP LTE системы с дуплексным частотным разделением каналов FDD (Frequency Division Duplex) [271]. Сфокусируем внимание на 1000 (1К), 5000 (5К) и 10000 (10К) устройств. Все устройства входят в систему на протяжении всей длительности моделирования и интересующие метрики собираются также в течение всего времени.

Будем использовать далее стандартную модель трафика МТС UL в соответствии с техническим 3GPP [273], что методологически определяет в [271] сетевую перегрузку, близкую к реальным моделям трафика устройств. Документ [273] предписывает, чтобы распределение длительности интервалов времени между поступлением пакетов было экспоненциальным со средним значением 30 с. Пакеты данных размером 256 бит передаются с фиксированным уровнем модуляции и кодирования 16-QAM с использованием точно одной пары ресурсных блоков [269] (1 подкадр во временном интервале) в диапазоне выше 5 МГц.

Поступающий поток данных рассматривается как пуассоновский поток без последствия, представляющий число поступлений пакетов в произвольный момент времени t . С учетом свойств этого процесса вероятность p_0 того, что число поступивших пакетов за время слота длины t_0 равна 0, определяется следующим образом:

$$p_0 = \Pr\{X(t, t + t_0) = 0\} = e^{-\lambda t_0}, \quad (5.17)$$

где $X(t, t + t_0)$ – число поступивших пакетов за интервал времени $[t, t + t_0)$ и λ интенсивность поступления пакетов.

Реализация процедур PUSCH и PRACH осуществляется в строгом соответствии с документами 3GPP. В частности, PUSCH базируется на значениях таймеров из таблицы В.1.2.1.1-1 из [274], а таймеры PRACH задаются в соответствии с таблицей В.1.1.1-1 из [274] и на основе [271].

Некоторые идеи процедуры COBALT были обсуждены в [272]. Важные параметры для оценки сведены в таблицу 5.4.

Таблица 5.4

Основные параметры системы

Обозначение	Параметр	Значение
λ	Интенсивность поступления для МТС устройств	$1/30 \text{ s}^{-1}$
t_0	Размер подкадра	1мс
M	Число устройств МТС	1К, 5К, 10К
s	Общее число преамбул /COBALT RB	54/4
L_1	Максимальный номер преамбулы-попытки передачи данных	10/4
W	PRACH/COBALT индикатор backoff	20мс
$1-p_{tx}$	HARQ вероятность повторной передачи для Msg3 и Msg4 (не адаптивный HARQ)	10%
L_3	Максимальное число HARQ Tx для Msg3 и Msg4 (не адаптивный HARQ)	5
b	Периодичность PRACH возможностей	5мс
K	PRACH/COBALT окно ответа	5/1мс
K_0	Период конкуренции Tx и пауза (PRACH/COBALT)	3мс
t_{pr}	Время обработки перед Msg3 Tx/COBALT Подтверждение, декодирование и обработка	5/3мс
t_{tx}	Длительность Msg3 Tx, ожидание и Msg4 Rx	6мс
T_0	Длительность процедуры PUSCH	8мс
T	PUSCH SR периодичность	10мс

Для энергетической модели МТС устройств будем использовать четыре различных состояния (рисунки 5.14–5.16) с потреблением мощности

возможных будущих МТС устройств из [275]. Будем дифференцировать эти состояния следующим образом: (i) *состояние* незанятости ($P_0 = 0.01$ mW), когда буфер трафика пустой и данные не передаются, (ii) *состояние ожидания* ($P_1 = 10$ mW), когда трафик поступает, но устройство не передает его, находясь в ожидании возможности передачи в конкурентной среде, (iii) *состояние приема* (Rx) ($P_2 = 100$ mW), когда устройство запросило базовую станцию о передаче и обрабатывает сообщения подтверждения, (iv) *состояние передачи* (Tx) ($P_3 = 300$ mW), когда устройство передает данные.

Важно отметить, что реальное потребление энергии в каждом из состояний может зависеть от специфики внедрения и приведенная модель, поэтому, рассматривается только как всеобъемлющий пример. Для того, чтобы дать возможность потенциального улучшения при реальном внедрении устройств, будем рассматривать далее оптимальное потребление энергии в том смысле, что устройство может немедленно вернуться в режим ожидания, т.е. оно не должно мониторить активность нисходящей линии связи в каждом подкадре. Другими словами, принимается идеальное предположение о том, что устройство знает, когда обратная связь поступит от базовой станции eNodeB и может соответствующим образом регулировать уровень мощности.

Ниже продолжим исследование важнейших метрик эффективности аналитически и с использованием имитационного моделирования. Сфокусируем внимание на потреблении энергии ресурсными блоками RB (включая число ресурсных блоков RB, доступных в системе для механизма доступа для данных и число реально используемых ресурсных блоков), потреблении мощности устройствами МТС и значениях задержки для доступа данных (определяемых как длительность интервала времени между поступлением данных и их успешной UL передачей).

5.8 Анализ доступа для малых данных в LTE

5.8.1 Основные положения

Последующий анализ трех схем доступа для данных показывает три различных подхода к этой проблеме. В то время как оценка механизма на базе PUSCH весьма тривиальна, анализ схем, основанных на конкурентном подходе (PRACH и COBALT), является существенно более новым. Из-за присущей конкурентному процессу памяти непосредственно решения не удалось получить даже для гораздо более простых протоколов ALOHA. Поэтому, попробуем адаптировать эквивалентные модели для потоков без последствия для обеих предложенных схем.

Благодаря большому числу преамбул, доступных для процедуры PRACH и принимая это во внимание, все переходы между состояниями системы можно считать излишне сложными. Тогда приступить к исследованию PRACH-основанного доступа для данных можно с точки зрения конкретного числа конкурирующих устройств МТС и их конкурентного поведения, абстрагируясь от большинства переходов через их усреднение. Предложенная аппроксимация подходит, в основном, для низко нагруженных систем, когда вероятность коллизий достаточно мала. В противном случае, вероятность коллизий при COBALT-основанном доступе высока вследствие более эффективного использования ресурсов. Поэтому, для получения лучших решений будем рассматривать все возможные переходы между состояниями системы и анализировать стационарные распределения.

5.8.2 Анализ доступа на базе PUSCH

Передача данных на базе PUSCH не чувствительна к коллизиям и не включает в себя периоды повторных передач. Длительность передачи пакета данных предполагается в течение 1 подкадра. Таким образом, средняя задержка пакета вычисляется следующим образом:

$$E[\tau] = T/2 + T_0 + 1, \quad (5.18)$$

где T_0 – длительность процедуры PUSCH (рисунок 5.14), T – периодичность передачи SR и 1 – позиции для времени передачи.

Для оценки потребляемой мощности будем рассматривать интервалы времени, в течение которых устройство находится в каждом из состояний:

$$q_1 = (T/2 + 3) \cdot \lambda, q_2 = (T_0 - 3) \cdot \lambda, q_3 = 2 \cdot \lambda. \quad (5.19)$$

Общая энергия, затраченная устройством, может быть тогда определена из :

$$\epsilon = P_2 q_2 + P_3 q_3 + P_1 q_1 + P_0 (1 - q_3 - q_2 - q_1). \quad (5.20)$$

5.8.3 Анализ доступа на базе PRACH

Рассмотрим систему PRACH, используя подход, предложенный в работе автора [276], который представляет конкурентно-основанную передачу с преамбул и данных от M МТС активных устройств в соответствии с процессом поступления, описанным выше.

Общая задержка в таких системах может быть декомпозирована на две независимые составляющие:

$$E[\tau] = E[\tau^{(1)}] + E[\tau^{(2)}]. \quad (5.21)$$

Здесь $E[\tau^{(1)}]$ – аппроксимация среднего времени между обработкой Msg3, которая может быть представлена как:

$$E[\tau^{(1)}] = \sum_{j=1}^M \frac{\binom{M-1}{j-1} \rho^{j-1} (1-\rho)^{M-j}}{(1-e^{-\lambda t_0}) + (a_j(K+K_0+\bar{w}) + \frac{b-K}{2} - \bar{w})^{-1}}, \quad (5.22)$$

где $\bar{w} = c_2(c_2 + 1) + (c_2 + b + bc_3)(W - bc_3 - c_2) + bc_3c_2$, $c_1 \cong 1.42$, $c_2 = b \lceil (K+K_0)/b \rceil - K - K_0$, $c_3 = \lfloor (W - c_2)/b \rfloor$, а другие параметры приведены ниже.

Переменные p_j^* , a_j и аппроксимация нагрузки устройства ρ могут быть вычислены следующим образом:

$$p_j^* = (1 - s^{-1}(K_0 + K + \bar{w})^{-1})^{j-1} (1 - (1 - e^{-\lambda t_0})s^{-1})^{M-j}, \quad (5.23)$$

$$a_j = p_j^* \sum_{n=1}^{L_1} n \left(1 - \frac{1}{e^n}\right) \prod_{i=1}^{n-1} \left(1 - p_j^* \left(1 - \frac{1}{e^i}\right)\right), \quad (5.24)$$

$$\rho = (1 - e^{-\lambda t_0})(c_1(K + K_0 + \bar{w}) + \frac{b-K}{2} - \bar{w}), \quad (5.25)$$

где M , L_1 и s – параметры из таблицы 5.4.

Средняя длительность Msg3 и Msg4 Tx $E[\tau^{(2)}]$ определяется следующим образом:

$$E[\tau^{(2)}] = t_{pr} + \frac{t_{tx}}{p_{tx}} \left[1 - (1 - p_{tx})^{L_3} (1 + L_3 p_{tx})\right]. \quad (5.26)$$

Что касается потребляемой мощности устройством в PRACH, подход к определению этих показателей для PUSCH может быть модифицирован без каких-либо существенных сложностей.

5.8.4 Анализ COBALT-основанного доступа для данных

Рассмотрим фазу передачи, когда устройство определяет разрешение на конкурентно-основанную передачу, выполняет обработку L2 и L1 пере-

даваемых данных и передает их на PUSCH. Предположим, что разрешение для процедуры COBALT может быть получено в каждом временном интервале. Заметим, что COBALT не поддерживает процедуру управления мощностью.

Предположим также, что коллизия происходит, когда два или более МТС устройств выбирают одно и то же разрешение COBALT и передают свои данные через PUSCH. Все подвергнутые коллизии передачи рассматриваются как отказы. Если передача неуспешна, устройство МТС равномерно выбирает счетчик backoff из множества $[0, W-1]$. Если устройство успешно завершает передачу, оно очищает свой буфер.

Во-первых, можно получить приближение средней задержки доступа для данных. Чтобы определить оценку задержки в системе с памятью, применим следующую упрощенную эквивалентную модель. Предположим, что каждое из устройств, требующих возможность передать данные, осуществляет попытки передачи в каждом подкадре с вероятностью $p = \frac{1}{K_0 + K + (W+1)/2}$, и в то же время вероятность получения хотя бы одного нового пакета равна $\sigma = 1 - p_0$, а p_0 определена в (5.17).

Пусть Let $N(t)$ есть случайный процесс, представляющий общее число устройств, требующих передачи данных, в системе в момент времени t . Для исследуемой модели $N(t)$ есть марковский процесс с пространством состояний $\{0, 1, \dots, M\}$. Элементы матрицы переходов $\Pi = \{\pi_{ij}\}_{i,j=0}^M$ для рассматриваемой марковской цепи представлены в таблице 5.5.

Ожидаемая доля успешно используемых каналов в состоянии i , $S_{out}(i)$ вычисляется следующим образом:

$$\begin{aligned}
S_{out}(i) = & (1-p)^i (M-i) \sigma (1-\sigma)^{M-i-1} + ip(1-p)^{i-1} (1-\sigma)^{M-i} + \\
& + \sum_{z=0}^i \sum_{k=\min(0, 3-z)}^{M-i} \frac{k+z}{2^{k+z-1}} \binom{i}{z} p^z (1-p)^{i-z} \binom{i}{z} \sigma^z (1-\sigma)^{i-z} + \\
& (M-i) \sigma (1-\sigma)^{M-i-1} ip(1-p)^{i-1} + (1-\sigma)^{M-i} \binom{i}{2} p^2 (1-p)^{i-2} + \\
& + \binom{M-i}{2} \sigma^2 (1-\sigma)^{M-i-2} (1-p)^i. \tag{5.27}
\end{aligned}$$

Вероятности переходов $\pi_{ij} = \Pr\{N(t) = j | N(t-1) = i\}$

j	Значение
$i-z$	0
$i-2$	$\frac{1}{2}(1-\sigma)^{M-i} \binom{i}{2} p^2 (1-p)^{i-2}$,
$i-1$	$\frac{1}{2}(M-i)\sigma(1-\sigma)^{M-i-1} ip(1-p)^{i-1} +$ $+(1-\sigma)^{M-i} [ip(1-p)^{i-1} + \sum_{k=3}^i \frac{k}{2^{k-1}} \binom{i}{k} p^k (1-p)^{i-k}]$
i	$\binom{M-i}{2} \sigma^2 (1-\sigma)^{M-i-2} (1-p)^i + (M-i)\sigma(1-\sigma)^{M-i-1} [(1-p)^i +$ $+\sum_{k=2}^i \frac{k+1}{2^k} \binom{i}{k} p^k (1-p)^{i-k}] + (1-\sigma)^{M-i} [\binom{i}{2} p^2 (1-p)^{i-2} +$ $+(1-p)^i + \frac{1}{2} \binom{i}{2} p^2 (1-p)^{i-2} + \sum_{k=3}^i (1 - \frac{k}{2^{k-1}}) \binom{i}{k} p^k (1-p)^{i-k}]$
$i+1$	$\binom{M-i}{2} \sigma^2 (1-\sigma)^{M-i-2} \sum_{k=1}^i \frac{k+2}{2^{k+1}} \binom{i}{k} p^k (1-p)^{i-k} +$ $(M-i)\sigma(1-\sigma)^{M-i-1} [\frac{1}{2} ip(1-p)^{i-1} + \sum_{k=2}^i (1 - \frac{k+1}{2^k}) \binom{i}{k} p^k (1-p)^{i-k}]$
$i+2$	$\binom{M-i}{2} \sigma^2 (1-\sigma)^{M-i-2} [\frac{1}{2} (1-p)^i + \sum_{k=1}^i (1 - \frac{k+2}{2^{k+1}}) \binom{i}{k} p^k (1-p)^{i-k}]$
$i+z$	$\binom{M-i}{z} \sigma (1-\sigma)^{M-i-z} \sum_{k=0}^i (1 - \frac{k+z}{2^{k+z-1}}) \binom{i}{k} p^k (1-p)^{i-k}$

Используя стационарное распределение процесса $\Omega = \{\omega_i\}_{i=0}^M$ можно найти среднее число устройств, требующих передачи, и стационарное значение доли успешных передач:

$$S_{out} = \sum_{i=0}^M S_{out}(i) \omega_i, \bar{n} = \sum_{i=0}^M i \omega_i. \quad (5.28)$$

Средняя задержка при доступе к системе определяется следующим образом:

$$E[\tau] = t_{pr} + K_0 + K + \bar{n}/S_{out} + 1/2. \quad (5.29)$$

Потребление энергии опять-таки возможно получить путем модификации уравнений для PUSCH.

5.9 Сравнение схем доступа для данных

В этом подразделе сравним предложенную схему COBALT с известными механизмами доступа для данных через PUSCH и PRACH в системе

длительной эволюции LTE. Результаты при этом базируются на имитационном моделировании на уровне протоколов (как минимум 100 минут времени LTE на цикл моделирования) и подтверждаются аналитическими изысканиями из раздела 5.8. Начнем сравнение с числа ресурсных блоков системы для доступа данных к каналу и числа ресурсных блоков, реально используемых для устройств МТС, таблица 5.6. Для PUSCH-основанного доступа сеть должна разместить избыточное число ресурсных блоков RB с учетом необходимости поддержки растущей популяции МТС, в то время как эффективность использования ресурсных блоков является очень низкой вследствие природы рассматриваемого МТС трафика, т.е. его редкого неперiodического появления. Важно отметить, что когда число устройств МТС очень велико, система не в состоянии поддерживать все устройства с заданными параметрами из-за слишком высокого уровня накладных расходов на заголовки. В частности, для диапазона 5 МГц число необходимых ресурсных блоков для поддержки 5000 или 10000 устройств выше, чем доступное в подкадре.

Таблица 5.6

Сравнение в потребности ресурсных блоков RB

Использование RB на сабфрейм [назнач. : использ.]	1К	5К	10К
PUSCH	12 : 0.0036	56 : 0.018	112 : 0.036
PRACH	2.4 : 0.50	2.4 : 1.66	2.4 : 2.18
COBALT	4 : 0.07	4 : 0.34	4 : 0.73

В противовес изложенному для PRACH-основанного доступа для данных применяется фиксированное число ресурсных блоков RB с очень высокой эффективностью их использования. Заметим, однако, что, относящиеся к процедуре PRACH данные в таблице 5.6 имеют лучшие оценки, поскольку в реальной ситуации Msg3 передачи также используют и ресурсы процедуры PUSCH. Кроме того, увеличение нагрузки PRACH выражается в росте числа коллизий преамбул. Это может подвергать опасности качество обслуживания для пользователей H2H, которые затем могут пострадать от чрезмерных задержек при входе в сеть, если не будут приняты превентив-

ные меры. Предложенная в диссертационной работе схема, как ожидается, COBALT избавит от перегруженности PRACH и позволит устройствам МТС получать более высокие вероятности доступа к сети и в то же время защитит трафик пользователей Н2Н. Из таблицы 5.6 видно, что даже при минимально возможном числе ресурсных блоков RB для COBALT, сеть LTE не испытывает трудностей для поддержки очень большой популяции устройств МТС.

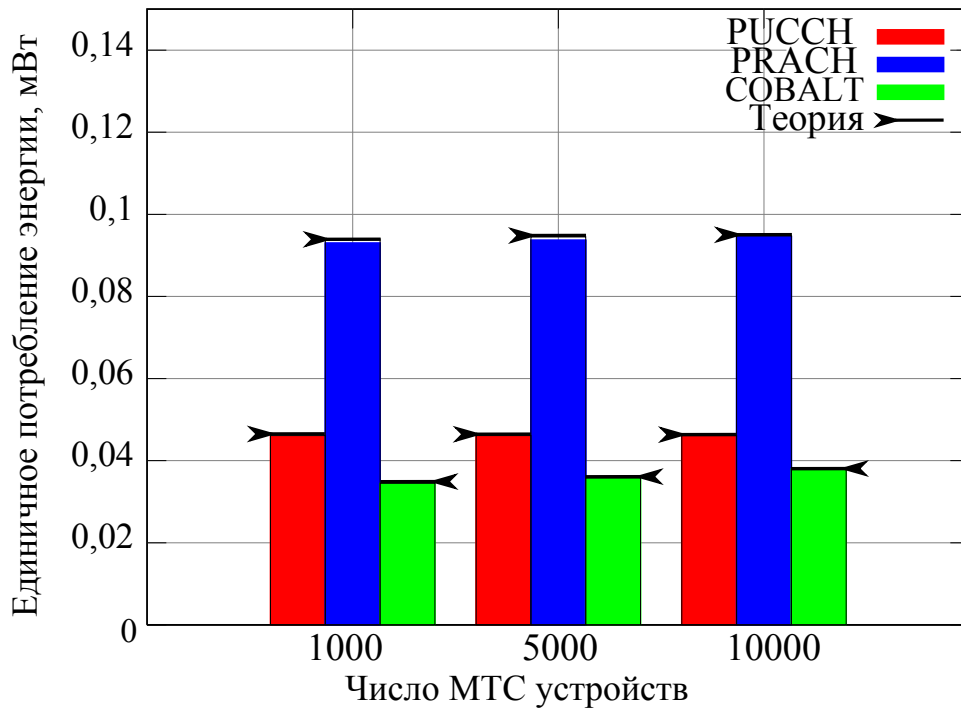


Рисунок 5.17 — Потребление мощности для различных схем доступа

Продолжим исследование и рассмотрим потребление энергии для механизма COBALT – доступа для малых данных. Анализируя рисунок 5.17, можно увидеть, что потребление энергии при процедуре COBALT существенно ниже, чем при использовании PRACH-основанной схемы, особенно при небольших нагрузках. Еще более интересно, что энергетическая эффективность COBALT даже немного лучше, чем PUCCH-основанного доступа, который свободен от конкуренции и, следовательно, высокоэффективен с точки зрения потребления энергии. Это происходит вследствие того, что уменьшается число сигнальных сообщений передаваемых/принимаемых устройством МТС при передаче нерегулярных малых данных посредством COBALT. Кроме того, подчеркнем, что рост потребления энер-

гии при увеличении популяции МТС минимален. Последнее объясняется очень небольшой вероятностью коллизий. Все это показывает потенциальные возможности механизма COBALT для поддержки малых данных при реализации МТС.

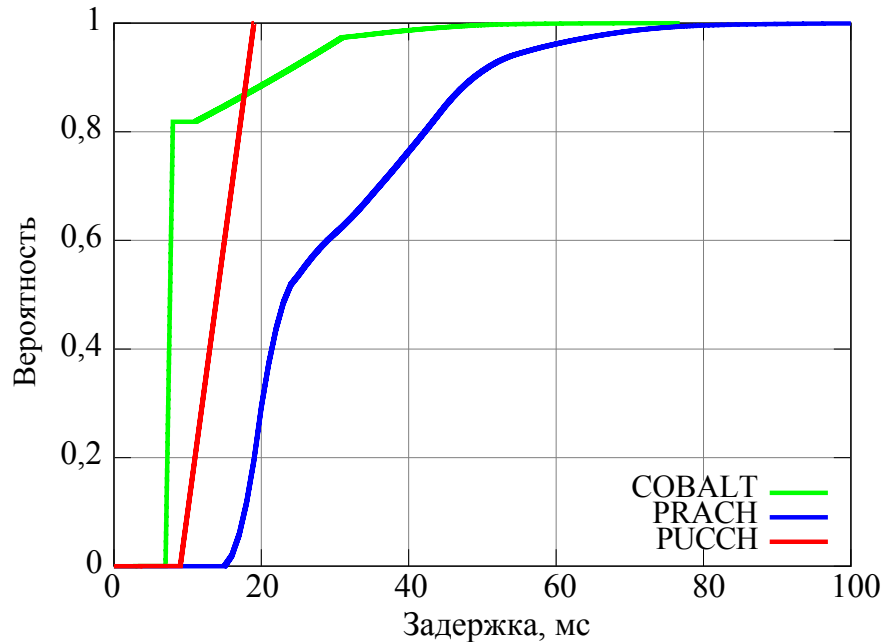


Рисунок 5.18 — Функция распределения задержки для доступа данных

Наконец, исследуем задержку для доступа данных для всех трех альтернативных механизмов (рисунок 5.18 и Таблица 5.7), делая вывод о том, что COBALT-основанный механизм доступа для данных обладает существенно меньшим значением задержки пакетов. Несмотря на то, что большинство видов трафика МТС относится к трафику, толерантному к задержкам, есть достаточно много ситуаций, когда задержка является критичной для качества восприятия пользователем (например, аварийные сообщения и приложения для транспорта), а предложенная в диссертационной работе схема COBALT улучшает задержку по сравнению со схемой PUCCH почти 85% случаев.

В этом разделе было подчеркнуто отсутствие всеобъемлющих систем для оценки эффективности механизма случайного доступа RACH для технологии 3GPP LTE-Advanced, при этом такая оценка должна опираться и на имитационное моделирование, и аналитические результаты. Более того, предыдущие результаты в области оценивания эффективности случайного доступа часто несовместимы и противоречивы из-за того, что унифициро-

Сравнение средней задержки на доступе [моделирование : анализ]

Тип	1К	5К	10К
PUSCH	14.00 : 14.00	14.00 : 14.00	14.00 : 14.00
PRACH	29.08 : 29.13	29.48 : 29.26	29.95 : 29.50
COBALT	7.72 : 7.74	8.81 : 8.81	10.51 : 10.38

ванная методология калибровки 3GPP была разработана совсем недавно. При этом в диссертационной работе были учтены последние стандарты, одобренные 3GPP, при использовании разработанного в диссертации усовершенствованного симулятора случайного доступа RACH на уровне протоколов.

В диссертационной работе также были проанализированы существующие механизмы доступа для данных, которые могли бы быть использованы устройствами 3GPP LTE для передачи своих данных. Было подчеркнuto, что ни используемая по умолчанию PUSCH-основанная схема, ни альтернативная PRACH-основанная схема не являются рациональными для поддержки большого множества устройств МТС, когда поступление трафика является нерегулярным и малым. Для того, чтобы уменьшить вероятность деградации качества обслуживания при использовании существующих схем, в диссертационной работе предложен новый конкурентно-основанный механизм передачи данных для LTE, названный COBALT. Предложенная в диссертационной работе схема имеет простую реализацию и требует меньшего числа сигнальных сообщений в LTE. Поэтому, она демонстрирует существенно лучшее использование сетевых ресурсов, более низкое энергопотребление для устройств МТС и может достаточно часто уменьшить и задержки. Полученные оценки характеристик эффективности на уровне протоколов аналитическими методами и с использованием имитационного моделирования подтверждают, что предложенный конкурентно-основанный механизм обладает необходимым потенциалом для улучшения доступа для малых данных в условиях возрастающего числа приложений LTE на основе МТС. Развитие исследовательских работ в этом направлении состоит в том, чтобы продолжить демонстрацию

преимуществ предложенного подхода в условиях сосуществования пользователей МТС и Н2Н, а также детализировать оптимальную реализацию механизма COBALT для сигнализации LTE.

5.10 Выводы

1. Разработана аналитическая модель для протокола IEEE 802.11ah, позволяющего обеспечить передачу данных для большого числа узлов в едином кластере. Все узлы такой сети представлены предсказуемыми характеристиками трафика, а именно: передачей одного малого сообщения за длительный промежуток времени. Предложенная модель отличается высокой скоростью работы по сравнению с известными моделями, что дает возможность получения результатов на основе имитационного моделирования с сохранением высокой точности.
2. Разработана математическая модель для анализа поведения канала множественного доступа IEEE 802.11ah, сформированы диаграмма состояний, матрица переходных вероятностей и соответствующий вектор стационарного распределения. Определены полезная пропускная способность канала (бит сообщения без учёта заголовков в секунду) и энергетические затраты устройств на передачу.
3. Разработаны Модель и метод доступа устройств M2M к ресурсам сети LTE на основе модифицированной процедуры COBALT, учитывающие особенности передачи данных малого объема, позволяющие уменьшить задержку пакетов по сравнению с методом PUSCH почти для 85% типов трафика и одновременно в 2-3 раза снизить энергопотребление по сравнению с методом PRACH.

Раздел 6. Разработка моделей и методов распределения ресурсов в гетерогенных сетях дальнего радиуса действия NB-IoT для приложений Интернета вещей

6.1 Введение

В последние десятилетия Интернет вещей (Internet of Things, IoT) стремительно развивается: начиная с технологии радиочастотной идентификации (Radio frequency identification, RFID) в начале 1980-х годов, в середине 1990-х произошел быстрый скачок в беспроводных сенсорных сетях (wireless sensor networks, WSNs). На сегодняшний день, несмотря на ограниченную функциональность, сегмент IoT становится все более массивным и взаимосвязанным с другими технологиями [277]. В дополнение к недорогому оборудованию, включая датчики, приводы и интеллектуальные счетчики, IoT следующего поколения готовится охватить более продвинутое сетевые устройства с повышенной степенью сложности и неоднородности, такие как самоуправляемые автомобили с подключением к сети Интернет, портативные гаджеты, промышленные роботы и (скоро) летающие квадрокоптеры [278–280].

Сегодня мы становимся свидетелями устойчивого и постоянного проникновения концепций IoT в автомобильную промышленность, где появляется возможность управлять *отдельными транспортными средствами (далее подвижные объекты) через облачные сети*, которые глубоко интегрированы с существующими системами кэширования и вычислительной техники [281;282]. Например, возможны комплексные решения между множеством разнородных IoT устройств, таких как интеллектуальные дорожные знаки, светофоры и уличные камеры.

Эта новая тенденция открывает двери к колоссальным взаимным выгодам. С одной стороны, подключенные транспортные средства (объекты) становятся способными получать важные, своевременные обновления и предупреждения от дорожной инфраструктуры, повышая безопасность и

удобство вождения. С другой стороны, появляется возможность управлять данными от сенсоров и значительно сэкономить заряд аккумуляторов.

Такое видение развития IoT выдвигает новую парадигму коллективного восприятия, так называемый *crowdsensing* [283], что означает краудсорсинг данных датчиков с мобильных устройств, которые совместно измеряют, сопоставляют, либо анализируют определенные сведения, данные [284]. Согласно краудсорсингу, глубокое слияние встроенных и подвижных сегментов обещает принести дополнительные выгоды этим ранее непересекающимся областям. Безопасный и надежный автомобильный транспорт трудно представить без разумного и широкого использования передаваемой датчиками информации, предоставляемой дорожной инфраструктурой [285], а транспортные средства, подключенные к сети Интернет, могут предложить свои уникальные преимущества [286].

Резкий скачок вперед в гибких приложениях краудсорсинг может произойти, если люди будут соответствующим образом заинтересованы вносить свой вклад, используя возможности радиосвязи своих подключенных транспортных средств. Следовательно, автоматическая ретрансляция чувствительных к задержке данных от датчиков IoT может повысить общую надежность и энергоэффективность передачи информации. Такое тесное взаимодействие между встроенными IoT устройствами и сетевыми транспортными средствами обещает не только поддерживать существующие приложения для краудсорсинга, но также позволяет предоставлять новые услуги в областях коллективной безопасности дорожного движения [287], экологичного транспорта [288] и многих других областях [289; 290].

6.2 Технологии дальнего радиуса действия для реализации решений Интернета вещей

Низкое энергопотребление, доступная цена на чипсет, высокая дальность связи и возможность размещения массивных развертываний устройств являются основными требованиями к новым решениям IoT, ко-

торые в совокупности можно назвать энергоэффективной сетью дальнего радиуса действия (Low-Power Wide-Area Networks, LPWANs). Разработка таких технологий была начата и развивалась параллельно в рамках нескольких ведущих организаций по стандартизации, включая IEEE, ETSI и 3GPP, а также в рамках отдельных компаний и небольших промышленных консорциумов [291;292]. Две из самых популярных технологий дальней связи LPWAN, основанных на запатентованных стандартах, это SIGFOX и LoRaWAN. В то время как оба варианта работают в зонах ISM, не требующих лицензирования, и топологически привязаны к сотовой структуре, но соответствующие принципы их работы принципиально различаются [293].

Технология SIGFOX работает с ультра-узкополосными сигналами: 100 Гц и 600 Гц для восходящей и нисходящей линии связи соответственно. Достижимая скорость передачи данных по восходящей линии связи составляет приблизительно 100 бит/с [294], и хотя полезная емкость кадра не может превышать 12 байт, эфирное время каждого кадра составляет около 1 секунды. Другими ограничениями SIGFOX являются использование единой схемы модуляции и кодирования (modulation and coding scheme, MCS) и ограниченное количество пакетов восходящей и нисходящей линии связи: 140 и 4 на устройство соответственно. Несмотря на то, что устройства SIGFOX отличаются низким энергопотреблением и длительным сроком службы батареи, присущие им ограничения сужают спектр использования данной технологии для будущих crowdsensing приложений.

В отличие от SIGFOX, LoRaWAN поддерживает несколько схем модуляции и кодирования (MCSs), результатом чего является увеличение скорости передачи данных до 50 Кбит/с, что значительно расширяет диапазон приложений [295]. Несмотря на то, что LoRaWAN, так же как и SIGFOX, слабо развернуты и реализованы некоторыми производителями чипсетов на рынке, промышленная поддержка этих технологий как со стороны телекоммуникационного, так и автомобильного сообщества довольно низка, так как ни одна из технологий не была глобально стандартизирована. Таким образом, в то время как эти решения могут найти свое место на ранних стадиях развития IoT, их долгосрочное глобальное проникновение в нашу жизнь может быть сомнительным.

И наоборот, семейство технологий IoT 3GPP в основном ориентировалось на масштабные сценарии с самых ранних этапов их развития. Во-первых, в LTE Release 12, упрощенная категория Cat. 0 пользовательского оборудования (User equipment, UE) представлена с пропускной способностью 20 МГц и пиковыми скоростями передачи данных 1 Мбит/с. Также у такой категории устройств примерно на 50% меньше логическая сложность по сравнению с устаревшей Cat.1. Новая категория пользовательских устройств, Cat. M1, была разработана для работы с полосой частот 1.4 МГц и позволила уменьшить максимальную мощность передачи до 20 дБм, при этом все еще обеспечивая существенное покрытие [296]. Эта категория является частью LTE Release 13 и, как ожидается, будет в несколько раз менее сложна, чем Cat. 1 UE. Кроме того, для развивающихся стран с низким уровнем проникновения LTE было разработано основанное на GSM решение IoT под названием extended coverage GSM (EC-GSM). EC-GSM теоретически может обеспечить емкость соты до 50К устройств, а ее покрытие увеличиться до 20 дБ.

Согласно последним разработкам 3GPP относительно сотового IoT (CIoT) [297], была утверждена работа над стандартом сотовой связи для устройств телеметрии с низкими объемами обмена данными (Narrowband IoT, NB-IoT). NB-IoT — решение, которое может быть легко интегрировано в существующие сети. Полоса пропускания дополнительно сужается до 180 кГц, обеспечивая более высокий коэффициент усиления 20 дБ. Процесс стандартизации NB-IoT был завершен (в июне 2016 года), и эта технология официально включена в LTE Release 13 [298].

В NB-IoT изменная полоса пропускания по-прежнему достаточна для обеспечения возможности соединения с тысячами подключенных машин: полоса может быть разделена на 12 поддиапазонов по 15 кГц каждый с общей скоростью передачи данных до 250 Кбит/с в восходящей линии связи и до 227 Кбит/с в нисходящем канале. Это существенное масштабирование в 5 МГц может обеспечить покрытие до 10 км в городе и до 30 км в сельских районах, что на порядок больше, чем в «ориентированном на человека» LTE [299]. Более суженная полоса пропускания также способствует упрощению радиоэлементов в устройствах конечного пользователя, что,

следовательно, ведет к снижению затрат на пользовательское оборудование. По той же причине размеры транспортных блоков как в восходящей, так и в нисходящей линиях снижаются, а также упрощаются процедуры сигнализации. В результате предполагается, что абонентское оборудование для NB-IoT будет значительно дешевле, чем для LTE, и по цене конкурировать с технологиями LoRaWAN и SIGFOX. Таким образом, NB-IoT является самой многообещающей технологией действительно массового развертывания услуг Интернета вещей.

Далее в этом разделе будут рассмотрены стратегии распределения ресурсов при развертывании технологии NB-IoT на сетях LTE, а также технологические аспекты для взаимодействий между транспортными средствами и датчиками, установленными оператором сети. Также будут рассмотрены и оценены преимущества использования ретрансляции для сетей NB-IoT.

6.3 Системная модель для анализа технологии NB-IoT

Комбинация двух или более гетерогенных сетей увеличивает интеллектуальность и надежность систем принятия решений. Например, сигнал об обнаружении чрезвычайных ситуаций, сообщаемый группой интеллектуальных счетчиков, может быть сразу проверен через визуальный канал. Кроме того, в комбинации таких гетерогенных сетей используется ряд задач, в которых надежность доставки потоков данных играет ключевую роль. Учитывая ограниченный объем радиоресурсов, выбор стратегии распределения радиоресурсов между меньшим числом высокоскоростных видеокамер и большим количеством низкоскоростных смарт-счетчиков является важной задачей для операторов беспроводных сетей. В рамках этой задачи необходимо сбалансировать вероятностно-временные характеристики (VBX), такие как вероятности потери сеанса для отдельных потоков данных.

В данном разделе разработана аналитическая модель для исследования ВВХ обработки трафика NB-IoT при наличии конкурирующего трафика LTE. Используя разработанный инструмент, далее сравним три возможные стратегии распределения радиоресурсов: статическую, динамическую и динамическую с резервированием. В качестве эталонной модели рассматривается внутриполосный режим, представленный в LTE Release 13 [296;297], где радиостанция LTE разделяет спектр частот с технологией NB-IoT. Эта комбинация была выбрана в силу того, что 1) LTE и NB-IoT – единственные два решения, которые используют один и тот же спектр частот; 2) эта комбинация получает наибольшую поддержку со стороны отрасли Интернета вещей и, как ожидается, станет основным стандартом для сбора больших данных в мире Интернета вещей.

В этом разделе вводится понятие оператора, реализующего службу наблюдения, используя возможности технологий LTE и NB-IoT. Далее формализуется модель LTE-соты с поддержкой технологии NB-IoT и вводятся стратегии распределения ресурсов.

Рассмотрим «оператора», который управляет системой зонального наблюдения. Система представляет собой гибридное решение, состоящее из 1) некоторого числа видеокамер для проведения видеомониторинга, 2) многочисленных датчиков для обнаружения вторжений, пожаров, а также для контроля температуры и давления воздуха и 3) системы обработки больших объемов данных, получаемых как с видеокамер, так и со смарт-датчиков.

Поскольку проводное и Wi-Fi соединения не всегда доступны в районах наблюдения, передача трафика от камер наблюдения производится при помощи сотовой сети LTE. Предполагается, что смарт-датчики подключены по NB-IoT, которая является одной из наиболее перспективных технологий IoT-отрасли. Такая комбинация решений может эффективно сосуществовать в рамках одной сети. Поскольку оба решения работают на лицензированных частотах, поставщик услуг, ориентированный на надежность связи, будет приобретать определенный частотный ресурс у оператора сети через механизм «Network slicing». Такой подход может обеспечить

бесперебойное обслуживание независимо от того, в каком состоянии находятся оставшиеся частотные полосы оператора.

Сценарий развертывания, описанный выше, предполагает совместное использование предоставленных ресурсов LTE между двумя типами конечных узлов: камерами наблюдения и датчиками NB-IoT, поднимая вопрос об эффективном распределении ресурсов. В этой связи характеристики надежности обоих потоков трафика приобретают первостепенное значение, поскольку сервер обработки не сможет принимать обоснованные решения, если часть потока станет временно недоступна. Одним из возможных примеров может служить система обнаружения пожара, которая развернута в определенных зонах, где любое ложное предупреждение является чрезвычайно дорогостоящим (например, для фабрики). Таким образом, если блок обработки получает предупреждение о пожаре от датчиков дыма/температуры, он сначала перекрестно проверяет его с помощью визуального канала (видеопоток от ближайшей камеры) и только затем передает сигнал аварийной эвакуации.

Для решения вышеперечисленных вопросов подробно рассмотрим модель системы, а затем приступим к формализации стратегий распределения ресурсов.

6.3.1 Модель соты LTE

Рассмотрим соту сети LTE с базовой станцией, расположенной в центре соты. Напомним, что радиус покрытия ячейки для услуг LTE, R_L , намного меньше по сравнению с покрытием технологии NB-IoT, R_N . Весь набор доступных ресурсов будем измерять в каналах NB-IoT. В дальнейшем будем называть это базовым каналом. В направлении восходящей линии для NB-IoT доступны C каналов, которые можно вычислить как $C = cS$, где S — число ресурсных блоков (RB), c — число базовых каналов в одном RB.

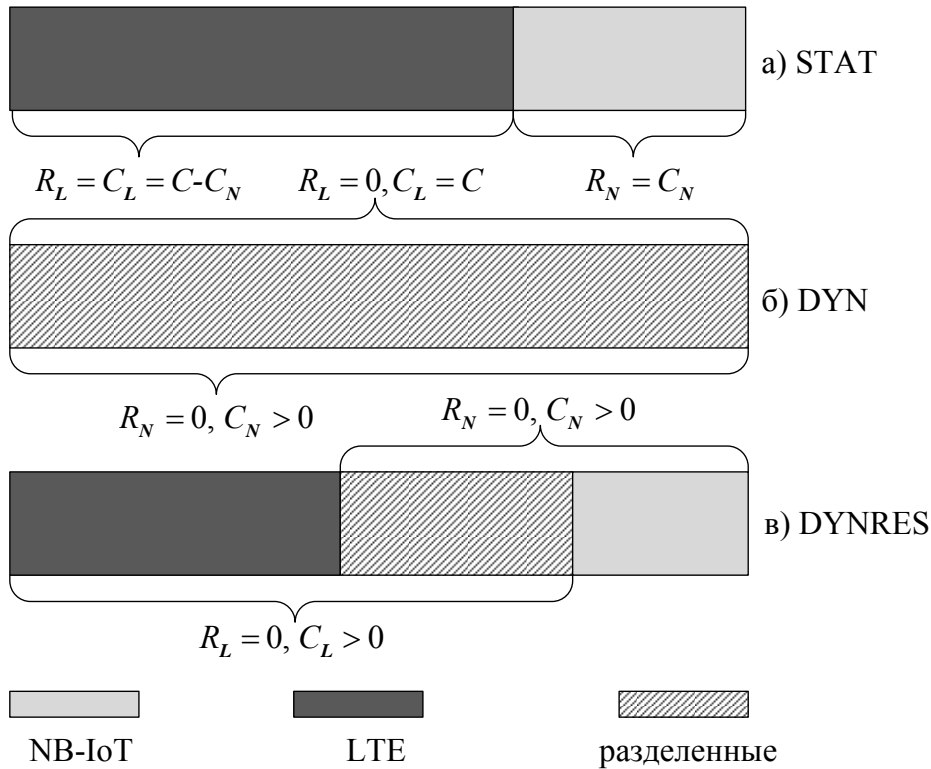


Рисунок 6.1 — Предлагаемые стратегии распределения ресурсов

Входящий поток от сессий LTE является пуассоновским с интенсивностью ν . Время предоставления LTE-услуги распределено по экспоненциальному закону со средним $1/\mu$, а минимальный объем ресурсов, запрошенных на установление соединения на каждом временном интервале восходящей линии, составляет d базовых каналов. Пусть $a = \nu/\mu$ — предложенная нагрузка LTE. Запросы на соединение от NB-IoT поступают в соответствии с процессом Пуассона с интенсивностью λ . Каждое поступление данных характеризуется экспоненциально распределенным временем обслуживания с параметром θ . Для каждого сеанса NB-IoT требуется b базовых каналов. Интенсивность предложенной нагрузки на передачу блоков данных от NB-IoT обозначим через $\rho = \lambda\theta$. Аналогично сессиям LTE, устройства NB-IoT считаются статичными в течение всей сессии.

6.3.2 Стратегии распределения ресурсов

Рассматриваемые стратегии распределения ресурсов между камерами LTE и измерительными датчиками NB-IoT показаны на рисунке 6.1. Отметим, что максимальное число базовых каналов, которые могут быть выделены для NB-IoT и LTE, составляет $C_N = C - R_L$ и $C_L = C - R_N$ соответственно, где R_L и R_N — минимальное число каналов, которые всегда доступны и зарезервированы для трафика LTE и NB-IoT соответственно. Далее в данном разделе представлено сравнение следующих трех стратегий распределения ресурсов:

- *Статическая стратегия (STAT)*. Эта стратегия соответствует случаю, когда минимальное и максимальное число базовых каналов, выделенных для NB-IoT и LTE, совпадают: $R_L = C_L$, $R_N = C_N$. Другими словами, все ресурсы строго разделены между NB-IoT и LTE.
- *Динамическая стратегия (DYN)*. В этом случае минимальный объем ресурсов не назначается NB-IoT и LTE, то есть $R_L = R_N = 0$. Однако максимальный объем ресурсов и $C_L = C_R = C$ полностью разделены между NB-IoT и LTE.
- *Динамическая стратегия с резервированием (DYNRES)*. В этой стратегии максимальные распределения ресурсов, доступные для NB-IoT и LTE, определены так, что $C_N = C - R_L > 0$ и $C_L = C - R_N$. При этом $R_N > 0$, $R_L > 0$ определяют минимальный объем ресурсов, предназначенных для NB-IoT и LTE соответственно. Остальные ресурсы динамически распределяются между двумя типами трафика.

Уникальной особенностью рассматриваемой системы с двумя разнотипными входящими потоками является тот факт, что базовые каналы будут последовательно выделяться с учетом особенностей внутрисполосной технологии NB-IoT. Процесс управления ресурсами показан на рисунке 6.2. Рассмотрим систему без активных сессий и предположим, что поступает новая сессия от NB-IoT. В этом случае RB становится доступным для

NB-IoT, где каждый запрос на принятие данных требует ровно b базовых каналов NB-IoT. Таким образом, новая сессия NB-IoT занимает базовые каналы в этом RB. Если при следующем установлении сессии NB-IoT все каналы в данном RB будут уже заняты, то будет выделен следующий RB, доступный для услуг NB-IoT.

6.4 Модель оценки показателей производительности

Система, описанная в разделе 6.3, может быть представлена в виде системы массового обслуживания с двумя входящими потоками, имеющими разные дисциплины обслуживания с выделением ресурсных блоков. Отметим, что система позволяет управлять потерями сессий как NB-IoT, так и LTE. Основное отличие модели, представленной ниже, заключается в ком-

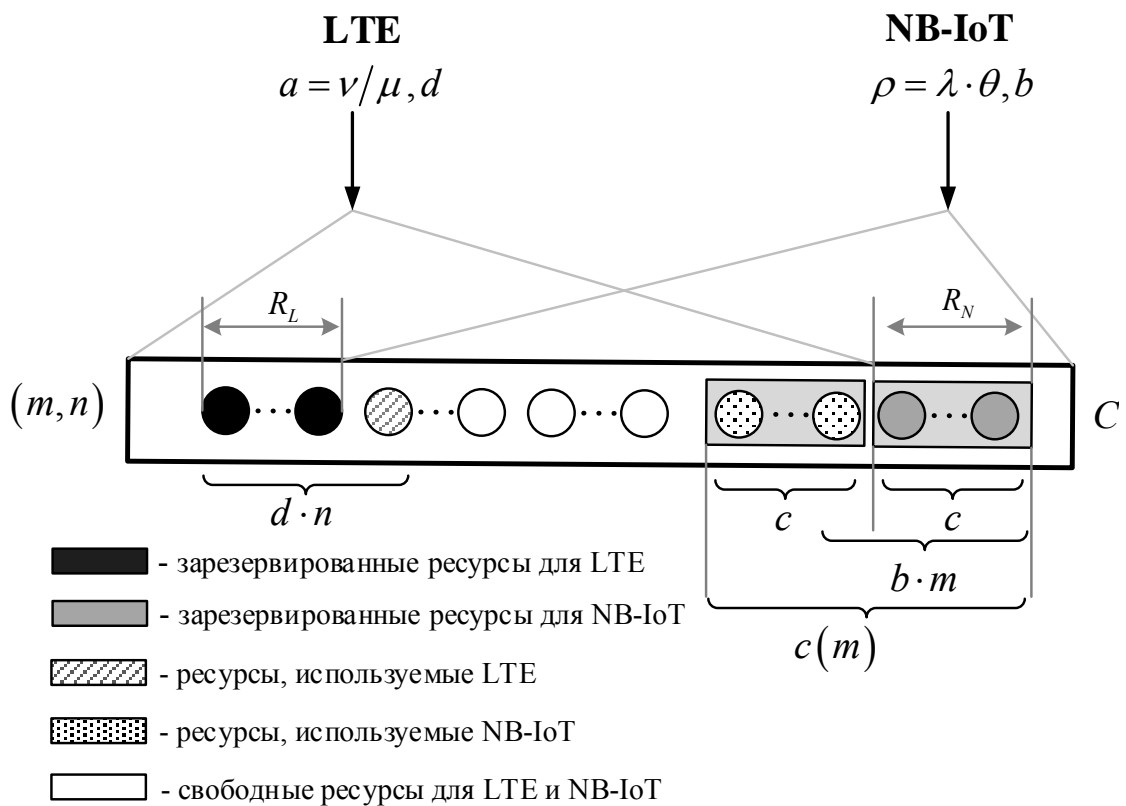


Рисунок 6.2 — Схема последовательного распределения ресурсов соты сети LTE

плексной последовательной стратегии распределения ресурсов на основе блоков.

Определим состояние системы. Отметим, что запрос на обслуживание от LTE принимается, если в момент его поступления доступно не менее d из $C_L = C - R_N$ каналов, таким образом уменьшается число доступных базовых каналов на величину d и число RB на $\lceil d/c \rceil$. Количество сессий LTE $\lceil d/c \rceil$ всегда является целым числом $\lceil d/c \rceil \geq 1$. Если при принятии запросов на соединение от NB-IoT число базовых каналов в текущем RB выше, чем b , то сессия NB-IoT принимается в этом RB. Если число доступных базовых каналов ниже, чем b , и новый RB доступен для трафика NB-IoT, то этот RB принимает на обслуживание поступивший запрос от NB-IoT, тем самым уменьшая число базовых каналов в этом RB на b . В других случаях сессия NB-IoT теряется.

Пусть $m(t)$, $n(t)$, $t > 0$, определяют число активных сессий NB-IoT и LTE соответственно. Следовательно, состояние системы соты сети LTE, обслуживающей трафик NB-IoT и LTE, может быть описано двумерным случайным процессом $\{m(t), n(t), t > 0\}$ со следующим пространством состояний:

$$\mathcal{X} = \{m \geq 0, n \geq 0 : nd \leq C - R_N, c(m) \leq C - R_L, nd + c(m) \leq C\}, \quad (6.1)$$

где $c(m) = c \lceil bm/M \rceil$ — число базовых каналов, занятых сессиями NB-IoT, M — максимальное число сессий в NB-IoT, которые могут быть обслужены в одном RB, $M = \lceil c/b \rceil$ — число базовых каналов, необходимых для обслуживания этих сессий.

Заметим, что $\{m(t), n(t), t > 0\}$ является марковским процессом. Введем $p(m, n)(t)$, $\{m, n\} \in \mathcal{X}$, — стационарное распределение вероятностей m сессий NB-IoT и n сессий LTE в системе в момент времени t :

$$p(m, n) = \lim_{t \rightarrow \infty} p(m, n)(t), \{m, n\} \in \mathcal{X}. \quad (6.2)$$

Чтобы получить уравнение локального баланса, рассмотрим произвольный контур в диаграмме перехода состояний. Стационарное распреде-

ление вероятностей случайного процесса $\{m(t), n(t), t > 0\}$ удовлетворяет следующим уравнениям локального баланса:

$$\begin{aligned} p(m, n) \frac{c(m)}{\theta} &= p(m-1, n) \lambda, \quad m > 0, (m, n) \in \mathcal{X}, \\ p(m, n) \mu n &= p(m, n-1) \nu, \quad n > 0, (m, n) \in \mathcal{X}, \end{aligned} \quad (6.3)$$

и представляется следующим выражением:

$$p(m, n) = G^{-1}(\mathcal{X}) \left(\frac{\rho}{Mb} \right)^m \left(\prod_{i=1}^m \left[\frac{i}{M} \right] \right)^{-1} \frac{a^n}{n!}, \quad (6.4)$$

где константа $G(\mathcal{X})$ имеет следующий вид:

$$G(\mathcal{X}) = \sum_{(m, n) \in \mathcal{X}} \frac{\rho}{Mb} \left(\prod_{i=1}^m \left[\frac{i}{M} \right] \right)^{-1} \frac{a^n}{n!}. \quad (6.5)$$

6.4.1 Характеристики обслуживания

Получение формул вероятностей блокировок сессий является трудоемким вычислительным процессом, так как пространство состояний системы довольно велико и число состояний может достигать нескольких десятков тысяч. Чтобы снять это ограничение, далее разработан специальный численный алгоритм, основанный на рекуррентной связи между ненормированными вероятностями макросостояний.

Определим разбиение пространства состояний \mathcal{X}

$$\mathcal{X} = \bigcup_{s=0}^S \mathcal{X}_s, \quad (6.6)$$

где $\mathcal{X}_s = \{(m, n) \in \mathcal{X} : c(m) = sc\}$.

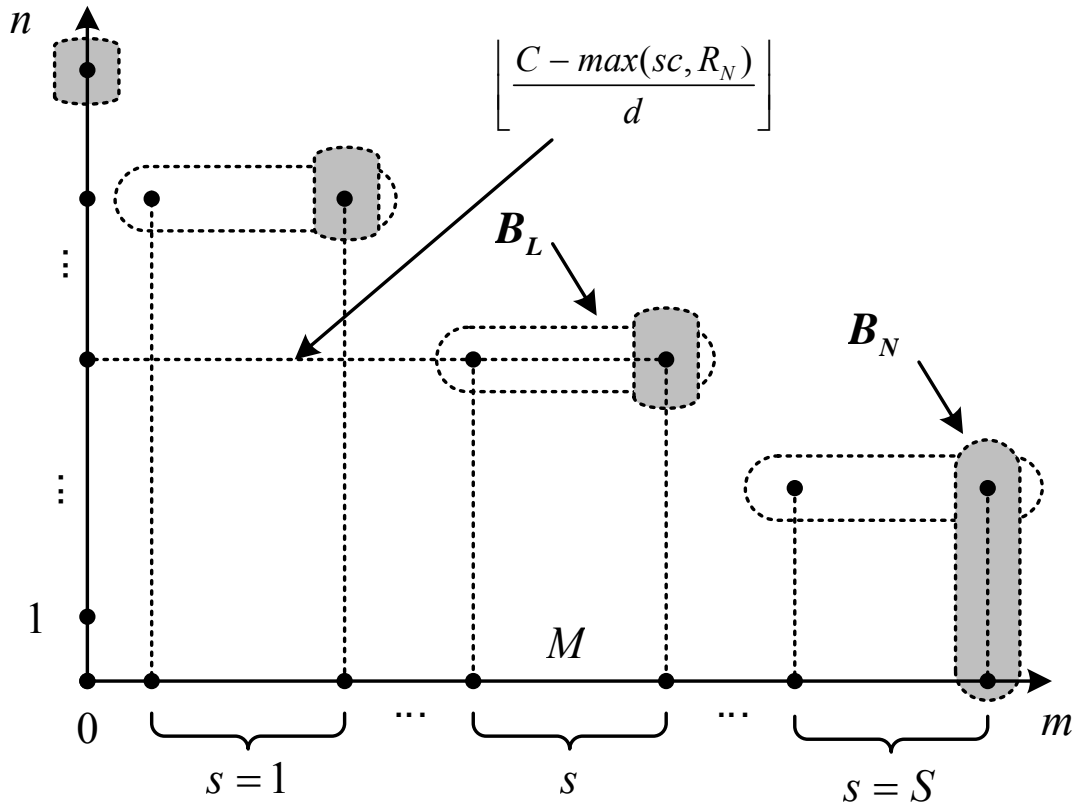


Рисунок 6.3 — Разбиение пространства \mathcal{X} на макросостояния

Разбиение пространства состояний на макросостояния показано на рисунке 6.3. Здесь также иллюстрируются подпространства, соответствующие состояниям блокировки сессий LTE и NB-IoT. Зная стационарное распределение состояний системы, представленное в (6.4) и (6.5) и действующее над (6.6), искомые вероятности блокировок можно легко вывести. В частности, используя (6.6), получим вероятность блокировки сессий LTE:

$$\begin{aligned}
 p_L = & p\left(0, \left\lfloor \frac{C - R_N}{d} \right\rfloor\right) + \\
 & + \sum_{s=1}^{\left\lfloor \frac{C - R_L}{c} \right\rfloor} \sum_{m=[s-1]M+1}^{sM} p\left(m, \left\lfloor \frac{C - \max[sc, R_N]}{d} \right\rfloor\right). \quad (6.7)
 \end{aligned}$$

Аналогичные вычисления проводятся для подмножества блокировок сессий NB-IoT, что в конечном итоге дает выражение для вычисления ве-

роятности блокировок сессий NB-IoT в следующем виде:

$$\begin{aligned}
 p_N = & \sum_{n=0}^{\lfloor \frac{R_L}{d} \rfloor} p \left(\left\lfloor \frac{C - R_L}{c} \right\rfloor M, n \right) + \\
 & + \sum_{s=\lfloor \frac{R_N}{c} \rfloor}^{\lfloor \frac{C-R_L}{c} \rfloor - 1} \sum_{n=\lfloor \frac{C-(s+1)c}{d} \rfloor}^{\lfloor \frac{C-\max(sc, R_N)}{d} \rfloor} p(sM, n). \tag{6.8}
 \end{aligned}$$

Средняя длительность сессии принимает вид

$$E[T_N] = \frac{\sum_{m=0}^{\lfloor \frac{C-R_L}{b} \rfloor} \sum_{n=0}^{\lfloor \frac{C-\max(\lceil \frac{m}{M} \rceil c, R_N)}{d} \rfloor} \lceil \frac{m}{M} \rceil p(m, n)}{\lambda (1 - p_N)}. \tag{6.9}$$

Среднее число базовых каналов, занятых одной сессией NB-IoT, имеет вид

$$E[b_N] = M \sum_{m=0}^{\lfloor \frac{C-R_L}{b} \rfloor} \sum_{n=0}^{\lfloor \frac{C-\max(\lceil \frac{m}{M} \rceil c, R_N)}{d} \rfloor} \lceil \frac{m}{M} \rceil p(m, n). \tag{6.10}$$

Среднее число каналов, занятых сессиями LTE, равно

$$E[b_L] = d \sum_{m=0}^{\lfloor \frac{C-R_L}{b} \rfloor} \sum_{n=0}^{\lfloor \frac{C-\max(\lceil \frac{m}{M} \rceil c, R_N)}{d} \rfloor} n p(m, n). \tag{6.11}$$

Среднее число базовых каналов, занятых обоими типами трафика, определяется через $E[b_{NL}] = E[b_N] + E[b_L]$.

6.4.2 Численный анализ

Для получения количественной и качественной оценки поведения рассматриваемой системы в данном разделе проводится численный анализ предложенных стратегий распределения ресурсов. Входные параметры для численного анализа представлены в таблице 6.1.

Таблица 6.1

Параметры системы для численного анализа

Параметр	Описание	Значение
C	Количество базовых каналов в соте LTE	100
c	Количество базовых каналов в RB	4
R_N	Количество каналов для NB-IoT	$[0, \dots, 100]$
R_L	Количество каналов для LTE	$[0, \dots, 100]$
b	Количество каналов для сессий NB-IoT	1
d	Количество каналов для LTE-сессий	4
θ	Средняя скорость сессий NB-IoT	100 кбит
$1/\mu$	Средняя продолжительность LTE-сессий	10 сек.
λ	Интенсивность входящего потока NB-IoT	10/мин.
ν	Интенсивность входящего потока LTE	1/мин.

На рисунке 6.4 рассматривается пример, иллюстрирующий значения вероятности блокировки сессии LTE в зависимости от общего числа подключенных датчиков NB-IoT при выбранном статическом распределении ресурсов STAT, в котором общий объем радиоресурсов строго разделен между двумя типами трафика. Для этой стратегии вводится коэффициент разделения ресурсов γ , который характеризует объем ресурсов, зарезервированных для NB-IoT. Заметим, что вследствие жесткого разделения ресурсов между LTE и NB-IoT, характеристики стратегии STAT зависят только от выбранного значения γ . Причиной этому является то, что стратегия STAT создает две независимые виртуальные беспроводные системы: одну для LTE, другую для NB-IoT. Хотя эта стратегия является самой простой с точки зрения ее реализации, ее можно рекомендовать только

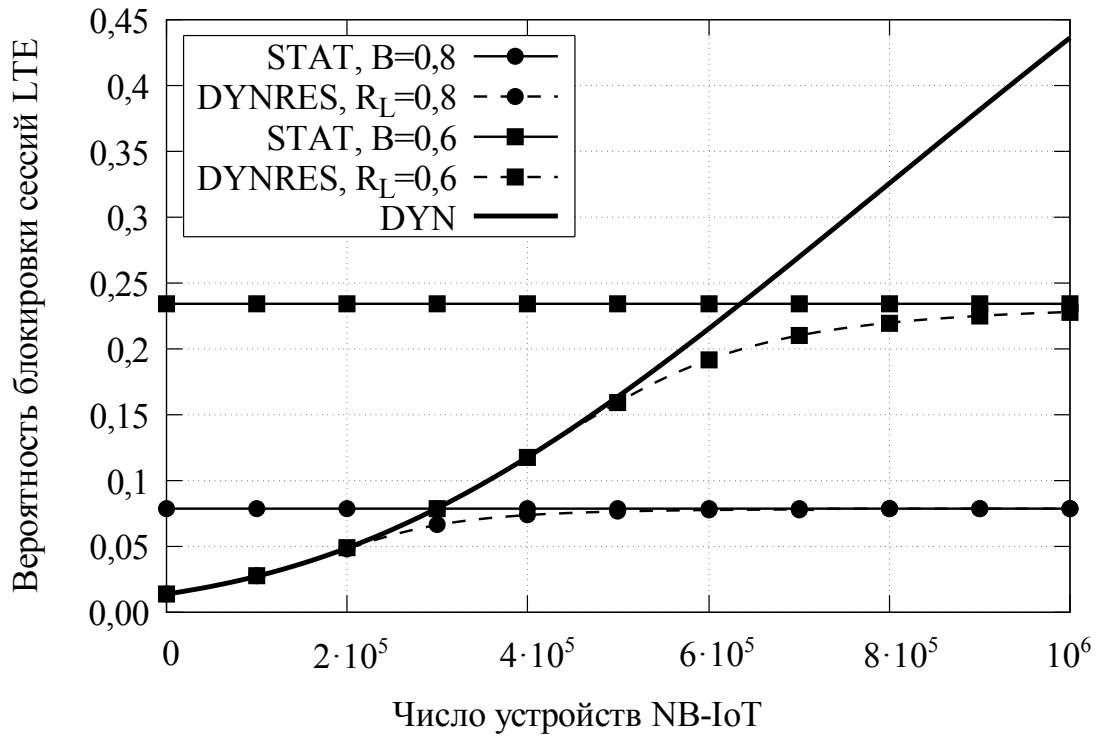


Рисунок 6.4 — Вероятность блокировки LTE в зависимости от предложенной нагрузки NB-IoT

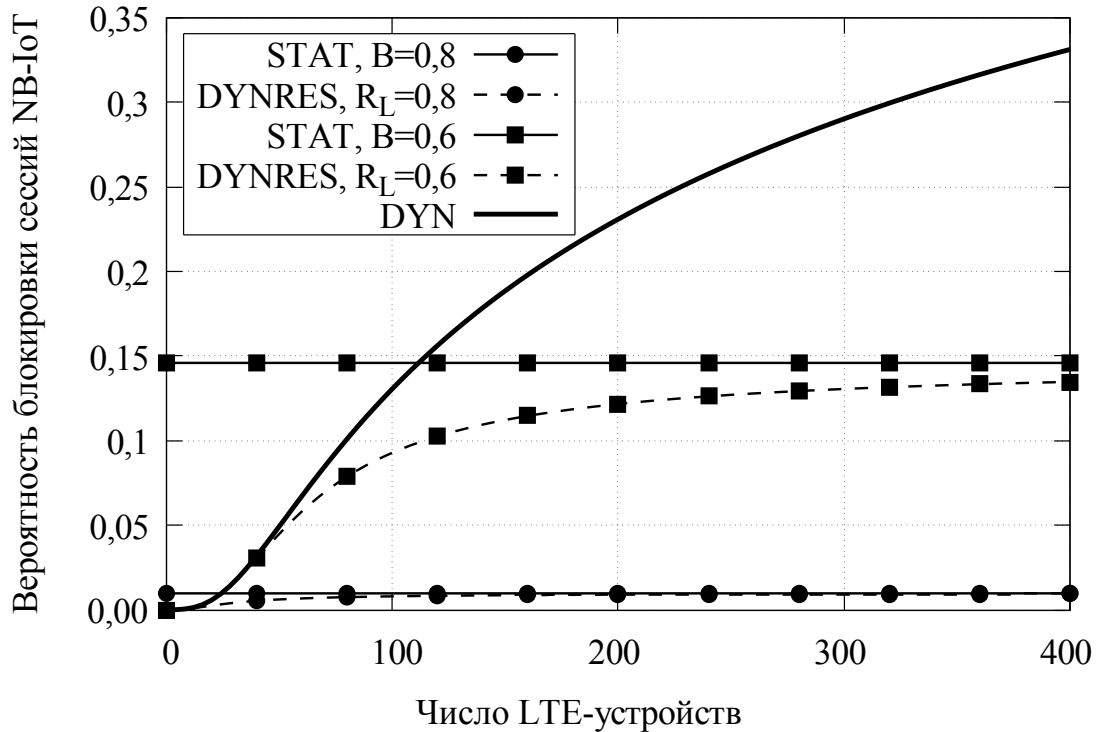


Рисунок 6.5 — Вероятность блокировки NB-IoT в зависимости от предложенной нагрузки LTE

для таких систем, где средняя нагрузка от всех параллельно выполняемых услуг 1) редко изменяется во времени, 2) известна заранее или 3) может быть хорошо предвиденным. Подобное поведение системы редко наблюдается на практике.

Значения вероятности блокировки в стратегии DYN, где общий объем ресурсов доступен для обоих трафиков, сильно зависит от текущей нагрузки из сегмента NB-IoT. Для меньших значений числа подключенных устройств NB-IoT стратегия DYN превосходит STAT, так как позволяет всю полосу частот переключить в режим доступности для LTE-устройств в отсутствие трафика NB-IoT. Рост трафика NB-IoT сопровождается увеличением значений вероятности блокировки для стратегии DYN, поскольку трафик NB-IoT начинает конкурировать за ресурсы с постоянной нагрузкой, исходящей от трафика LTE. Наконец, с поступлением большого числа устройств NB-IoT трафик начинает доминировать в общем диапазоне. В итоге вероятность блокировки будет выше, чем в стратегии STAT, независимо от выбранного значения коэффициента распределения ресурсов. Стратегия распределения ресурсов DYN может быть рекомендована для систем, где средние нагрузки, требуемые каждым потоком (измеренные в Гц/с), сопоставимы. В то же время стратегия DYN остается адекватным решением для нестабильных систем, где нагрузка может сильно изменяться во времени.

Перейдем к третьей стратегии, DYNRES, где определенный объем ресурсов резервируются для трафика каждого типа, а остальные ресурсы динамически распределяются между ними. Значения R_L выбираются эквивалентными соответствующим значениям γ в ранее рассмотренной стратегии выделения ресурсов STAT, а R_N считается нулевым. Другими словами, сравниваются статические и полудинамические варианты распределения. Заметим, что для меньшего числа устройств NB-IoT характеристики стратегии DYNRES идентичны характеристикам стратегии DYN. Однако увеличение значений вероятности блокировок для DYNRES растет медленнее, чем в случае DYN. Причиной этому является тот факт, что часть ресурсов, которые мог бы занять LTE-трафик, занята динамическим разделением. Кроме того, при более высоких нагрузках трафика NB-IoT вероятность

блокировки LTE для DYNRES никогда не будет превышать соответствующее значение для системы со стратегией STAT. Для крайне большого числа устройств NB-IoT кривая для DYNRES асимптотически приближается к кривой для системы STAT, но не пересекает ее.

Стратегия DYNRES не хуже, чем стратегия STAT и DYN. Объясняется это тем, что стратегия DYNRES включает основные свойства первых двух схем распределения ресурсов. Стратегия DYNRES ведет себя так же, как и стратегия DYN при низких нагрузках и стратегия STAT при высоких нагрузках. Следовательно, стратегия DYNRES может быть рекомендована для повышения надежности системы.

Представим в ортогональном разрезе рисунке 6.5, который показывает вероятность блокировки сеанса NB-IoT в сравнении с общим числом подключенных LTE-устройств. Как и в предыдущем случае, будем подбирать параметры для стратегии DYNRES таким образом, чтобы объемы ресурсов, зарезервированных для NB-IoT в STAT и DYNRES, были равными. Важно отметить, что в этом случае R_N равно $(1 - \gamma)$ вместо γ . Значения R_L считаются равными нулю. Для такого симметричного случая наблюдается то же качественное поведение, что и на рисунке 6.4. Тем не менее имеются заметные количественные различия между рисунке 6.4 и рисунке 6.5. Значения вероятности блокировки NB-IoT на порядок ниже, чем для LTE. Это связано с более низкими потребностями в ресурсах отдельной сессии NB-IoT, что, в свою очередь, ведет к увеличению шансов на прием обслуживания даже в почти полностью загруженной системе. Кроме того, способ выделения каналов NB-IoT во внутриполосном режиме также соответствует режиму движения датчика. Сессии LTE занимают весь RB, тогда как трафик NB-IoT получает полный RB, хотя одна сессия и не занимает его полностью. Поэтому следующие несколько сессий NB-IoT гарантированно будут обслуживаться, поскольку оставшаяся часть уже выделенного RB не может быть занята каким-либо потоком LTE. Независимо от количественных значений трафика LTE система DYNRES всегда находится в выигрышном положении.

6.4.3 Марковская модель

Далее проведен анализ производительности технологии NB-IoT с точки зрения вероятностных характеристик обслуживания, таких как вероятность потери сообщения, распределение и средняя задержка сообщения в системе. Представленная модель позволяет определить диапазоны нагрузок при которых выполняются заданные параметры качества обслуживания устройств.

Предположим, что зона обслуживания базовой станции имеет форму окружности с радиусом R с равномерно распределёнными в ней N устройствами. Плотность датчиков на 1 км^2 определяется формулой $\sigma = N/\pi R^2$. Интенсивность поступления заявки от одного устройства будем обозначать λ_t . Тогда $\lambda_I = \lambda_t \Delta$ интенсивность поступления сообщений за время одного слота, где Δ – длительность одного кадра LTE. Датчики отправляют сообщение размером 100 бит. Таким образом, для передачи любого сообщения требуется один ресурсный блок NB-IoT.

Рассматривается сценарий, представленный на рисунке 6.6. Предполагается, что датчик NB-IoT передает данные на базовую станцию. Запрос буферизируется пока не освободится канал для передачи данных. Как только появляется возможность для передачи данных, базовая станция выделяет один канал NB-IoT датчику и тот начинает передачу данных на установленной мощности передачи. Требуемая мощность передачи $P(d)$ зависит от расстояния от датчика до базовой станции и определяется формулой $P(d) = Ad^\gamma$, где A является константой, зависящей от частоты, а γ – экспонента затухания.

Пусть A обозначает количество сообщений, поступившие в течение одного слота. Тогда вероятность, что поступит ровно k сообщений определяется формулой (6.12).

$$A_k = \frac{(\lambda\Delta)^k}{k!} e^{-\lambda\Delta}, k = 0, 1, \dots \quad (6.12)$$

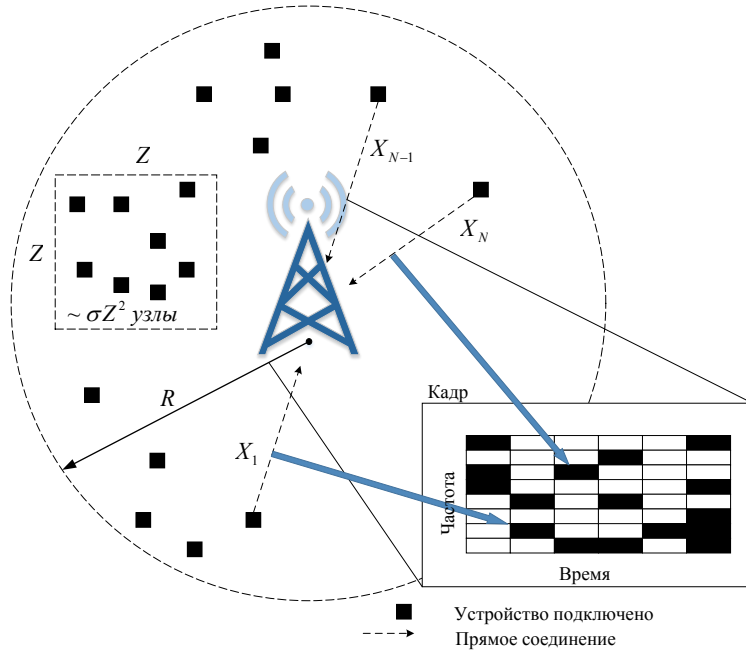


Рисунок 6.6 — Модель системы NB-IoT

В данном разделе введена вложенная цепь Маркова по моментам завершения обслуживания. Эволюция системы определяется уравнением (6.17).

$$Q(n) = \max(\min(Q(n-1) + A(n) - S(n-1), L), L), \quad (6.13)$$

где S обозначает количество обслуженных сообщений в слоте, при этом заявки обслуживаются пачками из не более чем K штук в момент окончания слота. Далее примем, что пришедшие заявки не могут моментально поступить на обслуживание в слоте, в котором они пришли, а будут обслужены только в следующем, а в случае, если в систему поступает пачка из i сообщений и при этом в системе уже находится k сообщений, так, что $k + i > L$, в очередь принимаются только $L - k$ сообщений, а оставшиеся теряются.

Обозначим $C = (c)_{i,j}, j \in \{0, 1, \dots, L\}$ матрицу переходных вероятностей для данной системы.

Пусть далее $\vec{x}_D = (x_{D,0}, x_{D,1}, \dots, x_{D,L})$, $x_{D,i} = \lim_{n \rightarrow \infty} Pr\{Q(n) = i\}$ – вектор стационарных вероятностей системы, описывающий количество сообщений в системе в момент времени непосредственно после окончания обслуживания. Вектор стационарных вероятностей может быть найден путем решения системы уравнений $\vec{x}_D C = \vec{x}_D$, $\vec{x}_D \vec{e} = 1$, где \vec{e} – единичный вектор.

В тексте раздела для описанной системы анализируются распределение вероятностей потери i сообщений, среднее значение времени задержки сообщения в системе.

Показано, что количество потерянных сообщений в слоте будет вычисляться по формуле (6.14).

$$f_L(i) = \frac{\sum_{k=0}^B x_{D,k} A_{K-k+i}}{Pr\{A(n) \geq 1\}} = \frac{\sum_{k=0}^B x_{D,k} A_{K-k+i}}{1 - e^{-\lambda\Delta}}, i = 1, 2, \dots, \quad (6.14)$$

где $Pr\{A(n) \geq 1\} = 1 - e^{-\lambda\Delta}$ – вероятность того, что в слот n поступает хотя бы одно сообщение.

Далее показано, что вероятность потери сообщений принимает вид (6.15)

$$p_L = \frac{\sum_{k=0}^B \sum_{i=1}^{\infty} x_{D,k} A_{L-k+i} \frac{i}{L-k+i}}{1 - e^{-\lambda\Delta}}. \quad (6.15)$$

6.4.4 Результаты и их анализ

Результаты численного анализа приведены на рисунке 6.7 и рисунке 6.8 для случая сот радиусом 10 и 30 км. В результате анализа показано, что технология NB-IoT может обеспечить надежную передачу данных (вероятность потери сообщений менее 10^{-3}) для 100 подключенных датчиков на квадратный километр, каждый из которых генерирует приблизительно $\lambda_I = 1$ сообщений в минуту. Как для указанных значений так и для меньшего количества подключенных IoT-устройств надежная связь может быть обеспечена.

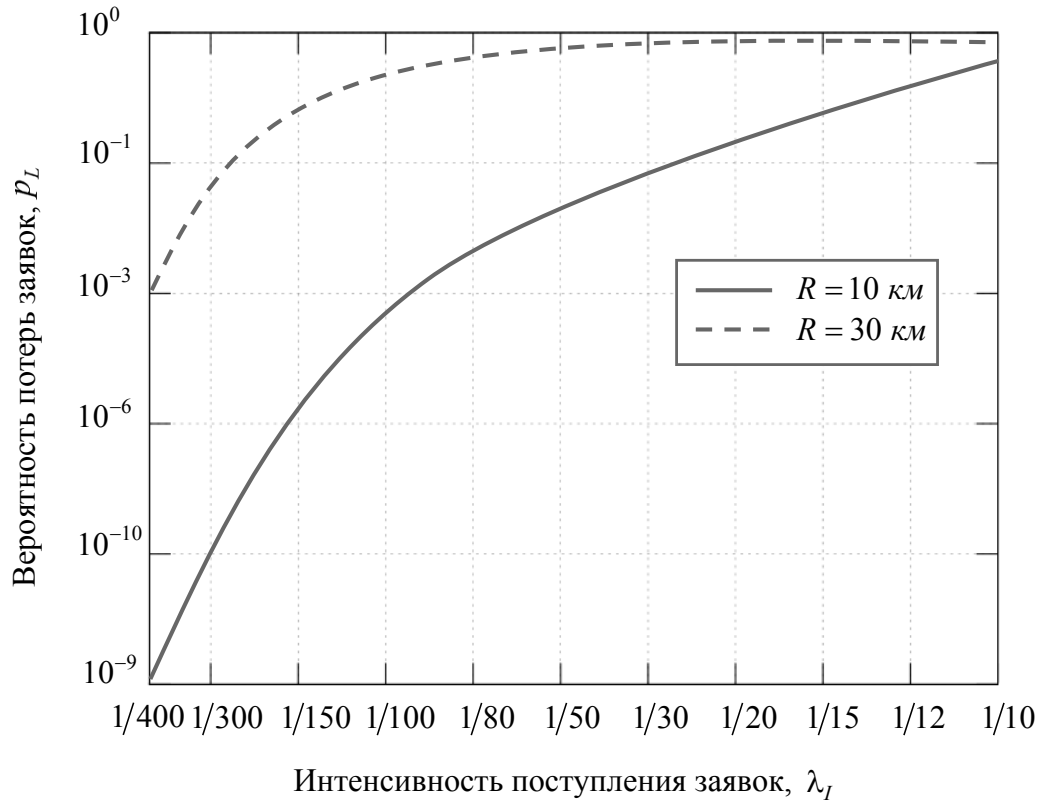


Рисунок 6.7 — Вероятность потери сообщения

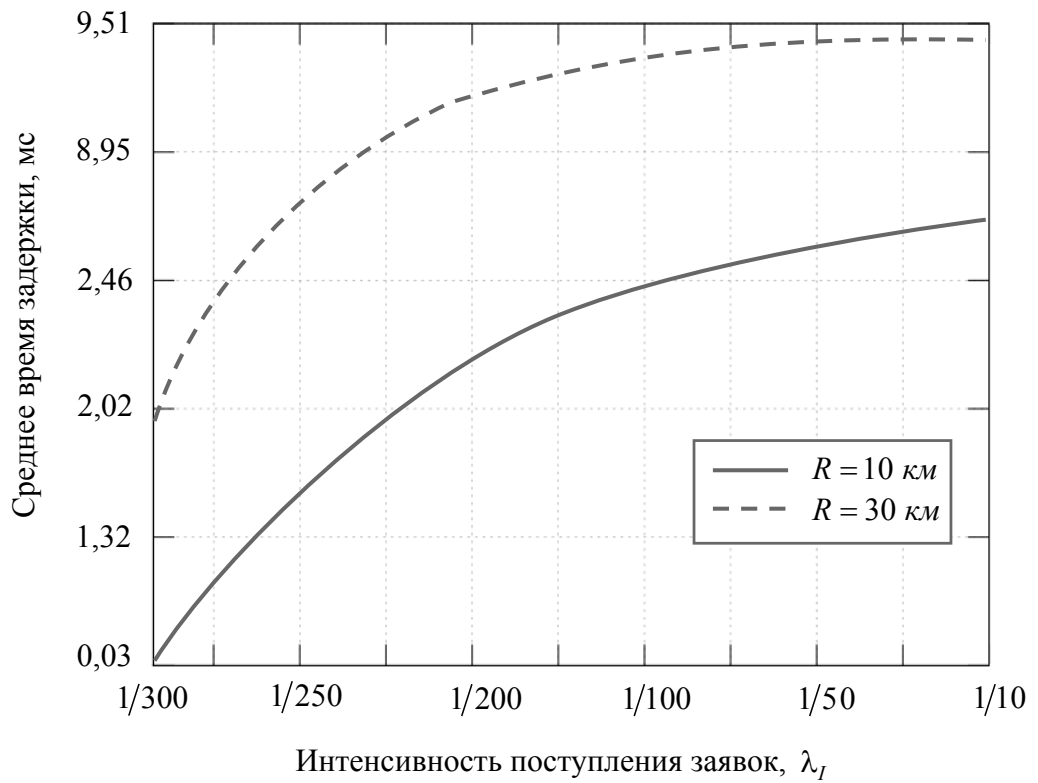


Рисунок 6.8 — Среднее время задержки

В разделе представлены методы эффективного сбора разнородных данных беспроводной сетью Интернета вещей. Основное внимание сосредоточено на передаче высокоскоростных видеопотоков, поступающих от камер наблюдения, подключенных к LTE, и низкоскоростных сенсорных данных, собираемых с многочисленных устройств, подключенных к технологии NB-IoT (с учетом внутриволнового развертывания NB-IoT) [300]. Разработана аналитическая модель оценки эффективности распределения радиоресурсов. Модель позволяет исследовать особенности совместного использования ресурсов LTE и NB-IoT. Наконец, на примере оператора наблюдения, реализующего систему мониторинга, проведена оценка трех альтернативных стратегий совместного использования ресурсов соты LTE с поддержкой технологии NB-IoT.

Проведенное численное исследование показало, что простейшая стратегия STAT очень чувствительна к предлагаемой нагрузке и требует точной информации об уровне поступающей нагрузки. Также показано, что стратегия DYN с полностью динамическим распределением ресурсов позволяет достичь наивысшей пропускной способности системы, но не может гарантировать требуемую надежность обслуживания. Предложенная стратегия динамического распределения ресурсов с резервированием, DYNRES, удовлетворяет требованиям надежности при сохранении коэффициента использования ресурсов на высоком уровне. Стратегия DYNRES рекомендуется для будущих развертываний IoT в мобильных сетях 5G, в то время как разработанная аналитическая модель может применяться для оценки эффективности будущих решений.

6.5 Узкополосная технология NB-IoT с ретрансляцией

6.5.1 Модель системы

Далее сосредоточимся на новой технологии NB-IoT и исследуем ее применимость для приложений краудсорсинга с помощью ретрансляторов, установленных на транспортных средствах [301]. Хотя в текущем релизе NB-IoT поддержка отслеживания и мобильности ограничена, но мы исследуем, есть ли перспективная выгода при включении этих функций в дополнительные версии спецификаций NB-IoT. Предполагается автономный режим работы NB-IoT, где его производительность не зависит от сотовой сети LTE.

Для достижения вышеуказанной цели ниже в данном разделе разрабатывается новая аналитическая модель, которая учитывает особенности краудсорсинга на основе ретрансляторов, мобильность вспомогательных транспортных средств и важные особенности новой технологии NB-IoT. Несмотря на то, что в NB-IoT уже имеет некоторое освещение в литературе [302; 303], новизна предлагаемого подхода заключается в объединении всех важных аспектов.

Далее в данном разделе представлены следующие исследования:

- ***Аналитическая модель NB-IoT***: несмотря на то, что в процессе стандартизации уже были зафиксированы некоторые базовые свойства работы NB-IoT, данное исследование сосредоточено на ориентированных на устройство показателях производительности, таких как вероятность потери сообщений, задержка, энергоэффективность и срок службы датчика [304]. Новые предложенные аналитические выражения для этих основных метрик получены с использованием инструментов теории массового обслуживания.
- ***Усовершенствованная конструкция системы NB-IoT с вспомогательными подвижными объектами***: отмечая значительное проникновение сотовых технологий в автомобильную

промышленность, ниже представлена архитектура системы, в которой пользовательские транспортные средства не только получают своевременные обновления от широко развернутых датчиков, но также помогают им передавать свои данные на принимающую базовую станцию. Поскольку подключенные транспортные средства могут быть оснащены более совершенными радиомодулями и не столь ограничены с точки зрения потребления энергии, предлагаемая модель может потенциально привести к значительным улучшениям в производительности.

- **Сравнение базового решения NB-IoT и улучшенного, основанного на применении ретранслятора:** ниже предложена аналитическая модель, которая учитывает использование транспортных средств в качестве ретрансляторов для доставки данных IoT. Численные результаты подтверждают, что массовое применение принадлежащих пользователю транспортных средств для приложений краудсорсинга позволяет существенно улучшить работу устройств IoT.

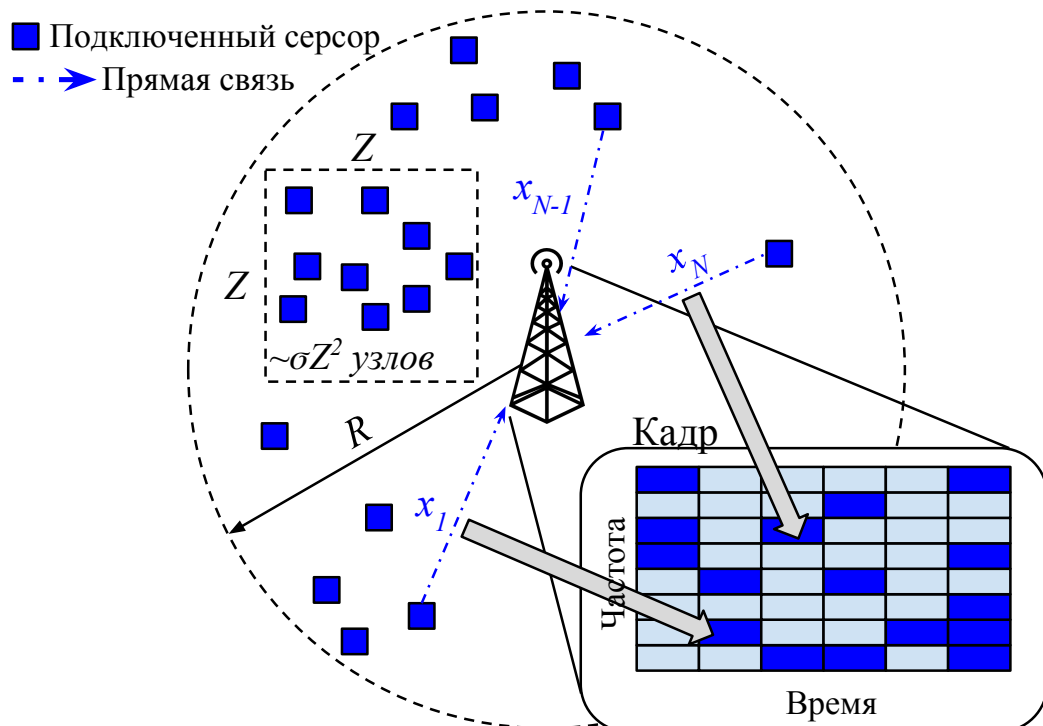


Рисунок 6.9 — Базовая NB-IoT система

В данной работе рассматривается зона обслуживания одной базовой станции (БС), имеющая форму окружности с фиксированным радиусом R . Предполагается, что в этой зоне равномерно распределены N датчиков. Плотность распределения датчиков, $\sigma = N/\pi R^2$, достаточно высока. Интервалы между временем поступления сообщений от датчиками имеют общее распределение с интенсивностью λ_t в s^{-1} (Гц). Тогда $\lambda_I = \lambda_t \Delta$ — интенсивность поступления сообщений за время одного слота, где Δ — длительность одного кадра. Датчики отправляют информацию о своем состоянии в коротком сообщении (100 бит). Таким образом, для передачи любого сообщения требуется один ресурсный блок NB-IoT (РБ, далее называемый виртуальным каналом NB-IoT), независимо от местоположения датчика в зоне обслуживания и выбранной MCS.

Мы рассмотрим два различных сценария, показанных на рисунке 4.8. Предполагается, что датчик, расположенный в зоне обслуживания, в случайный момент времени генерирует данные и, следуя стандартной процедуре случайного доступа, определенной в недавней спецификации 3GPP Rel. 13 [299], уведомляет базовую станцию о необходимости передачи данных. Планировщик сети принимает запрос и направляет его в очередь ожидания до того момента, пока не освободится канал для передачи данных. В это время, до непосредственной передачи данных, устройство переходит в

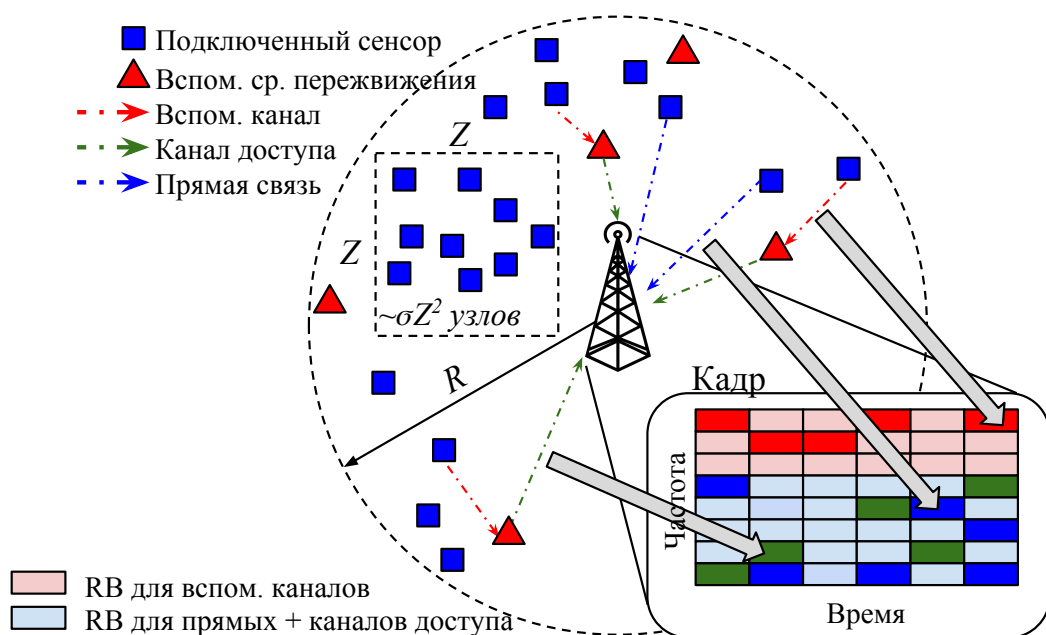


Рисунок 6.10 — Улучшенная система с ретрансляцией

режим ожидания передачи (Rx-ready). Введенная очередь предоставляет возможность контролировать максимальные задержки по передаче данных, т.к. сенсоры могут быть крайне чувствительны к задержкам, особенно когда необходимо передать важную информацию. Также это позволяет контролировать расход энергии сенсором, который все время ожидания до начала передачи данных находится в режиме прослушивания канала. В случае отсутствия мест в буфере, запрос на передачу сообщения считается потерянным. Как только появляется возможность для передачи данных, базовая станция выделяет один канал NB-IoT сенсором и тот начинает передачу данных на установленной мощности передачи. Требуемая мощность передачи $P(d)$ зависит от расстояния от датчика до базовой станции и определяется формулой $P(d) = Ad^\gamma$, где A является константой, зависящей от частоты, а γ — экспонента затухания.

Еще одна причина для использования виртуальной очереди ограниченного размера — это возможность контролировать расход энергии на датчики, поскольку они должны находиться в состоянии готовности Rx в течение всего времени ожидания. В случае отсутствия мест в буфере, запрос считается утерянным. Как только появляется возможность для передачи данных, БС выделяет для передачи данных РБ NB-IoT, и датчик отправляет свое сообщение на необходимом для достижения БС уровне мощности передачи. Требуемая мощность передачи, $P(d)$, зависит от расстояния между датчиками и БС. Предполагаем, что уровень мощности является инверсией модели распространения канала, т.е. $P(d) = Ad^\gamma$, где A — константа, а γ — экспонента затухания.

В дополнение к базовому сценарию рассмотрим расширенную модель на основе внедрения подвижных объектов процедуру обслуживания. Соответственно, датчики могут использовать подвижные объекты, которые перемещаются вокруг области в качестве узлов ретрансляции, чтобы доставлять информацию БС. Преимущество использования подвижных объектов в том, что они могут быть расположены гораздо ближе к датчикам, что позволяет снизить мощность передачи и сэкономить ресурс аккумулятора. Мы предполагаем, что в зоне обслуживания доступны M подвижных

объектов, которые будут служить узлами передачи данных от датчиков, назовем их ретрансляционными БС.

Чтобы зафиксировать эффект мобильности, моделируем ее, используя модель случайного направления (RDM, [305]) с постоянной скоростью v и средним временем обслуживания $E[\tau]$. В этом случае ретрансляционные БС функционируют как основные БС и реализуют виртуальную очередь, как описано выше. Чтобы уменьшить помехи между (i) датчиком и ретрансляционной БС и (ii) передачами от ретрансляционной БС к основной БС, K_1 каналов присвоены для основного БС, а K_2 каналов доступны для ретрансляционного БС. При генерации нового сообщения датчик сканирует беспроводные каналы для БС. Выбор обслуживающей БС производится на основе самого сильного уровня принимаемой мощности. Поскольку подвижные объекты являются мобильными, датчик не хранит информацию о прошлых передачах, а выбор предпочтительного адресата выполняется заново каждый раз, когда необходимо передать новое сообщение.

Мы сравниваем производительность базового и расширенного сценария на, рассматривая четыре интересующие метрики [306]: (i) вероятность того, что сообщение будет удалено из очереди (т. е. вероятность потери сообщений), (ii) распределение количества энергии, необходимой для доставки сообщений датчиком, (iii) распределение времени жизни датчика, (iv) распределение задержки сообщения в системе. Обозначения, используемые в работе, приведены в Таблице 6.2.

6.6 Аналитическая модель

В этом разделе мы разрабатываем математическую модель для возможности выполнения процедуры crowdsensing (краудсорсинг данных датчиков с мобильных устройств) над NB-IoT. Начнем с исследования базового сценария, где в процессе передачи сообщений участвует только статическая сетевая инфраструктура. Затем мы расширим построенный сценарий на случай применения подвижных объектов.

Обозначения, использованные в этой работе.

Параметр	Описание
λ_t	Интенсивность поступления сообщений от датчика, Гц
λ_I	Интенсивность поступления сообщений от датчика, слот ⁻¹
σ	Средняя плотность распределения датчиков в зоне обслуживания
T_I	Время генерирования входящего сообщения
$P(d)$	Мощность передачи, зависящая от расстояния
K	Общее число каналов NB-IoT
L	Количество позиций ожидания на БС
K_1	NB-IoT каналы, выделенные для БС
K_2	NB-IoT каналы, выделенные для ретрансляционных БС
γ_K	Доля каналов, выделенных подвижным объектам
L_1, L_2	Число позиций ожидания на БС и ретрансляционных БС
v	Скорость подвижных объектов
$E[\tau]$	Среднее время работы ретрансляционной БС в модели RDM
$Q(n)$	Число сообщений после завершения обслуживания
$A(n)$	Число поступлений в слот n
$S(n)$	Количество завершенных обслуживаний в слоте n
$C = [c]_{ij}$	Матрица переходных вероятностей цепи Маркова
$\vec{x}_D = [x]_k$	Вероятность наличия k сообщений в системе
$p_{L,i}$	Вероятность потери сообщения (модель ретрансляционных БС)
q_B, q_R	Время обслуживания датчика БС/ретрансляционной БС
p_i	Вероятность попадания сообщения в очередь i
Δ	Длительность кадра
λ_i	Интенсивность поступления сообщения в очередь i
$B(n)$	Число потерянных сообщений за слот n
W	Время ожидания в очереди
P	Позиция обслуживаемого сообщения
T_S, T_R, T_M	Продолжительность периодов времени поиска, доступа и передачи
T_A, T_S	Время доступа и обслуживания
P_{Rx}, P_{Tx}, P_I	Мощность получения, отправки и ожидания, Дж/с
E_T, E_S	Энергия за транзакцию и интервалы спящего режима, Дж
U	Мощность батареи, Дж
D	Расстояние между двумя случайными точками в окружности
D_{\min}	Минимальное расстояние до N равномерно распределенных точек в окружности
$f_X(x)$	ПРВ с.в. X
$F_X(x)$	ФРВ (функция распределения вероятностей) с.в. X

6.6.1 Базовая модель системы

Для анализа базовой модели представим процесс обслуживания передаваемого трафика от датчиков как сеть очередей. Параметризуя и решая эту постановку, приходим к распределению времени задержки сообщений. Кроме того, используя это распределение времени задержки, получаем распределения энергии, необходимое для доставки сообщений и распределение времени жизни датчика.

Напомним, что интенсивность поступления сообщения равна λ_I , а плотность распределения датчиков σ относительно высока. Также принимая во внимание, что согласно спецификации 3GPP Rel. 13, радиус R может достигать 30 км, мы можем применить теорему суперпозиции Палмы, в которой утверждается, что после наложения многих независимых точечных процессов результирующий агрегированный процесс будет локально пуассоновским [307]. Заметим, что этот важный результат выполняется для любого времени поступления сообщений (включая постоянные интервалы между сообщениями) от отдельных датчиков до тех пор, пока их плотность будет достаточно высокой. Следовательно, суммарный процесс поступления сообщений – это процесс Пуассона с интенсивностью $\lambda = N\lambda_I$.

Рассмотрим теперь обслуживание сообщений БС. Учитывая, что один датчик генерирует сообщения с интенсивностью λ_I , а также тот факт, что количество сенсоров в зоне обслуживания достаточно велико, суммарный поток поступления сообщений от всех датчиков зоне обслуживания соты является пуассоновским с интенсивностью $\lambda = N\lambda_I$. В рамках LTE Rel. 13 предполагаются изменения в беспроводных сетях, обслуживающих NB-IoT устройства – специально выделенная радиочастотная полоса NB-IoT в 180 кГц может быть разделена на 12 диапазонов по 15кГц, в каждом из которых имеется 6 каналов для передачи данных. Следовательно, под NB-IoT выделяется $K = 6 \times 12 = 72$ виртуальных каналов в одном слоте. Таким образом, в течении каждого слота обрабатывается до K сообщений. Таким образом, моделируется система массового обслуживания в дискретном времени с групповым поступлением и обслуживанием сообщений.

Время обслуживания каждой заявки постоянна. Поскольку исходным процессом поступления является пуассоновским с интенсивностью λ , то число сообщений, поступающих на каждый интервал, также имеет распределение Пуассона с параметром λ .

Системное время имеет временной интервал с длиной кадра Δ , определяемой продолжительностью одного слота. Согласно нашей модели, сообщения поступают пачками, а пачки сообщений поступают перед окончанием слота. Обслуживаются заявки пачками в момент окончания слота. Состояние системы рассматривается в момент времени сразу после окончания обслуживания одной пачки. При этом вновь пришедшие заявки не могут моментально поступить на обслуживание в слоте, в котором они пришли, а будут обслужены только в следующем слоте.

Легко видеть, что точки наблюдения образуют вложенную одномерную цепь Маркова, $\{Q(n), n = 0, 1, \dots\}$, описывающую количество сообщений в системе сразу после ухода. Эта система известна как модель запаздывающего поступления с задержкой [308; 309]. Время ожидания и время пребывания определяется как количество слотов, потраченных на сообщение в системе. Заметим, что время ожидания в этой модели обеспечивает плотную верхнюю границу фактического времени ожидания датчика. Мы рассматриваем частичную стратегию принятия и обслуживания пакета [310; 311]. В соответствии с описанной моделью, продолжительность одного такта равна Δ , и определяется продолжительностью одного слота LTE. Сообщения поступают в систему в момент времени непосредственно перед окончанием слота. Обслуживаются заявки пачками из не более чем K штук в момент окончания слота. Состояние системы рассматривается в момент времени непосредственно после окончания обслуживания очередной пачки. При этом вновь пришедшие заявки не могут моментально поступить на обслуживание в слоте, в котором они пришли, а будут обслужены только в следующем, а в случае, если в систему поступает пачка из i сообщений и при этом в системе уже находится k сообщений, так, что $k + i > L$, в очередь приримаются только $L - k$ сообщений, а оставшиеся теряются. то есть всякий раз, когда приходит партия из i сообщений и в

системе еще есть k сообщения, что $k + i > L$, в очередь принимаются $L - k$ сообщений, а остальные отклоняются.

Пусть A — с.в., обозначающая число сообщений, поступивших в течение одного слота. Тогда, т.к. процесс поступления является пуассоновским, вероятность A_k поступления ровно k сообщений за один слот вычисляется как

$$A_k = \frac{(\lambda\Delta)^k}{k!} e^{-\lambda\Delta}, k = 0, 1, \dots \quad (6.16)$$

Введем вложенную цепь Маркова по моментам завершения обслуживания. Эволюция системы определяется уравнением (6.17).

$$Q(n) = \max(\min(Q(n-1) + A(n) - S(n-1), L), L), \quad (6.17)$$

где S обозначает количество обслуженных сообщений в слоте.

Пусть $C = (c)_{i,j}, j \in \{0, 1, \dots, L\}$ — матрица переходных вероятностей для данной системы. Вероятности переходов определяются следующим образом:

$$\left\{ \begin{array}{l} c_{0,0} = A_0, \\ c_{0,j} = A_j, j = 1, 2, \dots, L-1, \\ c_{0,L-K} = \sum_{k=L}^{\infty} A_k, \\ c_{i,0} = A_0, i \leq K, \\ c_{i,0} = 0, i > K, \\ c_{i,L-K} = \sum_{k=L-i}^{\infty} A_k, i > K, \\ c_{i,j} = A_j, i \leq K, j = 1, \dots, L-i-1, \\ c_{i,j} = A_{j-i+K+1}, j = K+1, \dots, L-1, j = i-K, \dots, L-K-1, \\ c_{i,j} = 0, j = K+1, \dots, L, j = L-K+1, \dots, L, \\ c_{i,j} = 0, j = K+1, \dots, L, j = 0, \dots, i-K-1. \end{array} \right. \quad (6.18)$$

Пусть далее $\vec{x}_D = (x_{D,0}, x_{D,1}, \dots, x_{D,L}), x_{D,i} = \lim_{n \rightarrow \infty} Pr\{Q(n) = i\}$ — вектор стационарных вероятностей системы, описывающий количество со-

общений в системе в момент времени непосредственно после окончания обслуживания. Вектор стационарных вероятностей может быть найден путем решения системы уравнений $\vec{x}_D C = \vec{x}_D, \vec{x}_D \vec{e} = 1$, где \vec{e} – единичный вектор.

Для описанной системы анализируются распределение вероятностей потери i сообщений, среднее значение времени задержки сообщения в системе. Предельная вероятность равна $x_{D,i} = \lim_{n \rightarrow \infty} Pr\{Q(n) = i\}$.

Вероятность потери заявки

Пусть с.в. $B, B \in \{0, 1, \dots\}$, обозначает число потерянных сообщений за один слот и пусть $f_B(i) = Pr\{B(n) = i | A(n) \geq 1\}$, $i = 0, 1, \dots$, ее функция распределения при условии, что по крайней мере одно сообщение поступит за время одного слота.

Рассмотрим случай потери ровно $i, i = 1, 2, \dots$ сообщений в n -ый слот. Это происходит, когда одновременно в $(n - 1)$ -ом слоте есть ровно $k, k = 0, 1, \dots, L$ сообщений и число поступающих сообщений в n -ый слот равен $L - k + i$. Мы находим ряд распределения количества потерянных сообщений в слоте

$$f_B(i) = \frac{\sum_{k=0}^L x_{D,k} A_{L-k+i}}{Pr\{A \geq 1\}}, i = 1, 2, \dots, \quad (6.19)$$

где $x_{D,k}, k = 0, 1, \dots, L$ – вероятность того, что в системе есть ровно k сообщений непосредственно перед поступлением пачки. Для системы позднего прибытия с задержкой доступа эти моменты времени совпадают с распределением \vec{x}_D .

Вероятность поступления хотя бы одно сообщения за слот равна

$$Pr\{A \geq 1\} = \sum_{k=1}^{\infty} \frac{(\lambda \Delta)^k}{k!} e^{-\lambda \Delta} = 1 - e^{-\lambda \Delta}. \quad (6.20)$$

Заметим, что важный параметр, описывающий случай отсутствия потери сообщения в произвольном слоте, $f_B(0)$, можно получить как $1 - \sum_{i=1}^{infy} f_B(i)$. Используя ряд распределения $f_B(i)$, вероятность потери сообщения примет вид

$$p_L = \frac{\sum_{i=1}^{\infty} \sum_{k=0}^L x_{D,k} A_{L-k+i} \frac{i}{L-k+i}}{1 - e^{-\lambda\Delta}}, \quad (6.21)$$

где $i/(L-k+i)$ – это вероятность того, что случайно выбранное сообщение, поступившее в слот n , будет потеряно.

$$\begin{aligned} f_W(i) = & \frac{1}{\xi_A} \left(\sum_{j=0}^{\min(K, L-(K(i-1)+1))} x_{D,j} \right. \\ & \sum_{n=K(i-1)+1}^{\infty} \frac{\min(n+1 - (K(i-1)+1), K, L-j)}{n} A_n + \\ & + \sum_{j=K+1}^{\min(Ki, L-1)} x_{D,j} \sum_{n=Ki-j+1}^{\infty} \\ & \frac{\min(n+1 - (Ki-j+1), K, L-j)}{n} A_n + \\ & \sum_{j=Ki+1}^{\min(K(i+1)-1, L-1)} x_{D,j} \sum_{n=1}^{\infty} \\ & \left. \frac{\min(n, K(i+1) - j, L-j)}{n} A_n \right). \end{aligned} \quad (6.22)$$

Распределение времени ожидания

Рассмотрим произвольное сообщение, поступающий в слот n для обслуживания системой. Пусть W , $W \in \{0, 1, \dots, \lceil L/K \rceil\}$ – случайная величина, обозначающая задержку в очереди в системе. Тогда $f_W(i) =$

$Pr\{W(n) = i | A(n) > B(n)\}$ функция распределения W , описывающая случай прихода хотя бы одного сообщения в n -ый слот. Найдем распределение задержки следующим образом. Сначала определим распределение вероятностей приема выбранной заявки на обслуживание на место i в системе. Далее, учитывая пачечное обслуживание, определим количество слотов, проводимое выбранным сообщением в системе. Задержка ожидания в системе, связанная с поступлением сообщения в систему, показана на рисунке 6.11.

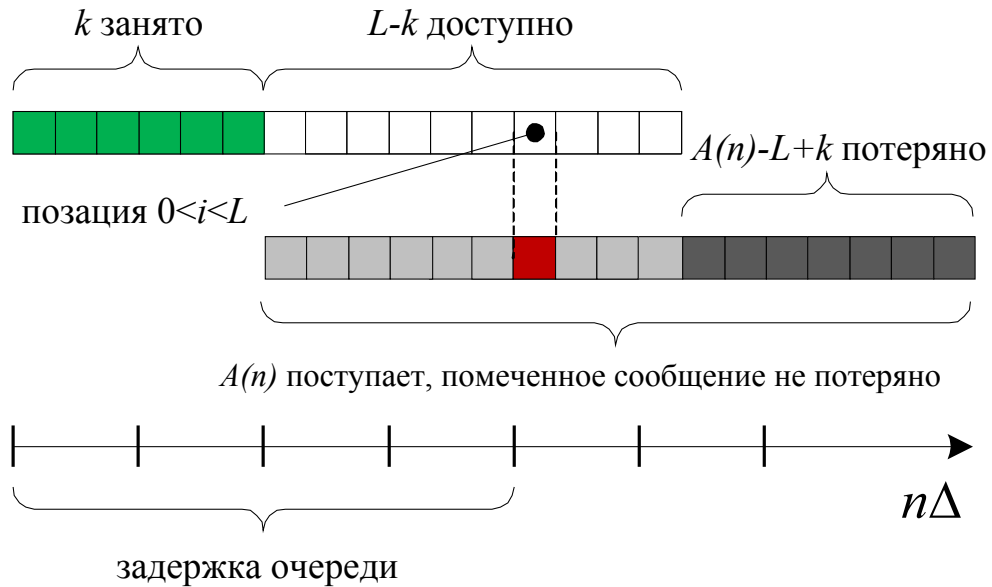


Рисунок 6.11 — Иллюстрация задержки ожидания в системе

Рассмотрим событие, когда сообщение проводит ровно один слот в системе. Это соответствует нулевому времени ожидания. Заметим, что это событие может произойти, когда (i) в системе не более K сообщений или (ii) в системе больше K сообщений, а рассматриваемое сообщение относится к числу первых из K сообщений в прибывшем пакете. Последнее является следствием системы запаздывающего поступления с немедленным доступом. В этих условиях мы имеем

$$f_W(0) = \frac{\sum_{i=0}^K x_{D,i} \left(A_1 + \frac{2A_2}{2} + \dots + \frac{KA_K}{K} + \frac{KA_{K+1}}{K+1} + \dots \right)}{\xi_A}, \quad (6.23)$$

где ξ_A — нормирующая константа, которая вычисляется как

$$\xi_A = \sum_{i=0}^{\lceil L/K \rceil - 1} f_W(i). \quad (6.24)$$

Аналогично, для задержки одного слота получаем

$$\begin{aligned} f_W(1) = & \frac{1}{\xi_A} \left(\sum_{i=0}^K x_{D,i} \times \right. \\ & \times \left(\frac{A_{K+1}}{K+1} + \frac{2A_{K+2}}{K+2} + \dots + \frac{KA_{2K}}{2K} + \frac{KA_{2K+1}}{2K+1} + \dots \right) + \\ & + x_{D,K+1} \left(\frac{A_K}{K} + \frac{2A_{K+1}}{K+1} + \dots + \frac{KA_{2K-1}}{2K-1} + \frac{KA_{2K}}{2K} + \dots \right) + \\ & + x_{D,K+2} \left(\frac{A_{K-1}}{K-1} + \frac{2A_K}{K} + \dots + \frac{KA_{2K-2}}{2K-2} + \frac{KA_{2K-1}}{2K-1} + \dots \right) + \\ & + \dots + \\ & + x_{D,2K} \left(A_1 + A_2 + \dots + A_K + \frac{KA_{K+1}}{K+1} + \dots \right) + \\ & + x_{D,2K+1} \left(A_1 + \dots + A_{K-1} + \frac{K-1}{K}A_K + \frac{K-1}{K+1}A_{K+1} + \dots \right) + \\ & + x_{D,2K+2} \left(A_1 + \dots + A_{K-2} + \frac{K-2}{K-1}A_{K-1} + \frac{K-2}{K}A_K + \dots \right) + \\ & + \dots + \\ & \left. + x_{D,3K-1} \left(A_1 + \frac{A_2}{2} + \frac{A_3}{3} + \dots \right) \right). \quad (6.25) \end{aligned}$$

основываясь на метод индукции получаем (6.22).

Предложенная модель обеспечивает приближение фактической задержки очередей. Для работы системы NB-IoT, когда потери сообщений происходят нечасто, задержки обычно относительно малы. Следовательно, наша построенная методология, где задержки округляются до ближайшего верхнего целого числа, может слегка переоценить реальную задержку в очереди.

Чтобы учесть этот факт, нам необходимо дополнительно рассмотреть два компонента, показанных на рисунке 6.12, где мы иллюстрируем интер-

валы планирования системы NB-IoT на соответствующим распределением диапазона. В дополнение к времени ожидания в очереди, мы также наблюдаем смещение обслуживания T_X — время, прошедшее с начала слота, где запланировано обслуживание сообщения, во время его передачи. Поскольку диапазон NB-IoT в каждом интервале планирования составляет всего 180 КГц, мы приближаем его к константе.

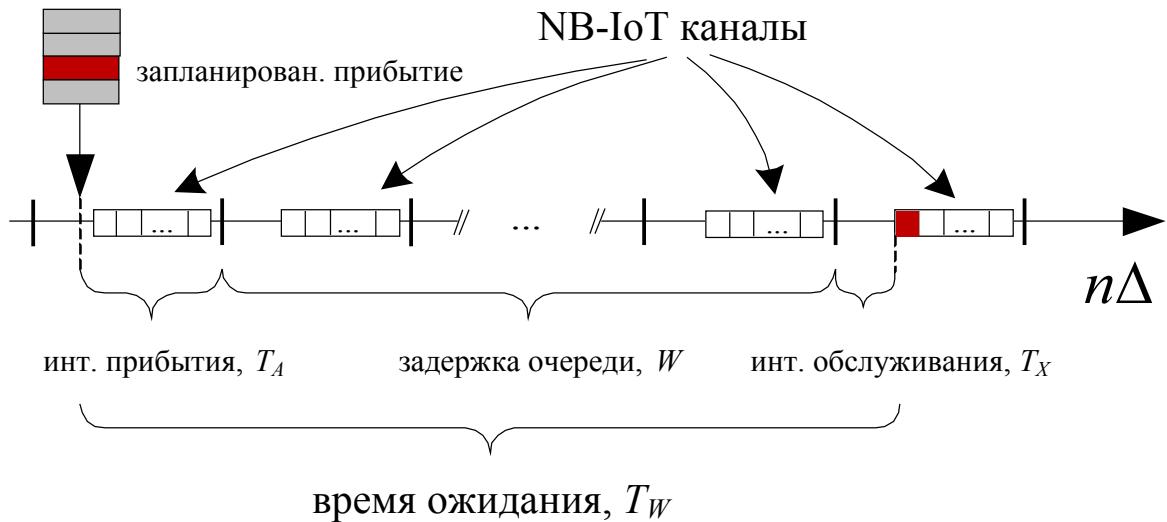


Рисунок 6.12 — Подробное описание времени ожидания для сообщения в системе

Таким образом, последним компонентом задержки является сдвиг поступления, T_A . Поскольку фактическое поступление сообщений является процессом Пуассона, этот сдвиг является непрерывным с.в. Напомним, что поступление сообщения из пуассоновского процесса, попадающего на отрезок постоянной длины, равномерно распределяется по этому интервалу, плотность распределения T_A равномерна на Δ , т.е.

$$f_{T_A}(x) = 1/\Delta, 0 < x < \Delta. \quad (6.13)$$

Поэтому фактическое время ожидания в системе обеспечивается путем перевода времени ожидания в системе очередности T_W в терминах T_S , а затем выполнением свертки с непрерывным треугольным распределением

T_A . В результате мы имеем

$$f_{T_W}(x) = \sum_{i=0}^{\infty} f_W(i + T_S) f_{T_A}(x). \quad (6.14)$$

Энергия на передачу

Результаты, полученные с помощью систем массового обслуживания, позволяют вычислить распределение энергии, необходимое для передачи сообщений. Рассмотрим рисунок 6.13, на котором показаны различные состояния сенсора, соответствующие уровни мощности и время пребывания в системе. Несмотря на то, что на практике у IoT устройства может быть больше состояний, мы определяем лишь три из них по расходу мощности: (i) состояние спящего режима, (ii) готовый к приёму (состояние Rx) и (iii) состояние Tx. Прежде чем характеризовать распределение энергии, затраченное на передачу, и распределение срока службы сенсора, необходимо определить продолжительность и уровни мощности этих состояний.

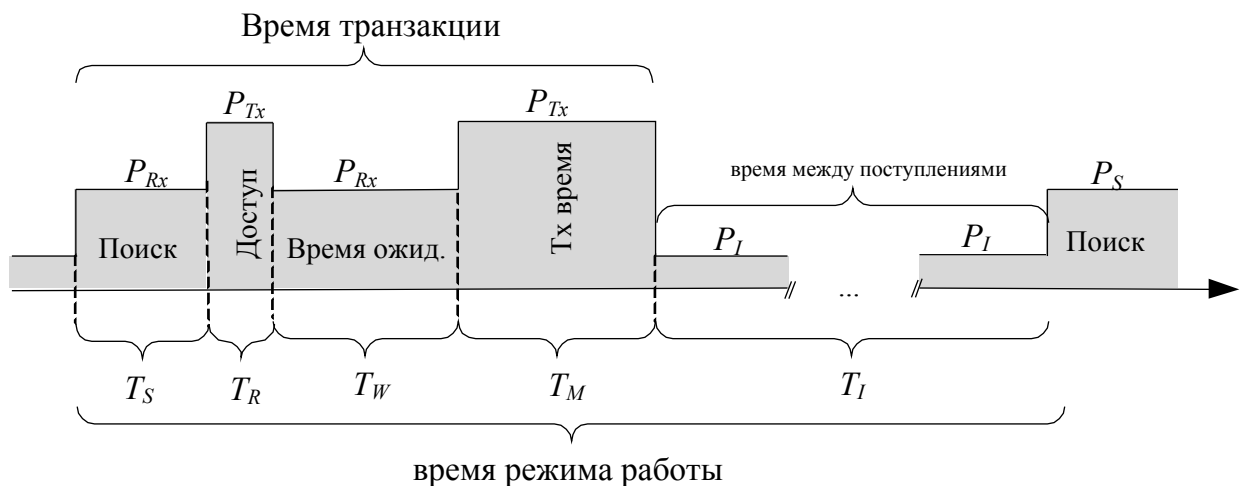


Рисунок 6.13 — Пример распределения уровней мощности для эксплуатируемого NB-IoT устройства

В дальнейшем, в соответствии с [312], мы предполагаем, что процедура произвольного доступа на общем канале хорошо подготовлена, то есть

нет никаких потерь или избыточных задержек. Действительно, процедура произвольного доступа, реализованная в NB-IoT, оставалась предметом обширных исследований последнее десятилетие и, таким образом, оптимизирована. Кроме того, мы считаем, что уровень мощности, необходимый для передачи сообщений, P_{Tx} , зависит от расстояния до БС, а мощность, требуемая для приема, P_{Rx} , постоянна. Обозначая продолжительности и уровни мощности T_* и P_* соответственно, принимаем следующие параметры:

- поиск БС: постоянные T_S и P_{Rx} ;
- произвольный доступ: постоянная T_R и случайная P_{Tx} ;
- время ожидания: случайная T_W и постоянная P_{Rx} ;
- передача сообщения: постоянная T_M и случайная P_{Rx} ;
- состояние сна: случайная T_I и постоянная P_I .

Можно заметить, энергия, необходимая для передачи сообщения (здесь: транзакция), представляет собой величину RV , состоящую из трех случайных составляющих и одного постоянного фактора. Чтобы охарактеризовать время жизни датчика, также необходимо учитывать случайное время сна, которое полностью определяется кумулятивной функцией распределения (CDF) общего времени послушания промежуточных сообщений, $F_{T_I}(x)$. Таким образом, энергия на транзакцию (в Джоулях) может быть определена как

$$E_T = \underbrace{P_{Rx}}_{Const.} \overbrace{(T_S + T_W)}^{RV} + \underbrace{P_{Tx}}_{RV} \overbrace{(T_R + T_M)}^{Const.}. \quad (6.15)$$

Плотность распределения вероятностей (ПРВ) $f_{T_S+T_W}(y)$ можно определить, используя линейное преобразование RV , т.е. если $Y = a + bX$ и $f_X(x)$ известны, то ПРВ Y , $f_Y(y)$, задается формулой

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{dx}{dy} \right| = f_X \left(\frac{y - a}{b} \right) \frac{1}{|b|}. \quad (6.16)$$

Объем мощности, требуемой в состоянии T_x , явно зависит от расстояния D до БС. Вспоминая предполагаемое обратное управление мощностью

в виде $P_{Tx}(D) = AD^\gamma$, где A — параметр, обеспечивающий определенную мощность Rx на БС, распределение P_{Tx} можно получить, применяя нелинейное преобразование RV. Заметим, что ПРВ RV Y , $f_Y(y)$, выражаемая монотонной функцией $y = \phi(x)$ другого RV X с ПРВ $f(x)$ [313]

$$f_Y(y) = f(\psi(y))|\psi'(y)|, \quad (6.17)$$

где $x = \psi(y) = \phi^{-1}(x)$ обратная функция.

ПРВ расстояния от точки, равномерно распределенной по окружности к центру этого круга равна $f_D(x) = 2x/R^2$, $0 < x < R$. Функция $P_{Tx}(D) = AD^\gamma$ монотонна на положительной части оси ox , а обратная ей функция равна $(A/y)^{1/\gamma}$. Тогда по модулю производной для обратного

$$|\psi'(y)| = \frac{(A/y)^{1/\gamma}}{\gamma y}. \quad (6.18)$$

Согласно (6.17), ПРВ P_{Tx} выражается как

$$f_{P_{Tx}}(y) = \frac{2(A/y)^{2/\gamma}}{\gamma R^2 y}, y > 0. \quad (6.19)$$

Согласно (6.15), преобразовывая $f_{T_S+T_W}(y)$ и $f_{P_{Tx}}(y)$ линейно с постоянными коэффициентами и затем свертывая полученные плотности, можно найти искомую ПРВ энергии на транзакцию, $f_{E_T}(x)$.

Время жизни датчика

Зная $f_{E_T}(x)$, можно продолжить определение распределения времени жизни для случайно выбранного датчика. Рассмотрим работу IoT устройства как процесс обновления, где цикл обновления представляет собой сумму случайного времени транзакции и времени ожидания. Количество энер-

гии, затрачиваемой в течение цикла обновления, зависит от продолжительности интервала. ПРВ объема энергии, потребляемой в спящем режиме, может быть получена путем линейного преобразования ПРВ времени сна, T_I , в виде $E_S = P_I T_I$, где P_I — постоянная, согласно (6.16). В ПРВ объема энергии, потребляемой в течение цикла обновления, выдается свертка $f_{E_S}(x)$ и $f_{E_T}(x)$.

Пусть емкость батареи равна U Джоулей, а $f_E(t, x)$ — ПРВ величины объема энергии, затрачиваемой за время t . Поскольку мы предполагаем, что $U \gg E[E_C]$, где E — математическое ожидание, имеет место центральная предельная теорема, гласящая, что при больших t распределение потребляемой энергии близко к нормальному. Пренебрегая временем возвращения в последнем цикле обновления, среднее и дисперсия задаются [314]

$$\lim_{t \rightarrow \infty} \mu_E(t) = \frac{tE[E_C]}{E[T_C]}, \quad \lim_{t \rightarrow \infty} \sigma_E^2[t] = \sigma^2[T_C] \frac{t}{(E[T_C])^3}. \quad (6.20)$$

Вероятность того, что время жизни датчика больше, чем t , равна

$$Pr\{T_L > t\} = Pr\{E(t) - C < 0\} = \int_{-\infty}^0 f_E(t, x + U) dx, \quad (6.21)$$

которую можно рассчитать численно.

6.6.2 Усовершенствованная система с ретрансляцией

Теперь рассмотрим настройку NB-IoT системы с M мобильными ретрансляторами, установленными на машинах, которые помогают доставлять сообщения датчиков. Согласно модели RDM, каждая подвижная БС сначала выбирает случайное направление α равномерно между 0 и 2π , а затем движется по прямой какое-то экспоненциально распределенное время τ с постоянной скоростью v . Наблюдаемая система находится в равнове-

сии. Известно, что в устойчивом состоянии положение подвижной БС в пределах круговой области радиуса R равномерно распределяется по πR^2 , независимо от исходного распределения подвижной БС, также как и эффекты границы [305].

Моделирование процесса доставки сообщений предлагаем проводить с использованием сети массового обслуживания с дискретным распределением временем. Сеть изображена на рисунке 6.14 и состоит из $M + 1$ систем массового обслуживания (СМО). На входе существуют M СМО, представляющие процесс обслуживания на подвижных БС. Коэффициенты p_i , $i = 1, 2, \dots, M$ отражают общую интенсивность входящего потока сообщений λ , обслуживаемую соответствующими подвижными БС. Затем выходящие отсюда потоки суммируются с потоками, обслужившимися непосредственно на БС, $p_{M+1}\lambda$. В конце суммарный поток попадает в $M + 1$ -ю СМО для обслуживания. Поскольку данная сеть не имеет никаких циклов, можем проанализировать ее применяя метод декомпозиции, т.е. рассматривая каждую очередь отдельно.

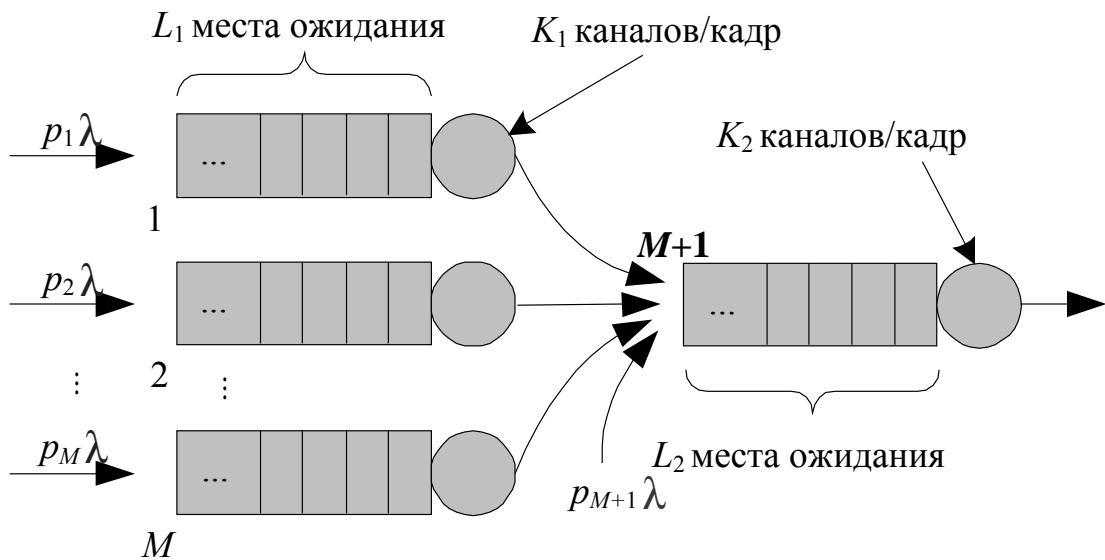


Рисунок 6.14 — Иллюстрация рассматриваемой СеМО

Задержки и потери сообщений

Опираясь на результаты предыдущего подраздела, параметризуем интенсивности очередей с помощью p_i , $i = 1, 2, \dots, M + 1$. Рассматривая эволюцию процесса обслуживания на подвижных БС и помня, что равновесное распределение их местоположений равно $f(x, y) = 1/\pi R^2$, видим, что все подвижные БС неразличимы друг от друга в том смысле, что доля трафика, которую они обслуживают, остается неизменной, то есть $p_i = p_j = p$, $\forall i, j = 1, 2, \dots, M$. Чтобы определить p и p_{M+1} , рассмотрим M равномерно распределенных точек в круге радиуса R , а новая точка P , представляющая поступление сообщения, равномерно падает в круг.

Напомним, что наша модель распространения является монотонно убывающей функцией расстояния. Следовательно, для захвата p и p_{M+1} достаточно определить, будет ли P ближе к центру круга или к любой подвижной БС. ПРВ расстояния между двумя точками, равномерно распределенными в круге радиуса R , равна [315]

$$f_D(x, R) = \frac{2x}{R^2} \left(\frac{2}{\pi} \cos^{-1} \left(\frac{x}{2R} \right) - \frac{x}{\pi R} \sqrt{1 - \frac{x^2}{4R^2}} \right), \quad (6.22)$$

где $0 < x < 2R$.

Известно, что КФР минимального M независимого и тождественно распределенного RVs, $Y = \min(X_1, X_2, \dots, X_M)$, задается следующим образом [313]

$$\begin{aligned} F_Y(y) &= 1 - Pr\{(X_1 > y), \dots, (X_M > y)\} = \\ &= 1 - \prod_{i=1}^M Pr\{X_i > y\} = 1 - \prod_{i=1}^M [1 - F_i(y)], \end{aligned} \quad (6.23)$$

где $F_i(y) = Pr\{X_i \leq y\}$ — КФР X_i .

Интегрируя (6.22), определяем КФР как

$$F_D(x, R) = \frac{2}{\pi} \csc^{-1} \left(\frac{2R}{x} \right) - \frac{2R^2x + x^3}{4\pi R^3} \sqrt{4 - \frac{x^2}{R^2}} + 2Rx^2 \sec^{-1} \left(\frac{2R}{x} \right), 0 < x < 2R. \quad (6.24)$$

Из (6.23) и (6.24), мы приходим к КФР расстояния до ближайшей подвижной БС в виде

$$F_{D_{\min}}(x, R) = 1 - \left(1 - \frac{2}{\pi} \csc^{-1} \left(\frac{2R}{x} \right) - \frac{2R^2x + x^3}{4\pi R^3} \sqrt{4 - \frac{x^2}{R^2}} + 2Rx^2 \sec^{-1} \left(\frac{2R}{x} \right) \right)^M, 0 < x < 2R. \quad (6.25)$$

Пусть D_C — RV, обозначающее расстояние от равномерно распределенной точки внутри круга до центра этой окружности. Известно, что ПРВ D_C это [315]

$$F_{D_C}(x) = \frac{2x}{R^2}, 0 < x < R. \quad (6.26)$$

Вероятность того, что по прибытии сообщения к случайно выбранному датчику БС будет ближе к этому датчику, чем к любой из подвижных БС, задается формулой

$$p_{M+1} = Pr\{D_{\min} > D_C\} = Pr\{D_{\min} - D_C > 0\}, \quad (6.27)$$

где разница между RV получается путем применения свертки ПРВ D_{\min} и $-D_C$.

Кроме того, входные данные во входящие потоки

$$\lambda_i = \frac{(1 - p_{M+1})}{M} \lambda, i = 1, 2, \dots, M, \quad (6.28)$$

в то время как входящие потоки — пуассонлвские, из-за прямого применения теоремы разложения Райкова [307]. Каждый входящий поток затем

классифицируется как очередная СМО с теми же параметрами ввода и обслуживания. В предыдущем разделе он был подробно проанализирован в отношении его характеристик задержки и потерь.

Рассмотрим теперь последнюю СеМО. Заметим, что она обслуживает мультиплексный трафик из всех входящих потоков M , а также долю трафика $p_{M+1}\lambda$, поступающую непосредственно на БС. Последний процесс также пуассоновский из-за теоремы суперпозиции Райкова. Кроме того, поскольку радиус зоны обслуживания R ратифицирован 3GPP на величину, равную 30 км, предполагается, что количество подвижных БС также велико, а процессы отправления из всех входящих потоков стохастически эквивалентны. Таким образом, можно аппроксимировать совокупный трафик, выходящий из всех входящих потоков M пуассоновским процессом с интенсивностью

$$\begin{aligned}\lambda_{M+1} &= M(1 - p_L) \frac{(1 - p_{M+1})}{M} \lambda + p_{M+1} \lambda = \\ &= \lambda[(1 - p_L)(1 - p_{M+1}) + p_{M+1}],\end{aligned}\quad (6.29)$$

где p_L — вероятность потери сообщения для любого входящего потока. Как можно заметить, очередь $M + 1$ имеет тот же тип, что и используемые для описания сдвигов, и также была проанализирована в предыдущем подразделе.

В результате общая вероятность потери сообщений равна

$$\tilde{p}_L = 1 - (1 - p_L)(1 - p_{L,M+1}),\quad (6.30)$$

где $p_{L,M+1}$ — вероятность потери сообщений для очереди $M + 1$. ПРВ совместного времени ожидания, $f_{T_W}(x)$, обеспечивается прямой сверткой ПРВ задержек, которые возникают в произвольно выбранном входящем потоке и в очереди $M + 1$.

Энергия на транзакцию и время жизни датчика

Изменяя подход, применяемый для базовой модели, могут быть улучшены некоторые параметры времени жизни датчика и энергия на транзакцию. Главное различие заключается в том, что теперь датчик может использовать подвижные БС для передачи определенной части своих сообщений. Вводя q_R и q_B в качестве пропорций времени, когда случайно выбранный датчик отправляет свои сообщения через ретрансляционные БС и основные БС соответственно, требуемые ПРВ энергии за транзакцию и срок службы датчика можно выразить как

$$\begin{aligned} f_E(x) &= q_R f_{E_R}(x) + q_B f_{E_B}(x), \\ f_L(x) &= q_R f_{L_R}(x) + q_B f_{L_B}(x), \end{aligned} \quad (6.31)$$

где индексы R и B относятся к ПРВ энергии за транзакцию и времени жизни датчика, связанным с использованием подвижных БС и БС соответственно. Заметим, что ПРВ $f_{E_B}(x)$ и $f_{L_B}(x)$ были получены ранее. Тогда ПРВ $f_{E_R}(x)$ и $f_{L_R}(x)$ можно легко получить, заменив ПРВ расстояния до БС на ПРВ расстояния до ближайшей подвижной БС как показано в (6.25).

Наконец, установив периоды времени, когда датчик передает свои сообщения через подвижные БС и БС, q_R и q_B , замечаем, что они совпадают с вероятностями, когда случайное поступающее сообщение отправляется через соответствующую БС. Эти вероятности также были рассчитаны ранее в (6.27), что приводит к

$$q_R = 1 - p_{M+1}, q_B = p_{M+1}. \quad (6.32)$$

6.7 Численные результаты

В этом разделе мы представляем численные результаты, которые показывают производительность предлагаемой модели системы, когда ретрансляторы, установленные на транспортном средстве, помогают передавать данные на БС. Сначала мы фиксируем вероятность потери пакета и эффективность использования энергии для базовой установки NB-IoT без помощи ретрансляционных устройств. Затем мы добавляем ретрансляторы в систему и фиксируем те же показатели, что подчеркивает важность разделения радиоресурсов между мобильными ретрансляционными БС и сотовой БС. Наконец, мы оцениваем достижимые улучшения в производительности, которые вносит усовершенствованная система оптимально распределяющая ресурсы между БС установленной на автомобиле и стационарной БС. Для всесторонней проверки нашей аналитической методологии мы также сравниваем полученные характеристики с использованием предлагаемой математической модели с результатами, полученными общесистемными оценками. Указанная проверка проводится с использованием системы имитационного моделирования WINTERsim [316], который был широко использован в наших предыдущих исследованиях для оценки различных сценариев IoT [277].

6.7.1 Оценка базовой системы без ретрансляции

Сначала рассмотрим вероятность потери сообщений в базовом сценарии как функцию интенсивности поступления сообщений от одного датчика λ_I , который проиллюстрирован на рисунке 6.15. Мы меняем эту интенсивность, когда один пакет генерируется каждые 5 минут, и до того момента, когда пакет прибывает каждые 10 секунд. Это соответствует ситуациям с непостоянными обновлениями (например, мониторинг окружающей среды для температуры, давления воздуха и т. д.), а также слу-

Обозначения используемые установкой моделирования.

Параметр	Значение
Радиус городской зоны обслуживания, R	10км
Радиус сельской зоны обслуживания, R	30км
Число подвижных БС в зоне, M	$0, \dots, 1000$
Интенсивность генерации сообщений датчика, λ_I	4 за мин
Средняя плотность датчиков в области, σ	$10, \dots, 1000$ на 2
Общее число NB-IoT каналов, K	72
Позиции ожидания на БС, L	$3 \times K$
Доля каналов, расположенных на транспортных средствах, γ_K	$0, \dots, 0.95$
Длительность кадра, Δ	10 мс

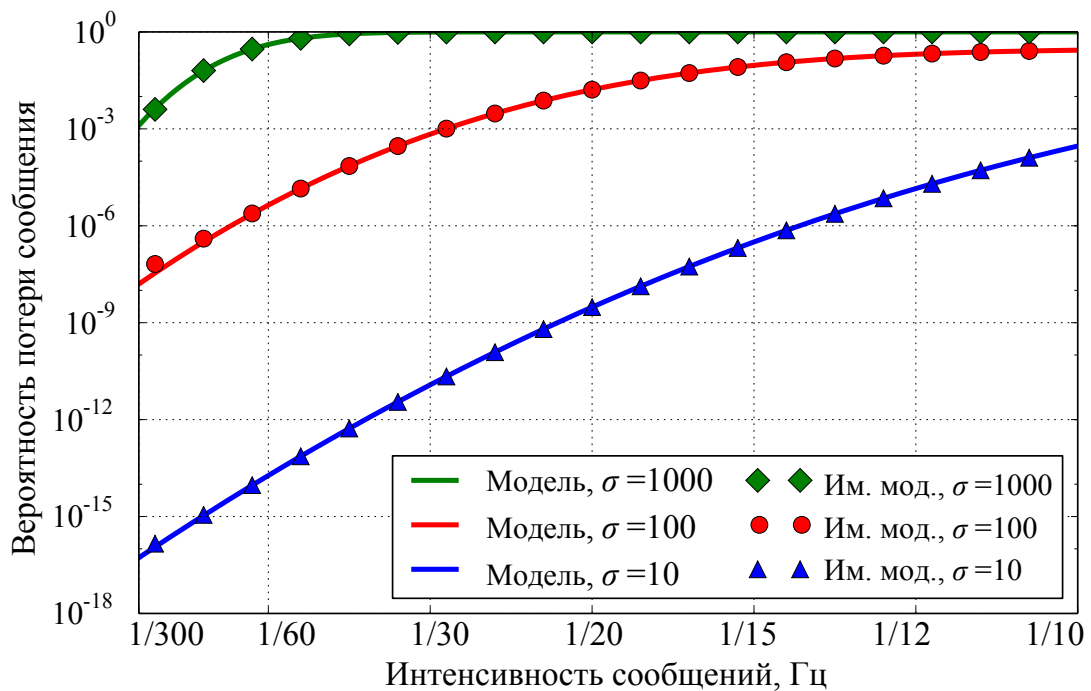


Рисунок 6.15 — Вероятность потери сообщения в базовом сценарии

$$R=10 \text{ км}$$

чаям относительно интенсивного трафика (например, аварийные сигналы обнаружения вторжений). На рисунке 6.15 показана вероятность потери сообщений для городского развертывания сети с радиусом действия соты

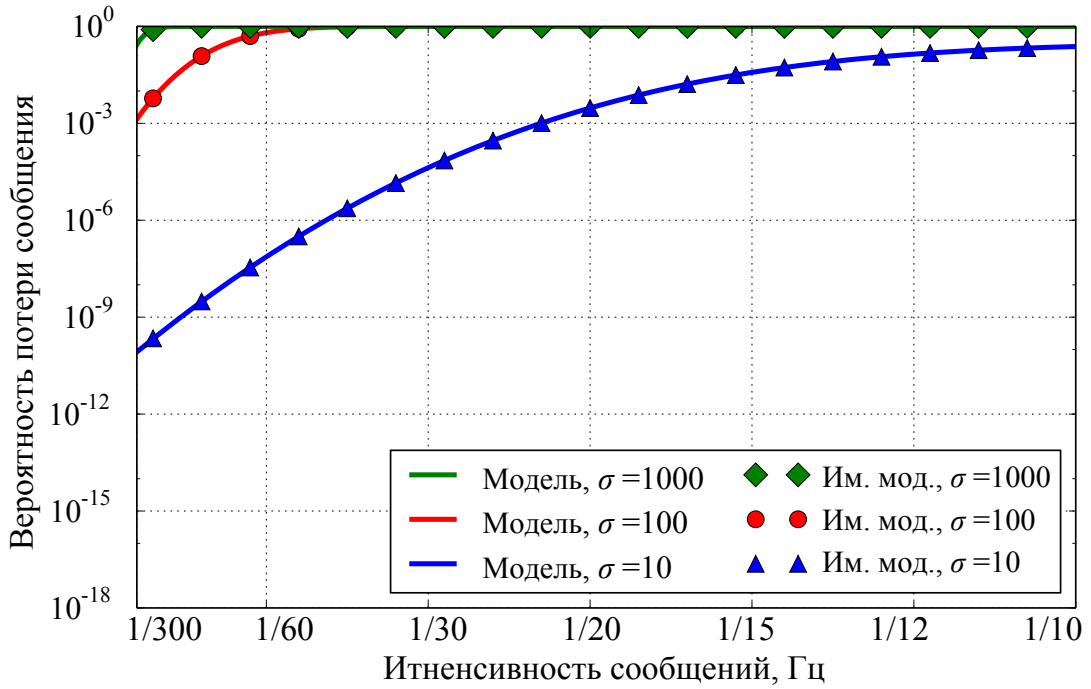


Рисунок 6.16 — Вероятность потери сообщения в базовом сценарии $R=30$ км

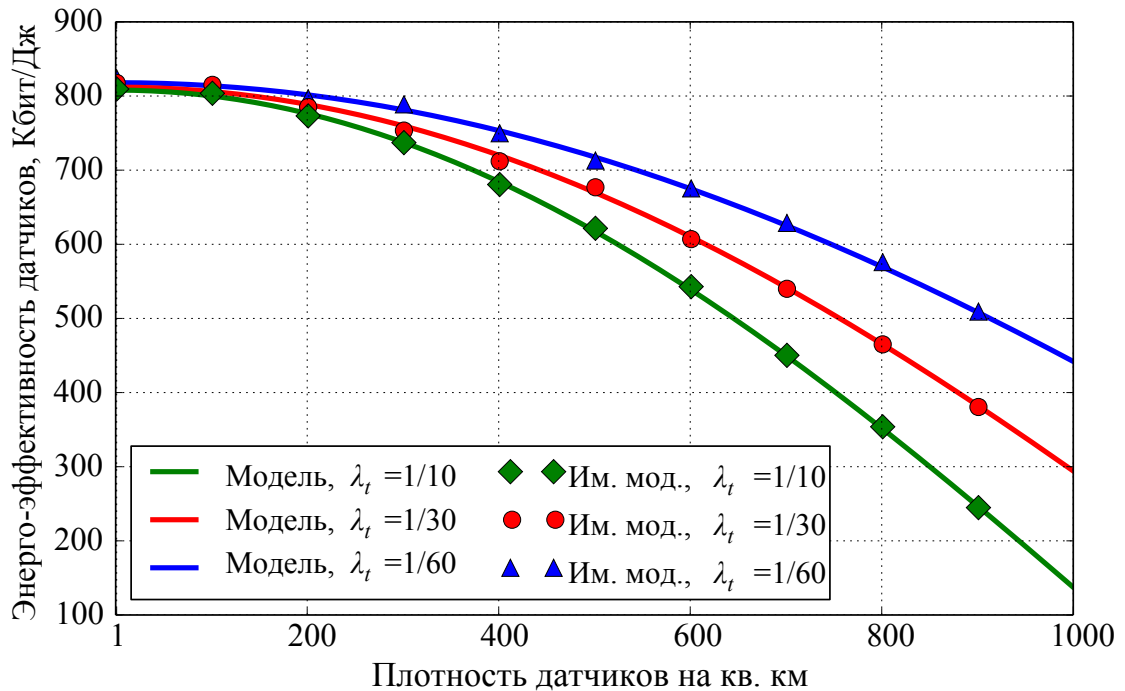


Рисунок 6.17 — Энергоэффективность устройств IoT в базовом сценарии $R=10$ км

10 км, а на рисунке 6.16 показывает то же самое в сценарии сот с радиусом действия 30 км [299].

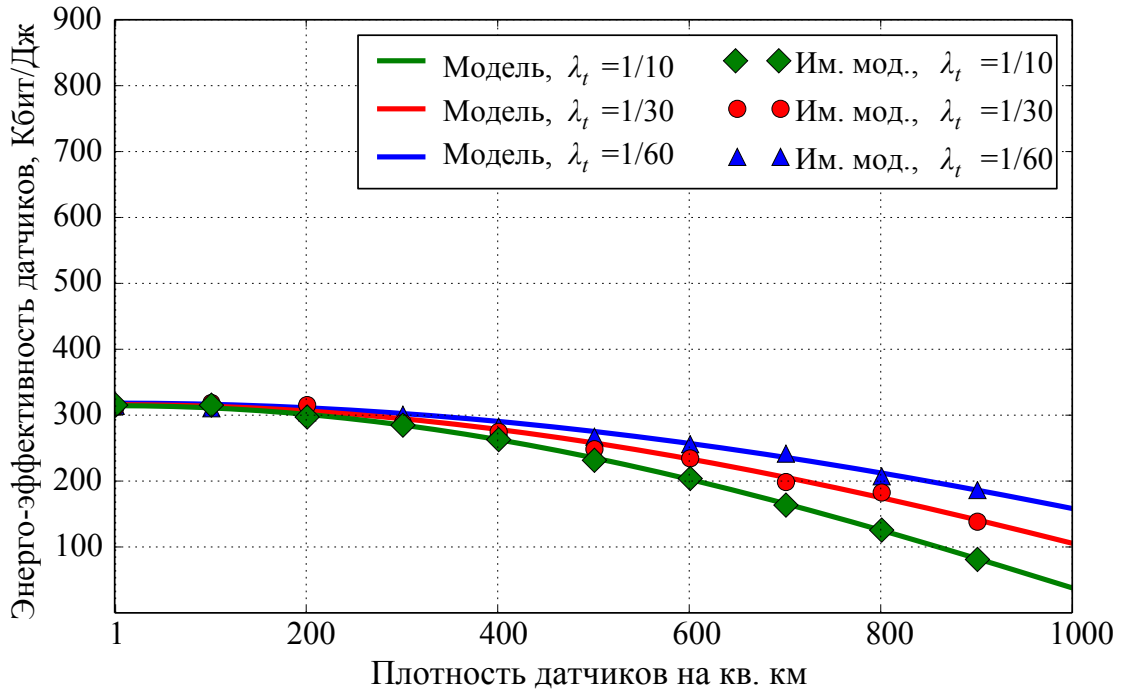


Рисунок 6.18 — Энергоэффективность устройств IoT в базовом сценарии $R=30$ км

Оценивая поведение системы в городском сценарии, на рисунке 6.15 показано, что БС NB-IoT может обеспечить надежную передачу данных (то есть вероятность потери пакета менее 10^{-6}) для $\sigma = 10$ подключенных датчиков на км^2 , которые генерируют приблизительно $\lambda_I = 4$ пакетов в минуту от одного устройства IoT. Для более высокой плотности подключенных датчиков ($\sigma = 100$ подключенных устройств на км^2) порог потери 10^{-6} достигается при $\lambda_I = 1$ пакете в минуту. Для плотности 1000 датчиков на квадратный километр любая надежность передачи пакетов не может быть гарантирована (вероятность потери пакета превышает 10^{-3}). Поэтому базовая система не может справиться с доставкой чувствительных к задержке данных при массовом использовании системы. Однако для меньшего количества подключенных устройств IoT может быть обеспечена надежная связь. Эта ситуация становится еще хуже в сельской местности, где плотность $\sigma = 100$ датчиков на км^2 уже приводит к более 10^{-3} вероятности потери пакетов.

Соответствующее среднее значение энергоэффективности в описанном сценарии представлено на рисунке 6.17 отражается городской случай,

а рисунок 6.18 соответствует сельскому случаю. Как можно заметить, передачи коротких пакетов по NB-IoT по меньшей мере на порядок превосходят по производительности аналогичную в LTE [317; 318], хотя радиус действия соты примерно в 50 раз больше (200 против 10 км). Тем не менее, желаемый срок службы батареи 10+ лет вряд ли может быть достигнут в рассматриваемом сценарии для типичного недорогого датчика, оснащенного батареей емкостью 50 мАч и напряжением 3 В и передающего сообщения всего в 10 байт. Фактическая энергоэффективность приводит к максимальному времени автономной работы 1 – 8 лет, хотя предполагается, что вся емкость аккумулятора предназначена исключительно для радиосвязи (подробнее см. Раздел 6.6). Вследствие больших (средних) расстояний между БС и устройствами IoT в сельском сценарии соответствующие значения энергоэффективности примерно в 2.6 раза ниже. Несмотря на то, что радиус ячейки в 3 раза больше, коэффициент энергоэффективности здесь не равен 3, так как зависимость между количествами потребляемой энергии и радиусом ячейки не является линейной.

6.7.2 Оценка усовершенствованной системы с ретрансляцией

После понимания показателей, представляющих интерес в базовом сценарии, мы приступаем к характеристике производительности усовершенствованной системы, которая помогает доставлять пакеты от датчиков к БС с помощью передвижных БС, установленных на транспортных средствах. Во-первых, напомним, что пул доступных каналов NB-IoT в кадре K должен быть разделен между ретрансляционными БС (транспортными средствами), K_1 и стационарными БС, K_2 , $K_1 + K_2 = K$, так чтобы не было чрезмерных помех вызванных их одновременной эксплуатацией. Введем здесь отношение $\gamma_K = K_1/K$, определяющее долю от общего количества каналов, выделенных для транспортных средств. Если $\gamma_K = 0$, улучшенный сценарий сводится к базовому случаю, так как нет ресурсов, выделенных для передвижных БС, то есть поддержка мобильных БС от-

ключена. Также, когда $\gamma_K = 1$, все каналы назначаются транспортным средствам.

Важность оптимального распределения между радиоресурсами, выделяемыми вспомогательными транспортными средствами, и для стационарных БС, выделена на рисунке 6.19, которая представляет вероятность потери сообщений в зависимости от γ_K для сельской местности с $\sigma = 10$ подключенными устройствами IoT на км², а также интенсивность поступления одного пакета в минуту от одного датчика, $\lambda_I = 1$. Как можно заметить, улучшенная система ведет себя аналогично базовому сценарию при $\gamma_K = 0$. Затем, когда γ_K увеличивается, надежность доставки пакетов улучшается до тех пор, пока не будет хватать ресурсов для поддержки всех датчиков без достаточного количества вспомогательных транспортных средств в непосредственной близости. После этого момента вероятность потери сообщения начинает расти и в конечном итоге может превышать ее в базовой системе.

Анализируя результаты на рисунке 6.19, заметим, что для любого набора входных параметров существует оптимальное значение γ_K , которое дает наименьшую возможную вероятность потери пакета. Когда γ_K становится больше оптимальных значений, производительность уменьшается, и в конечном итоге вероятность потери пакета сходится к 1 для всех рассмотренных входных параметров. Это связано с тем, что в этом случае все каналы назначаются мобильным БС, и нет каналов для передачи генерируемого трафика на NB-IoT БС. Например, для $M = 1000$ оптимальное значение γ_K находится около $\gamma_K = 0.55$, что приводит к значительно более высокой надежности передачи сообщений.

Подводя итог этому подразделу, заключаем, что использование оптимального значения γ_K имеет решающее значение для улучшения общей производительности системы. Зная количество доступных вспомогательных транспортных средств в пределах БС, этот параметр можно настроить динамически, чтобы ресурсы распределялись в зависимости от ситуации.

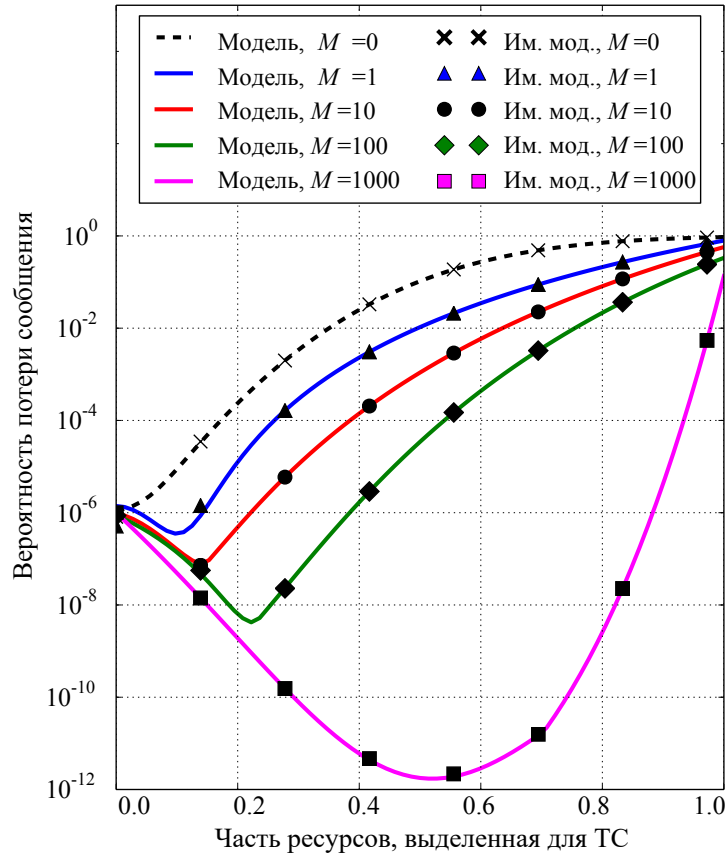


Рисунок 6.19 — Результат разделения ресурсов между кооператорами и БС

6.7.3 Оценка расширенной системы

Теперь мы сосредоточим внимание на прогнозируемых показателях производительности в расширенной системе, где γ_K чтобы уменьшить вероятность потери пакетов устанавливается на наиболее эффективное значение, определяемое как значение γ_K , основываясь на результатах на рисунке 6.19. Здесь представлены три первичные метрики, представляющие интерес: вероятность потери пакетов, энергоэффективность и время работы датчика. В дальнейшем мы концентрируемся на сельское размещение с сотой радиусом $R = 30$. Городское размещение характеризуется подобным поведением для рассматриваемых параметров. На рисунке 6.20 подтверждается, что вероятность потери сообщения быстро уменьшается по мере увеличения количества вспомогательных транспортных средств. Как указано в подразделе 6.7.1, эта тенденция не только повышает надежность

при подключении до сотни датчиков на км², но также обеспечивает более надежную доставку данных в сильно перегруженных случаях при тысячи подключенных устройств на км², что не может быть достигнуто базовом сценарии.

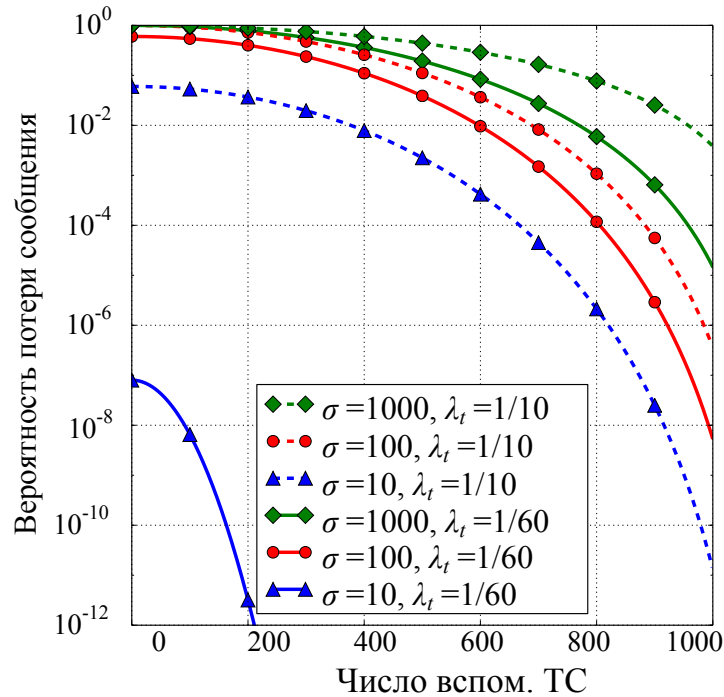


Рисунок 6.20 — Зависимость вероятности потери сообщения от количества автомобилей в сети

Наряду с значительным повышением надежности усовершенствованная система характеризуется заметным повышением энергоэффективности, как подчеркивается в рисунке 6.21. Начиная с сотен кБит/Дж для небольшого количества мобильных БС (до $M = 300$) улучшенная система превосходит свой базовый аналог с точки зрения энергоэффективности в 6 раз, при $M = 1000$. Это улучшение, в свою очередь, позволяет значительно повысить время работы от батареи, как показано на рисунке 6.22.

Наш предлагаемый подход — это возможная стратегия, которая не только удовлетворяет требованиям к сроку службы батареи до 10 лет, но и еще больше повышает ключевой показатель производительности при различных плотностях развертывания датчиков. Это включает в себя крайние случаи с $\sigma = 1000$ подключенных устройств IoT на км², которые приводят к большому объему (2.8 миллиона) датчиков в пределах одной БС.

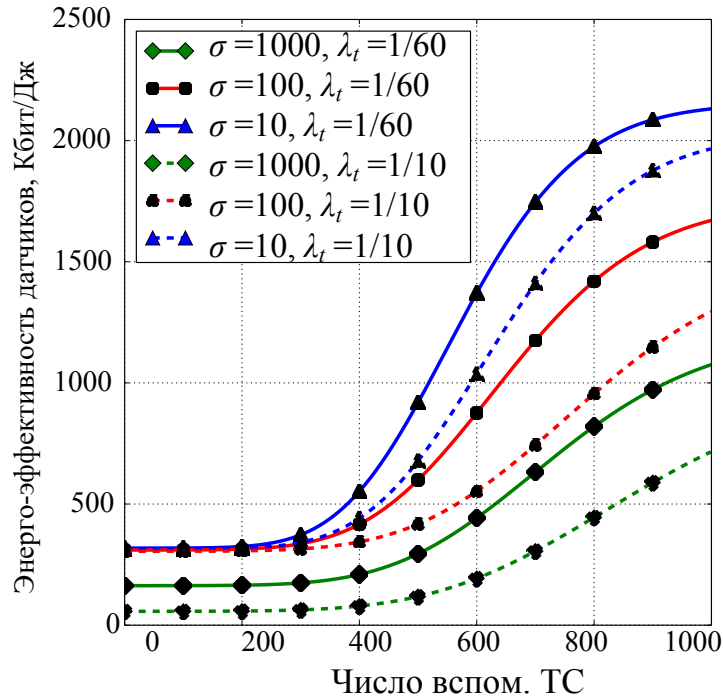


Рисунок 6.21 — Зависимость энергоэффективности от количества автомобилей в сети

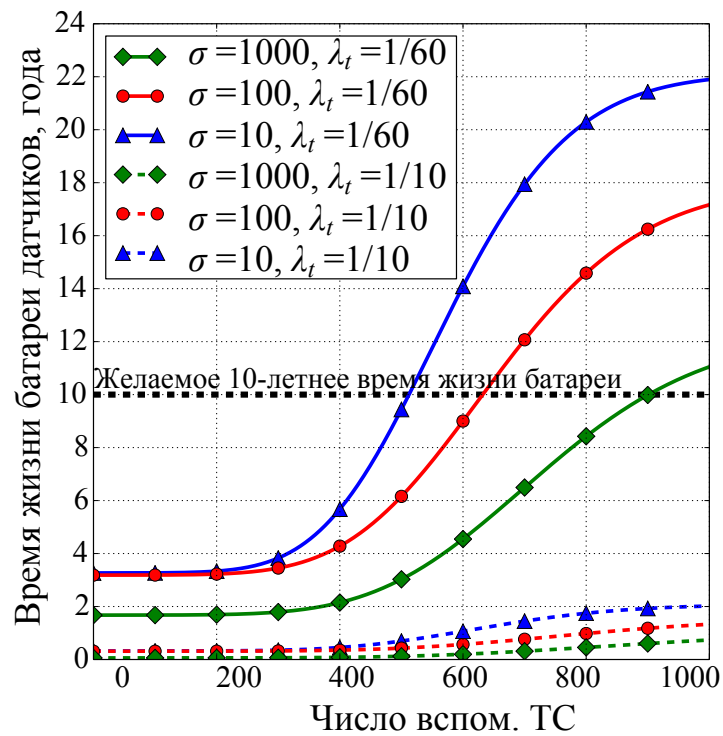


Рисунок 6.22 — Зависимость времени жизни батареи от количества автомобилей в сети

Как показано на рисунках 6.20 и 6.22 оптимально спланированное повторное использование радиоресурсов, предоставляемое нашей расширен-

ной системой, приводит не только к возможности надежного размещения таких массивных систем, но и к существенному улучшению времени жизни батареи. В то же время, независимо от фактической плотности подключенных устройств IoT, целевой 10-летний срок службы батареи вряд ли может быть достигнут при использовании датчика с батареей 50 мАч и 3 В с любым разумным количеством вспомогательных транспортных средств, если этот датчик передает свои сообщения чаще, чем раз в минуту.

6.8 Выводы

1. Разработана модель для анализа показателей производительности системы NB-IoT и проведен анализ производительности технологии NB-IoT с точки зрения вероятностных характеристик обслуживания, таких как вероятность потери сообщения, распределение и средняя задержка сообщения в системе. Представленная модель позволила определить диапазоны нагрузок, при которых выполняются заданные параметры качества обслуживания устройств. Получены численные результаты для сот радиусов 10 и 30 километров.
2. Разработана модель и метод доставки данных от M2M устройств в гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяющие для ряда сценариев как повысить вероятность доставки сообщений, так и повысить энергоэффективность передающих узлов.

Заключение

Основные научные результаты работы заключаются в следующем:

1. Анализ развития новых технологий, концепций развития сетей и систем телекоммуникаций, таких как Интернет Вещей, показывает, что в обозримой перспективе решающую роль в реализации сетей и систем пятого поколения и последующих будут играть гетерогенные структуры, основанные на интеграции множества используемых радио технологий в условиях сверхплотных сетей связи. При этом для создания перспективных сетей и систем связи требуется как разработка новой методологии исследования гетерогенных сетей, так и комплекса взаимоувязанных адекватных новому этапу развития сетей моделей и методов распределения ресурсов.
2. Предложен новый критерий оптимизации распределения ресурсов в гетерогенных сетях связи, отличающийся от известных тем, что справедливое распределение ресурсов для гетерогенных сетей с учетом требуемого баланса между сетевыми возможностями и пропускной способностью достигается за счет использования критерия $\max\text{-min}$, основанного на полосе пропускания. Целевая функция при этом дополняется весовыми коэффициентами на основе мгновенных значений спектральной эффективности.
3. Использование предложенного критерия позволяет сформулировать задачи оптимизации как задачи линейного программирования, предложенные решения хорошо масштабируемы с ростом числа пользователей и инфраструктурных узлов, даже в условиях создания сетей сверх высокой плотности.
4. Разработан метод управления производительностью гетерогенных радиотехнологий доступа H-CRAN, отличающийся от известных тем, что используется кооперативное управление посредством менеджера кооперативных радиоресурсов CRRM в H-CRAN в реальном времени на основе динамического управления ресурсами двух альтернативных метрик: справедливого распределения ресур-

сов для пользователей всех доступных сетей радиодоступа и общей производительности системы.

5. Разработана тестовая среда для H-CRAN, которая использует современную архитектуру программно-конфигурируемой сети SDN и протокол OpenFlow для динамического и эффективного управления связностью пользовательского оборудования. Экспериментальные исследования на тестовой среде подтвердили основные преимущества, которые получает облачная сеть доступа при использовании менеджера кооперативных радио ресурсов.
6. Анализ особенностей перехода к сетям связи пятого поколения показал, что эти сети количественно и качественно будут принципиально отличаться от всех существующих сетей связи как по плотности размещения устройств сети, так и по требуемым задержкам. При этом сети связи пятого поколения характеризуются как сверхплотные сети (до 1 млн устройств на кв.км) и сети с ультра малыми задержками. Это требует разработки новой методологии распределения ресурсов в таких сетях и новых технологий взаимодействия устройств на относительно небольших расстояниях.
7. Предложена новая классификация гетерогенных беспроводных сетей связи, подразделяющая все гетерогенные беспроводные сети на три типа.
8. Предложен новый шлюз входа в сеть доступа (AAGW), который в отличие от предыдущих архитектурных решений не требует обеспечения доступа ни в операторскую магистральную сеть, ни в интернет и отражает функциональность решений 3GPP во WLAN радио технологиях RAT (например, WiFi) и наоборот. Это, в свою очередь, позволяет пользовательскому оборудованию максимально использовать доступные радио интерфейсы без необходимости применения отдельных решений для мобильности для сетей доступа не-3GPP.
9. Разработан алгоритм выбора мощности передачи мобильного устройства в условиях наличия двух и более радиотехнологий, способных функционировать одновременно. Показано, что предлага-

емый алгоритм максимизируют энергоэффективность мобильного устройства при обеспечении минимально требуемой скорости передачи и ограничении сверху на мощность передачи.

10. Анализ кооперативных коммуникаций вкупе с анализом особенностей машино-ориентированных коммуникаций и взаимодействий машина-машина M2M показал, что до настоящего времени возможности использования клиентской ретрансляции для трафика M2M практически не изучены.
11. Исследована трехузловая модель клиентской ретрансляции, а также возможности ее облачной реализации, что дало возможность предложить такие модели для обеспечения приемлемого уровня качества обслуживания в сетях M2M.
12. Разработаны модель и метод облачной клиентской ретрансляции, учитывающие особенности трафика для устройств M2M, расположенных на границе соты, позволяющие увеличить производительность системы в 3 - 4 раза при уменьшении в несколько раз задержки пакетов.
13. Анализ возможностей технологии взаимодействия устройство-устройство D2D показал, что функциональное расширение возможностей технологии D2D может быть достигнуто за счет использования поддержки со стороны сети, т.е. при совместном использовании ресурсов D2D и сотовой сети.
14. С учетом функционирования устройств D2D, как правило, в условиях сетей сверхвысокой плотности и в нелицензируемом спектре, определены основополагающие теоретические зависимости между плотностью передающих устройств, размещенных на некоторой плоскости и интерференционным расстоянием для взаимодействующих устройств D2D. При этом также получена зависимость математического ожидания и среднеквадратического отклонения величины отношения сигнала к интерференции SIR от параметра распределения дистанции до интерферирующего передатчика.
15. Предложена новая архитектура сетей связи следующего поколения с широким использованием технологии D2D, отличающаяся от

типовой архитектуры 3GPP единственным дополнительным объектом — сервером D2D, находящимся в EPC фрагменте сети. Такое размещение позволяет серверу D2D связываться с центром местоположения (SMLC), чтобы позиционировать пользовательское оборудование, а также эффективно взаимодействовать с серверами приложений внешних провайдеров и операторов.

16. Разработаны модели и методы выгрузки трафика в гетерогенных сетях, использующих для выгрузки трафика технологию D2D на базе WiFi Direct с сетевой поддержкой, обеспечивающие удвоение пропускной способности соты и повышение энергоэффективности до 6 раз.
17. Для натуральных экспериментов была использована модельная сотовая сеть на базе департамента телекоммуникаций Технологического Университета Брно (BUT) в Чехии. BUT LTE зона полностью коммерчески пригодная версия для внедрения всех важнейших подсистем, входящих в современные сети 4G. Сервер D2D был реализован как виртуальное устройство. При натуральных экспериментах наблюдался значимый синергетический эффект при совместном использовании мобильной сети и сети D2D.
18. Разработана аналитическая модель для протокола IEEE 802.11ah, позволяющего обеспечить передачу данных для большого числа узлов в едином кластере. Все узлы такой сети представлены предсказуемыми характеристиками трафика, а именно: передачей одного малого сообщения за длительный промежуток времени. Предложенная модель отличается высокой скоростью работы по сравнению с известными моделями, что дает возможность получения результатов на основе имитационного моделирования с сохранением высокой точности.
19. Разработана математическая модель для анализа поведения канала множественного доступа IEEE 802.11ah, сформированы диаграмма состояний, матрица переходных вероятностей и соответствующий вектор стационарного распределения. Определены полезная пропускная способность канала (бит сообщения без учёта

заголовков в секунду) и энергетические затраты устройств на передачу.

20. Разработаны Модель и метод доступа устройств M2M к ресурсам сети LTE на основе модифицированной процедуры COBALT, учитывающие особенности передачи данных малого объема, позволяющие уменьшить задержку пакетов по сравнению с методом PUSCH почти для 85% типов трафика и одновременно в 2-3 раза снизить энергопотребление по сравнению с методом PRACH.
21. Разработана модель для анализа показателей производительности системы NB-IoT и проведен анализ производительности технологии NB-IoT с точки зрения вероятностных характеристик обслуживания, таких как вероятность потери сообщения, распределение и средняя задержка сообщения в системе. Представленная модель позволила определить диапазоны нагрузок, при которых выполняются заданные параметры качества обслуживания устройств. Получены численные результаты для сот радиусов 10 и 30 километров.
22. Разработана модель и метод доставки данных от M2M устройств в гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяющие для ряда сценариев как повысить вероятность доставки сообщений, так и повысить энергоэффективность передающих узлов.

Список терминов

Русскоязычные сокращения

БС — базовая станция.

ККТ — подход Каруша-Куна-Таккера.

МС — мобильная станция.

ПРВ — плотность распределения вероятностей.

СМО — система массового обслуживания.

ССОП — сеть связи общего пользования.

ТфОП — телефонные сети общего пользования.

Англоязычные сокращения

3GPP — 3rd Generation Partnership Project — партнерский проект третьего поколения.

AAGW — alien access gateway — шлюз входа в сеть доступа.

ACK — acknowledgement — подтверждение.

ANDSF — access network discovery and selection function — функция обнаружения и выбора сети доступа.

AP — access point — точка доступа.

BBU — base band unit — централизованный узел.

BC — backoff counter — счетчик выдержки.

BEB — binary exponential backoff — алгоритмом двоичной экспоненциальной выдержки.

BDT — bi-directional TXOP — реверсивная схема TXOP.

BLE — Bluetooth low energy — Bluetooth с низким энергопотреблением.

BSS — basic service set — базовый набор услуг.

CC — conventional cooperation — обычная кооперация.

CDF — cumulative distribution function — кумулятивная функция распределения.

COOP — cooperation opportunistic — добровольная кооперация.

COBALT — contention-based LTE transmission — конкурентный доступ для систем LTE.

Cloud RAN — cloud radio access network — облачная сеть радиодоступа.

CRRM — cooperative radio resource manager — менеджер кооперативных радио ресурсов.

CSI — channel state information — информация о состоянии канала.

CW — collision window — окно коллизии.

CR — client relay — клиентская ретрансляция.

DL — downlink — нисходящий канал.

D2D — device-to-device — прямое взаимодействие «устройство-устройство».

DIFS — DCF Interframe Space — промежуток времени между кадрами.

EAB — Extended Access Barring — схема расширенного исключительного доступа.

ECP — equivalent constrained problem — эквивалентная задача с ограничениями.

EDCA — enhanced distributed channel access — расширенный распределенный канальный доступ.

EPC — evolved packet core — эволюционирующее ядро сети.

EUP — equivalent unconstrained problem — эквивалентная задача без ограничений.

E-UTRAN — evolved universal terrestrial radio access network — эволюционирующая универсальная наземная сеть радиодоступа.

eWaitTimer — extended wait timer — механизм таймера расширенного ожидания.

EWMA — exponentially weighted moving average — экспоненциально-взвешенное скользящее среднее.

IoT — Internet of Things — Интернет вещей.

GW — gateway — шлюз.

FDD — frequency division duplex — дуплексное частотное разделение каналов.

H2H — human-to-human — трафик от человека к человеку.

H-CRAN — heterogeneous cloud radio access network — гетерогенные облачные радиотехнологии доступа.

HARQ — Hybrid Automatic Repeat reQuests — автоматический повтор запросов.

HetNet — heterogeneous network — гетерогенная сеть.

QoS — quality of service — качество обслуживания.

IEEE — Institute of Electrical and Electronics Engineers — Институт инженеров электротехники и электроники.

LAА — license assisted access — поддержка доступа в лицензируемом спектре.

LPN — low-power node — низко потребляющий узел.

LPWAN — low-power WAN — глобальная сеть с низким потреблением энергии.

LTE — long term evolution — системы связи долгосрочной эволюции.

MAC — medium access control — уровень управления доступа к каналу.

MCS — modulation and coding scheme — схема модуляции и кодирования.

MIMO — multiple input multiple output — многоантенная система.

MIP — manual input processing — смешанно-целочисленное программирование.

MME — mobility management entity — модуль управления мобильностью.

MPR — multi-Packet Reception — многопакетный прием.

MTC — machine type communications — межмашинное взаимодействие.

MU-MIMO — multi-user MIMO — многопользовательский MIMO.

MR — multiray reception — многолучевой прием.

M2M — machine-to-machine — межмашинное взаимодействие.

NAT — network address translation — преобразование сетевых адресов.

NB-IoT — narrow band Internet of Things — узкополосный Интернет Вещей.

OFDM — orthogonal frequency-division multiplexing — ортогональное частотное разделение каналов с мультиплексированием.

OFDM-MIMO — система, использующая технологии MIMO для сигналов OFDM.

P2P — peer-to-peer — пиринговые сети.

PDN — packet data network — сеть пакетной связи.

PHY — physical layer — физический уровень.

PLC — power line communications — связь по линиям электропередач.

PPP — Poisson point process — пуассоновский точечный поток.

RA — resource allocation — распределение ресурсов.

RACH — random access channel — канал случайного доступа.

RAT — radio access network — сеть радиодоступа.

RDM — random direction model — модель случайного направления.

RR — round robin — алгоритм распределения ресурсов с помощью циклического опроса.

RRH — remote radio head — удаленный головной радиоузел.

ProSe — proximity services — сервисы ближнего радиуса действия.

PUCCH — physical uplink control channel — физический управляющий канал для восходящей линии.

PRACH — physical random access channel — физический канал случайного доступа.

RAR — random access resolution — решение проблемы случайного доступа.

RAW — restricted access window — окно ограничивающего доступа.

SDN — software-defined network — программно-конфигурируемая сеть.

SE — spectral efficiency — спектральная эффективность.

SIFS — Short interframe spaces — короткое пространство между фреймами.

SINR — signal-to-interference-plus-noise ratio — отношения сигнала к интерференции плюс шум.

SGSN — serving GPRS support node — узел поддержки обслуживающего GPRS.

SMLC — serving mobile location center — центр местоположения.

SNR — signal to noise ratio — отношение сигнал/шум.

SS — subscriber station — абонентская станция.

SSL — secure sockets layer — уровень безопасного порта.

SST — sub-channel selective transmission — подканальная выборочная передача.

STA — station — станция.

TIM — traffic indication map — карта индикации трафика.

TCP — transmission control protocol — протокол доставки данных с контролем.

TXOP — transmission opportunity — возможность передачи.

TWT — target wakeup time — целевое время пробуждения.

UE — user equipment — мобильная станция.

VPN — virtual private network — виртуальная частная сеть.

UL — uplink — восходящий канал.

WiFi Direct — технология прямой связи WiFi.

WiMAX — worldwide Interoperability for Microwave Access — протокол беспроводного доступа IEEE серии 802.16.

WMAN — wireless metropolitan area network — беспроводная городская сеть.

WPAN — wireless personal area network — беспроводная персональная сеть.

WWAN — wireless wide area network — беспроводная глобальная сеть.

Список литературы

1. Evolving Wireless Communications: Addressing the Challenges and Expectations of the Future / Robert Baldemair, Erik Dahlman, Gabor Fodor et al. // *IEEE Vehicular Technology Magazine*. — 2013. — Vol. 8, no. 1. — Pp. 24–30. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6470755>.
2. Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks / Jian Li, Mugen Peng, Aolin Cheng et al. // *IEEE Systems Journal*. — 2014.
3. Intelligent Access Network Selection in Converged Multi-Radio Heterogeneous Networks / Sergey Andreev, Mikhail Gerasimenko, Olga Galinina et al. // *IEEE Wireless Communications Magazine*. — 2014. — Vol. TBD. — Pp. 1–10.
4. What will 5G be? / Jeffrey G Andrews, Stefano Buzzi, Wan Choi et al. // *IEEE Journal on selected areas in communications*. — 2014. — Vol. 32, no. 6. — Pp. 1065–1082. <http://arxiv.org/abs/1405.2957>.
5. Andrews Jeffrey G. Seven ways that HetNets are a cellular paradigm shift // *IEEE Communications Magazine*. — 2013. — Pp. 136–144. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6476878.
6. Network densification: the dominant theme for wireless evolution into 5G / Naga Bhushan, Junyi Li, Durga Malladi et al. // *IEEE Communications Magazine*. — 2014. — Vol. 52, no. 2. — Pp. 82–89. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Network+Densification+:+The+Dominant+Theme+for+Wireless+Evolution+into+5G#0http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6736747>.
7. Femtocells: Past, Present, and Future / Jeffrey Andrews, Holger Claussen, Mischa Dohler et al. // *IEEE Journal on Selected Areas in Communications*. — 2012. — Vol. 30, no. 3. — Pp. 497–

508. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6171992
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6171992>.
8. *Singh Sarabjot, Dhillon Harpreet, Andrews Jeffrey*. Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights // *IEEE Transactions on Wireless Communications*. — 2013. — Vol. 12, no. 5. — Pp. 2484–2497. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6497002>.
 9. When Cellular Meets WiFi in Wireless Small Cell Networks / Mehdi Bennis, Meryem Simsek, Andreas Czylik et al. // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 6. — Pp. 44–50.
 10. LTE Release 12 and beyond / David Astely, Erik Dahlman, Gabor Fodor et al. // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 7. — Pp. 154–160. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6553692.
 11. *Кучерявый А.Е., Кучерявый Е.А., Харью Я.* Качество обслуживания в сети Интернет // *Электросвязь*. — 2002. — no. 1. — Pp. 9–14.
 12. An Overview of Load Balancing in HetNets: Old Myths and Open Problems / Jeffrey G Andrews, Sarabjot Singh, Qiaoyang Ye et al. // *IEEE Wireless Communications*. — 2013. — Vol. 21, no. 2. — Pp. 18–25.
 13. Networks and Devices for the 5G Era / Boyd Bangerter, Shilpa Talwar, Reza Arefi, Ken Stewart // *IEEE Communications Magazine*. — 2014. — no. February. — Pp. 90–96.
 14. *Shuminoski Tomislav, Janevski Toni*. Radio Network Aggregation for 5G Mobile Terminals in Heterogeneous Wireless and Mobile Networks // *Wireless Personal Communications*. — 2014. — Vol. TBD. — Pp. 1–19. <http://link.springer.com/10.1007/s11277-014-1813-0>.
 15. Architecture enhancements for non-3GPP accesses // *3GPP Technical specification (TS) 23.402*. — 2013.

16. *Chin Woon, Fan Zhong, Haines Russell*. Emerging technologies and research challenges for 5G wireless networks // *IEEE Wireless Communications*. — 2014. — Vol. 21, no. 2. — Pp. 106–112. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6812298<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6812298>.
17. Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies / Mugen Peng, Yuan Li, Jiamo Jiang et al. // *IEEE Wireless Communications*. — 2014. — Vol. 21, no. 6. — Pp. 126–135.
18. <http://winter-group.net/rohde-schwarz-tutorial/>.
19. *Bertsekas Dimitri P, Gallager Robert G, Humblet Pierre*. Data networks. — Prentice Hall, 1992.
20. *Pióro Michal, Medhi Deep*. Routing, Flow, and Capacity Design in Communication and Computer Networks. — 1 edition. — 2004.
21. *Maros István*. Computational Techniques of the Simplex Method. — 2003.
22. *Massoulié Laurent, Roberts James*. Bandwidth Sharing: Objectives and Algorithms // *IEEE/ACM Transactions on Networking*. — 2002. — Vol. 10, no. 3. — Pp. 320–328.
23. *Кучерявый Евгений Андреевич*. Управление трафиком и качество обслуживания в сети Интернет. — 2004. — P. 336.
24. *Омётов А.Я., Андреев С.Д., Кучерявый Е.А.* Алгоритм кластеризации мобильных станций на основании пространственной и социальной метрик // *Информационные технологии и телекоммуникации*. — 2015. — Vol. 11, no. 3. — Pp. 45–53.
25. *Кучерявый Е.А., Самуйлов К.Е.* Методология распределения ресурсов в гетерогенных сетях // *Электросвязь*. — 2008. — no. 4. — Pp. 34–40.

26. Path optimization for elastic traffic under fairness constraints / G. Fodor, G. Malicsco, M. Pioro, T. Szymanski // ITC-17. — 2001. — Pp. 667–680.
27. Study on WLAN/3GPP Radio Interworking // *3GPP Technical Report (TR) 37.834*. — 2013.
28. Further advancements for E-UTRA physical layer aspects // *3GPP Technical Report (TR) 36.814*. — 2010.
29. Coordinated multi-point operation for LTE physical layer aspects // *3GPP Technical Report (TR) 36.819*. — 2011.
30. *ITU-R M.2135*. Guidelines for evaluation of radio interface technologies for IMT-Advanced. — 2009.
31. Characterizing Performance of Load-Aware Network Selection in Multi-Radio (WiFi/LTE) Heterogeneous Networks / Mikhail Gerasimenko, Nageen Himayat, Shu-ping Yeh et al. // Proc. of the Globecom Workshops. — 2013.
32. Multi-Radio Heterogeneous Networks: Architectures and Performance / Nageen Himayat, Shu-ping Yeh, Ali Panah et al. // Proc. of IEEE ICNC. — 2013.
33. WINTERSim system-level simulator, 2015 // <http://winter-group.net/downloads/>.
34. Evolution Towards 5G Multi-tier Cellular Wireless Networks: An Interference Management Perspective / Ekram Hossain, Mehdi Rasti, Hina Tabassum, Amr Abdelnasser // *IEEE Wireless Communications Magazine*, to appear. — 2014. — Vol. TBD. — Pp. 1–10. <http://arxiv.org/abs/1401.5530>.
35. C-RAN The Road Towards Green RAN // China Mobile Research Institute. — 2011. — oct.
36. Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory / Seok-Hwan Park,

- Oswaldo Simeone, Onur Sahin, Shlomo Shamai Shitz // *IEEE Signal Processing Magazine*. — 2014. — Vol. 31, no. 6. — Pp. 69–79.
37. Yang Yufei, Quek Tony QS, Duan Lingjie. Backhaul-constrained small cell networks: Refunding and QoS provisioning // *IEEE Transactions on Wireless Communications*. — 2014. — Vol. 13, no. 9. — Pp. 5148–5161.
38. Zhou Yuhan, Yu Wei. Optimized backhaul compression for uplink cloud radio access network // *IEEE Journal on Selected Areas in Communications*. — 2014. — Vol. 32, no. 6. — Pp. 1295–1307.
39. Investigation of cell association techniques in uplink cloud radio access networks / Shi Yan, Wenbo Wang, Zhongyuan Zhao, Abana Ahmed // *Transactions on Emerging Telecommunications Technologies*. — 2016. — Vol. 27, no. 8. — Pp. 1044–1054.
40. Peng Mugen, Yan Shi, Poor H Vincent. Ergodic capacity analysis of remote radio head associations in cloud radio access networks // *IEEE Wireless Communications Letters*. — 2014. — Vol. 3, no. 4. — Pp. 365–368.
41. Ding Zhiguo, Poor H Vincent. The use of spatially random base stations in cloud radio access networks // *IEEE signal processing letters*. — 2013. — Vol. 20, no. 11. — Pp. 1138–1141.
42. Liu An, Lau Vincent KN. Joint power and antenna selection optimization in large cloud radio access networks // *IEEE Transactions on Signal Processing*. — 2014. — Vol. 62, no. 5. — Pp. 1319–1328.
43. ONF White Paper. — Software-Defined Networking: The New Norm for Networks, 2012. — April.
44. Raychaudhuri D., Mandayam Narayan B. Frontiers of Wireless and Mobile Communications // *Proceedings of the IEEE*. — 2012. — Vol. 100, no. 4. — Pp. 824 – 840.
45. *GSMA Intelligence*. Global Mobile Economy Report 2015. — 2015.

46. Scenarios for 5G mobile and wireless communications: the vision of the METIS project / Afif Osseiran, Federico Boccardi, Volker Braun et al. // *IEEE Communications Magazine*. — 2014. — Vol. 52, no. 5. — Pp. 26–35. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6815890>.
47. *Hossain Ekram, Hasan Monowar*. 5G cellular: key enabling technologies and research challenges // *IEEE Instrumentation & Measurement Magazine*. — 2015. — Vol. 18, no. 3. — Pp. 11–21.
48. 5G radio access / Erik Dahlman, Gunnar Mildh, Stefan Parkvall et al. // *Ericsson Review*. — 2014. — Vol. June. — Pp. 2–7. http://sixtysix.wirelab.ericsson.net/res/thecompany/docs/publications/ericsson_review/2014/er-5g-radio-access.pdf.
49. Balancing spectral efficiency, energy consumption, and fairness in future heterogeneous wireless systems with reconfigurable devices / R. Amir, J. Martin, J. Deaton et al. // *IEEE Journal on Selected Areas in Communications*. — 2013. — Vol. 31, no. 5. — Pp. 969 – 980.
50. Enhanced intercell interference coordination challenges in heterogeneous networks / David Lopez-Perez, Ismail Guvenc, Guillaume De la Roche et al. // *IEEE Wireless Communications*. — 2011. — Vol. 18, no. 3.
51. *Cheung Wang Chi, Quek Tony Q. S., Kountouris Marios*. Throughput Optimization, Spectrum Allocation, and Access Control in Two-Tier Femtocell Networks // *IEEE Journal on Selected Areas in Communications*. — 2012. — Vol. 30, no. 3. — Pp. 561–574. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6171997<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6171997>.
52. *Ni Wei, Collings Iain*. A New Adaptive Small-Cell Architecture // *IEEE Journal on Selected Areas in Communications*. — 2013. — Vol. 31, no. 5. — Pp. 829–839. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6502478>.

53. Trends in small cell enhancements in LTE Advanced / Takehiro Nakamura, Satoshi Nagata, Anass Benjebbour et al. // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 2. — Pp. 98–105. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6461192<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6461192>.
54. Characterizing Performance of Load-Aware Network Selection in Multi-Radio (WiFi / LTE) Heterogeneous Networks / Mikhail Gerasimenko, Nageen Himayat, Shu-ping Yeh et al. // GLOBECOM Workshops (GC Wkshps). — 2013.
55. Utility-based radio link assignment in multi-radio heterogeneous networks / Ali Y. Panah, Shu-Ping Yeh, Nageen Himayat, Shilpa Talwar // Proc. of International Workshop on Emerging Technologies for LTE-Advanced and Beyond-4G on IEEE Globecom. — 2012. — Pp. 618–623. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6477645>.
56. *VNI Cisco*. Global Mobile Data Traffic Forecast 2016–2021. — White Paper. — 2017.
57. Direct communication between terminals in infrastructure based networks / JJ J Lehtomäki, Johanna Vartiainen, Mehdi Bennis et al. // ICT-MobileSummit 2008. — 2008. — Pp. 1–8. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Direct+Communication+Between+Terminals+in+Infrastructure+Based+Networks#0>.
58. Design aspects of network assisted device-to-device communications / Gábor Fodor, Erik Dahlman, Gunnar Mildh et al. // *IEEE Communications Magazine*. — 2012. — Vol. 50, no. 3.
59. Energy efficient communications for future broadband cellular networks / Sergey Andreev, Pavel Gonchukov, Nageen Himayat et al. // *Computer Communications*. — 2012. — Vol. 35, no. 14. — Pp. 1662–1671.

60. Operator controlled device-to-device communications in LTE-advanced networks / Lei Lei, Zhangdui Zhong, Chuang Lin, Xuemin Shen // *IEEE Wireless Communications*. — 2012. — Vol. 19, no. 3.
61. *Andersson Karl, Ahlund Christer*. Optimized Access Network Selection in a Combined WLAN/LTE Environment // *Wireless Personal Communications*. — 2011. — Vol. 61, no. 4. — Pp. 739–751.
62. *Hongseok Kim, de Veciana Gustavo*. Leveraging Dynamic Spare Capacity in Wireless Systems to Conserve Mobile Terminals' Energy // *IEEE/ACM Transactions on Networking*. — 2010. — Vol. 18, no. 3. — Pp. 802–815. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5339109>.
63. *Singh Sarabjot, Andrews Jeffrey*. Joint Resource Partitioning and Offloading in Heterogeneous Cellular Networks // *IEEE Transactions on Wireless Communications*. — 2014. — Vol. 13, no. 2. — Pp. 888–901. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6684550>.
64. Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks // *3GPP Technical specification (TS) 24.302*. — 2013.
65. A SyMPHOnY of Integrated IoT Businesses: Closing the Gap between Availability and Adoption / Jiri Hosek, Pavel Masek, Sergey Andreev et al. // *IEEE Communications Magazine*. — 2017. — Vol. 55, no. 12. — Pp. 156–164.
66. *Yeh SP, Talwar S, Wu G*. Capacity and coverage enhancement in heterogeneous networks // *Wireless Communications, IEEE*. — 2011. — Vol. 18, no. June. — Pp. 32–38. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5876498.
67. *Pentikousis K*. In search of energy-efficient mobile networking // *IEEE Communications Magazine*. — 2010. — Vol. 48, no. 1. — Pp. 95–103.

68. The evolution to 4G cellular systems: LTE-Advanced / I. Akyildiz, D. Gutierrez-Estevez, E. Reyes et al. // *Physical Communication*. — 2010. — Vol. 3, no. 4. — Pp. 217–244.
69. Multi-radio coexistence: challenges and opportunities / J. Zhu, A. Waltho, X. Yang, X. Guo // Proc. of the 16th International Conference on Computer Communications and Networks (ICCCN). — 2007. — Pp. 358–364.
70. White paper: Power consumption & energy efficiency. — 2003.
71. Miao Guowang, Himayat Nageen, Li Geoffrey Ye. Energy-efficient link adaptation in frequency-selective channels // *IEEE Transactions on communications*. — 2010. — Vol. 58, no. 2.
72. Raychaudhuri Dipankar, Mandayam Narayan B. Frontiers of wireless and mobile communications // *Proceedings of the IEEE*. — 2012. — Vol. 100, no. 4. — Pp. 824–840.
73. Song Guocong, Li Ye. Asymptotic throughput analysis for channel-aware scheduling // *IEEE Transactions on Communications*. — 2006. — Vol. 54, no. 10. — Pp. 1827–1834.
74. Lin X., Shroff N., Srikant R. A Tutorial on Cross-Layer Optimization in Wireless Networks // *IEEE Journal on Selected Areas in Communications*. — 2006. — Vol. 24, no. 8. — Pp. 1452–1463.
75. Benini Luca, Bogliolo Alessandro, De Micheli Giovanni. A survey of design techniques for system-level dynamic power management // *IEEE transactions on very large scale integration (VLSI) systems*. — 2000. — Vol. 8, no. 3. — Pp. 299–316.
76. Rodoplu V., Meng T. Bits-per-Joule Capacity of Energy-limited Wireless Networks // *IEEE Transactions on Wireless Communications*. — 2007. — Vol. 6, no. 3. — Pp. 857–865.
77. Cross-layer optimization for energy-efficient wireless communications: a survey / G. Miao, N. Himayat, G. Li, A. Swami // *Journal on Wireless*

- Communications and Mobile Computing.* — 2009. — Vol. 9, no. 4. — Pp. 529–542.
78. *Song G.* Cross-Layer Optimization for Spectral and Energy Efficiency: Ph.D. thesis / School of Electrical and Computer Engineering, Georgia Institute of Technology. — 2005.
 79. *Song G., Li G.* Cross-layer optimization for OFDM wireless networks - part II: algorithm development // *IEEE Transactions on Wireless Communications.* — 2005. — Vol. 4, no. 2. — Pp. 625–634.
 80. Interference-aware energy-efficient power optimization / G. Miao, N. Himayat, G. Li et al. // Proc. of the IEEE International Conference on Communications (ICC). — 2009.
 81. *Verdu S.* Spectral efficiency in the wideband regime // *IEEE Transactions on Information Theory.* — 2002. — Vol. 48, no. 6. — Pp. 1319–1343.
 82. An energy-efficient approach to power control and receiver design in wireless networks / F. Meshkati, H. Poor, S. Schwartz, N. Mandayam // *IEEE Transactions on Communications.* — 2005. — Vol. 53, no. 11. — Pp. 1885–1894.
 83. Energy efficient modulation and MAC for asymmetric RF microsensor systems / A.Y. Wang, SeongHwan Cho, C.G. Sodini, A.P. Chandrakasan // ISLPED'01: Proceedings of the 2001 International Symposium on Low Power Electronics and Design (IEEE Cat. No.01TH8581). — ACM, 2001. — Pp. 106–111. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=945384>.
 84. *Cui S., Goldsmith A., Bahai A.* Energy-constrained modulation optimization // *IEEE Transactions on Wireless Communications.* — 2005. — Vol. 4, no. 5. — Pp. 2349–2360.
 85. *Kim Hongseok.* Exploring tradeoffs in wireless networks under flow-level traffic: Energy, capacity and QoS. — The University of Texas at Austin, 2009.

86. *Grover P., Woyach K., Sahai A.* Towards a Communication-Theoretic Understanding of System-Level Power Consumption // *IEEE Journal on Selected Areas in Communications*. — 2011. — Vol. 29, no. 8. — Pp. 1744–1755.
87. Issues in integrating cellular networks WLANs, and MANETs: a futuristic heterogeneous wireless network / D. Cavalcanti, D. Agrawal, C. Cordeiro et al. // *IEEE Wireless Communications*. — 2005. — Vol. 12, no. 3. — Pp. 30–41.
88. *Walke B., Mangold S., Berlemann L.* IEEE 802 Wireless Systems: Protocols, Multi-Hop Mesh/Relaying, Performance and Spectrum Coexistence. — Wiley, 2007.
89. Multi-radio coexistence and collaboration on an SDR platform / Tommi Zetterman, Antti Piipponen, Kalle Raiskila, Sverre Slotte // *Analog Integrated Circuits and Signal Processing*. — 2011. — Vol. 69, no. 2-3. — Pp. 329–339. <http://link.springer.com/10.1007/s10470-011-9713-7>.
90. *L. Wang, G.-S. Kuo.* Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks - A Tutorial // *IEEE Communications Surveys & Tutorials*. — 2013. — Vol. 15. — Pp. 271–292.
91. *Andreev Sergey, Galinina Olga, Koucheryavy Yevgeni.* Energy-efficient client relay scheme for machine-to-machine communication // Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE / IEEE. — 2011. — Pp. 1–5.
92. RAT selection games in HetNets / Ehsan Aryafar, Alireza Keshavarz-Haddad, Michael Wang, Mung Chiang // *Proc. of IEEE INFOCOM*. — 2013. — Pp. 1–11.
93. Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks / Mikhail Gerasimenko, Dmitri Moltchanov, Sergey Andreev et al. // *IEEE Access*. — 2017. — Vol. 5. — Pp. 219–235.

94. *Zafeiris V., Giakoumakis E.* Optimized traffic flow assignment in multi-homed, multi-radio mobile hosts // *Computer Networks*. — 2011. — Vol. 55, no. 5. — Pp. 1114–1131.
95. *Zangwill W.* Nonlinear programming: a unified approach. — Englewood Cliffs, N.J : Prentice-Hall, 1969.
96. Optimizing energy efficiency of a multi-radio mobile device in heterogeneous beyond-4G networks / Olga Galinina, Sergey Andreev, Andrey Turlikov, Yevgeni Koucheryavy // *Performance Evaluation*. — 2014. — Vol. 78. — Pp. 18–41. <http://dx.doi.org/10.1016/j.peva.2014.06.002>.
97. Random-access latency optimization and stability of highly-populated LTE-based M2M deployments / Olga Galinina, Andrey Turlikov, Tuomas Tirronen et al. // *Communications (ICC), 2016 IEEE International Conference on / IEEE*. — 2016. — Pp. 1–6.
98. *Andreev Sergey, Galinina Olga, Koucheryavy Yevgeni.* Energy-Efficient Client Relay Scheme for Machine-to-Machine Communication // *Proceedings of GLOBECOM*. — 2011.
99. *Galinina Olga, Andreev Sergey, Koucheryavy Yevgeni.* Performance analysis of client relay cloud in wireless cellular networks // *Proc. of International Conference on Wired/Wireless Internet Communications / Springer*. — 2012. — Pp. 40–51.
100. On Lambert's W function: Tech. Rep. CS-93-03 / R. Corless, G. Gonnet, D. Hare, D. Jeffrey: University of Waterloo, 1993.
101. An overview of load balancing in HetNets: old myths and open problems / Jeffrey G. Andrews, Sarabjot Singh, Qiaoyang Ye et al. // *IEEE Wireless Communications*. — 2014. — Vol. 21. — Pp. 18–25.
102. *Пяттаев А.В., Кучерявый Е.А., Андреев С.Д.* Методы оценки характеристик сетей связи с клиентской ретрансляцией // *Информационные технологии моделирования и управления*. — 2014. — Vol. 3, no. 87. — Pp. 247–255.

103. LTE Release 10 & beyond (LTE-Advanced).
104. *Van Der Meulen Edward C.* Three-Terminal Communication Channels // *Advances in Applied Probability*. — 1971. — Vol. 3. — Pp. 120–154.
105. *Cover Thomas M., El Gamal Abbas A.* Capacity Theorems for the Relay Channel // *IEEE Transactions on Information Theory*. — 1979. — Vol. 25(5). — Pp. 572–584.
106. *Kramer G., Gastpar M., Gupta P.* Cooperative Strategies and Capacity Theorems for Relay Networks // *IEEE Transactions on Information Theory*. — 2005. — Vol. 51(9). — Pp. 3037–3063.
107. *Laneman J.N., Tse D.N.C., Wornell G.W.* Cooperative diversity in wireless networks: Efficient protocols and outage behavior // *IEEE Transactions on Information Theory*. — 2004. — Vol. 50(12). — Pp. 3062–3080.
108. *Sendonaris A., Erkip E., Aazhang B.* User cooperation diversity. Part I, II // *IEEE Transactions on Communications*. — 2003. — Vol. 51(11). — Pp. 1927–1938.
109. *Nosratinia A., Hunter T.E., Hedayat A.* Cooperative communication in wireless networks // *IEEE Communications Magazine*. — 2004. — Vol. 42 (10). — Pp. 74 – 80.
110. *Nosratinia Aria, Hunter Todd E.* Grouping and partner selection in cooperative wireless networks // *IEEE Journal on Selected Areas in Communications*. — 2007. — Vol. 25, no. 2.
111. *Tannious Ramy, Nosratinia Aria.* Spectrally-efficient relay selection with limited feedback // *IEEE Journal on Selected Areas in Communications*. — 2008. — Vol. 26, no. 8.
112. IEEE Std 802.16m (D9), Amendment to IEEE Standard for Local and metropolitan area networks. Advanced Air Interface.

113. Active-Mode Power Optimization in OFDMA-Based Wireless Networks / S. Andreev, Y. Koucheryavy, N. Himayat et al. // Proc. of the IEEE BWA Workshop of Globecom. — 2010.
114. *Schurgers C.* Energy-Aware Wireless Communications: Ph.D. thesis / University of California, Los Angeles. — 2002.
115. *Miao G.* Cross-Layer Optimization for Spectral and Energy Efficiency: Ph.D. thesis / School of Electrical and Computer Engineering, Georgia Institute of Technology. — 2008.
116. *Kim H.* Exploring Tradeoffs in Wireless Networks under Flow-Level Traffic: Energy, Capacity and QoS: Ph.D. thesis / University of Texas at Austin. — 2009.
117. *Rong Beiyu, Ephremides Anthony.* On Opportunistic Cooperation for Improving the Stability Region with Multipacket Reception // Third Euro-NF Conference, NET-COOP. — Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2009. — Pp. 45–59.
118. *Cui Shuguang, Goldsmith Andrea J, Bahai Ahmad.* Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks // *IEEE Journal on selected areas in communications.* — 2004. — Vol. 22, no. 6. — Pp. 1089–1098.
119. *Andreev Sergey, Galinina Olga, Vinel Alexey.* Performance Evaluation of a Three Node Client Relay System // *International Journal of Wireless Networks and Broadband Technologies, IJWNBT.* — 2011. — Vol. 1(1). — Pp. 73–84.
120. IEEE Std 802.16j-2009, IEEE Standard for Local and metropolitan area networks - Multihop Relay Specification.
121. *Stuber G. L.* Principles of Mobile Communication. — Norwell, MA: Kluwer Academic Publishers, 2001. — P. 776.
122. *Кучерявый А.Е., Парамонов А.И., Кучерявый Е.А.* Сети связи общего пользования. — ФГУП ЦНИИС, ISBN: 5-00-000831-6., 2008.

123. *Andreev S., Galinina O., Turlikov A.* Basic Client Relay Model for Wireless Cellular Networks // Proc. of the ICUMT Congress. — 2010.
124. *Пяттаев Александр Владиславович, Кучерявый Евгений Андреевич, Андреев Сергей Дмитриевич.* Подходы к моделированию сетей связи с клиентской ретрансляцией // *Системы управления и информационные технологии.* — 2013. — Т. 2, № 52. — С. 98–102.
125. Some modeling approaches for client relay networks / Alexander Pyattaev, Sergey Andreev, Yevgeni Koucheryavy, Dmitri Moltchanov // Computer Aided Modeling, Analysis and Design of Communication Links and Networks (CAMAD), 2010 15th IEEE International Workshop on / IEEE. — 2010. — Pp. 116–120.
126. *Cho H., Puthenkulam J.* — Machine to Machine (M2M) Communication Study Report, IEEE 802.16ppc-10/0002r6, 2010. — May.
127. Machine-To-Machine (M2M) & Smart Systems Forecast 2010-2014. — 2009.
128. 3GPP Technical Report. — System Improvements for Machine-Type Communications (Release 10), TR 23.888, 2010. — July.
129. *Jaiswal N.* Priority Queues. — Academic Press, New York, 1968. — P. 240.
130. *Kleinrock L.* Queueing Systems: Volume 1. Theory. — John Wiley & Sons, 1975. — P. 417.
131. *Himayat N., Talwar S., Johnsson K. et al.* — Proposed IEEE 802.16p Performance Requirements for Network Entry by Large Number of Devices, IEEE 802.16p-10/0006, 2010. — November.
132. QUALCOMM Incorporated. — Simulation Assumptions for MTC and RACH Load Simulation Results for UMTS, R2-105619, 2010. — October.
133. *Norros I.* A storage model with self-similar input // *Queueing Systems.* — 1994. — Vol. 16, no. 3-4. — Pp. 387–396.

134. *Taqqu M. S., Teverovsky V., Willinger W.* Estimators for Long-Range Dependence: An Empirical Study // *Fractals*. — 1995. — Vol. 3. — Pp. 785–798.
135. *Bianchi G.* Performance analysis of the IEEE 802.11 distributed coordination function // *IEEE Journal on Selected Areas on Communications*. — 2000. — Vol. 18. — Pp. 535–547.
136. Client Relay Simulation Model for Centralized Wireless Networks / A. Pyattaev, S. Andreev, A. Vinel, B. Sokolov // Proc. of the EUROSIM Congress. — 2010.
137. Performance of the IEEE 802.16e Sleep Mode Mechanism in the Presence of Bidirectional Traffic / K. De Turck, S. Andreev, S. De Vuyst et al. // Proc. of the IEEE ICC Conference. — 2009.
138. *Himayat N., Johnsson K., Talwar S., Wang X.* — Functional Requirements for Network Entry and Random Access by Large Number of Devices, IEEE 802.16ppc-10/0049r1, 2010. — August.
139. Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets / Supratim Deb, Pantelis Monogioudis, Jerzy Miernik, James P Seymour // *IEEE/ACM transactions on networking*. — 2014. — Vol. 22, no. 1. — Pp. 137–150.
140. *ElSawy Hesham, Hossain Ekram.* Two-tier HetNets with cognitive femtocells: Downlink performance modeling and analysis in a multichannel environment // *IEEE Transactions on Mobile Computing*. — 2014. — Vol. 13, no. 3. — Pp. 649–663.
141. *Hwang Insoo, Song Bongyong, Soliman Samir S.* A holistic view on hyperdense heterogeneous and small cell networks // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 6. — Pp. 20–27.
142. *Кучерявый А. Е., Прокопьев А. В., Кучерявый Е. А.* Самоорганизующиеся сети. — 2011. — P. 312.

143. On the integration of cooperative relaying into the winner system concept / Klaus Doppler, Afif Osseiran, Michal Wodczak, Peter Rost // Mobile and Wireless Communications Summit, 2007. 16th IST / IEEE. — 2007. — Pp. 1–5.
144. Dynamic Resource Assignment and Cooperative Relaying in Cellular Networks: Concept and Performance Assessment / Klaus Doppler, Simone Redana, Michał Wódczak et al. // *EURASIP Journal on Wireless Communications and Networking*. — 2009. — Vol. 2009, no. 1. — Pp. 1–14. <http://jwcn.urasipjournals.com/content/2009/1/475281>.
145. System-level evaluation of opportunistic client cooperation in wireless cellular networks / Alexander Pyattaev, Sergey Andreev, Olga Galinina, Yevgeni Koucheryavy // Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on / IEEE. — 2011. — Pp. 1–6.
146. *Pyattaev Alexander*. System Level Performance Evaluation of Client Cooperation in Wireless Cellular Networks: Master of Science Thesis / Tampere University of Technology. — 2011. — P. 74. <http://urn.fi/URN:NBN:fi:tty-2011122214964>.
147. Device-to-device communication as an underlay to LTE-advanced networks / K Doppler, M Rinne, C Wijting et al. // *Communications Magazine, IEEE*. — 2009. — no. 12. — Pp. 42–49. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5350367.
148. FlashLinQ: A Synchronous Distributed Scheduler for Peer-to-Peer Ad Hoc Networks / Xinzhou Wu, Saurabh Tavildar, Sanjay Shakkottai et al. // Proc. of the 48th Annual Allerton Conference on Communication, Control, and Computing. — 2010. — Pp. 514–521. </home/headhunter/documents/papers/mac/flashlinq.pdf>.
149. Toward proximity-aware internetworking / M Scott Corson, Rajiv Laroia, Junyi Li et al. // *IEEE Wireless Communications*. — 2010. — Vol. 17, no. 6.

150. *Vigato Alberto, Vangelista Lorenzo*. Joint discovery in synchronous wireless networks // *IEEE Transactions on Communications*. — 2011. — Vol. 59, no. 8. — Pp. 2296–2305. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5928643.
151. *Belleschi M, Fodor G, Abrardo A*. Performance analysis of a distributed resource allocation scheme for D2D communications // In Proc. of GLOBECOM Workshops. — 2011. — Pp. 358–362.
152. Device-to-Device Communication Underlying Cellular Communications Systems / Pekka Janis, Chia-Hao Yu, Klaus Doppler et al. // *International Journal of Communications, Network and System Sciences*. — 2009. — Vol. 02, no. 03. — Pp. 169–178. <http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/ijcns.2009.23019>.
153. *Yu Chia-Hao, Tirkkonen Olav*. Device-to-device underlay cellular network based on rate splitting // Wireless Communications and Networking Conference (WCNC), 2012 IEEE / IEEE. — 2012. — Pp. 262–266.
154. *Kaufman Brett, Lilleberg Jorma, Aazhang B*. Spectrum Sharing Scheme Between Cellular Users and Ad-hoc Device-to-Device Users // *IEEE Transactions on Wireless Communications*. — 2013. — Vol. 12, no. 3. — Pp. 1038–1049. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6415102.
155. Power optimization of device-to-device communication underlying cellular communication / C-H Yu, Olav Tirkkonen, Klaus Doppler, Cássio Ribeiro // *Communications, 2009. ICC'09. IEEE International Conference on* / IEEE. — 2009. — Pp. 1–5.
156. Interference and sinr in millimeter wave and terahertz communication systems with blocking and directional antennas / Vitaly Petrov, Mikhail Komarov, Dmitri Moltchanov et al. // *IEEE Transactions on Wireless Communications*. — 2017. — Vol. 16, no. 3. — Pp. 1791–1808.
157. *Ефимушкина Т.В., Молчанов Д.А., Кучерявый Е.А.* Исследование вероятностно временных характеристик функционирования соты

- WIMAX с несколькими режимами модуляции и эластичным трафиком данных // *T-Comm — Телекоммуникации и транспорт*. — 2010. — Vol. 4, no. 7. — Pp. 203–204.
158. Resource sharing optimization for device-to-device communication underlying cellular networks / Chia-Hao Yu, Klaus Doppler, Cassio B Ribeiro, Olav Tirkkonen // *IEEE Transactions on Wireless communications*. — 2011. — Vol. 10, no. 8. — Pp. 2752–2763.
159. Network controlled device-to-device (D2D) and cluster multicast concept for LTE and LTE-A networks / Jaakko Seppälä, Timo Koskela, Tao Chen, Sami Hakola // *Wireless Communications and Networking Conference (WCNC), 2011 IEEE / IEEE*. — 2011. — Pp. 986–991.
160. *Min H, Seo W, Lee J*. Reliability improvement using receive mode selection in the device-to-device uplink period underlying cellular networks // *IEEE Transactions on Wireless Communications*. — 2011. — Vol. 10, no. 2. — Pp. 413–418. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5675650.
161. *Chen Tao, Charbit Gilles, Hakola Sami*. Time Hopping for Device-To-Device Communication in LTE Cellular System // *Wireless Communications and Networking Conference (IEEE WCNC)*. — 2010.
162. Dynamic resource allocation for Device-to-Device communication underlying cellular networks / Yanfang Xu, Rui Yin, Tao Han, Guanding Yu // *International Journal of Communication Systems*. — 2012. — no. 38. <http://onlinelibrary.wiley.com/doi/10.1002/dac.2485/full>.
163. *Jung Minchae, Hwang Kyuho, Choi Sooyong*. Joint Mode Selection and Power Allocation Scheme for Power-Efficient Device-to-Device (D2D) Communication // *IEEE Vehicular Technology Conference*. — 2012. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6240196.
164. *Mukherjee Amitav, Hottinen Ari*. Energy-efficient device-to-device MIMO underlay network with interference constraints // *Smart Antennas (WSA), 2012 International ITG Workshop on / IEEE*. — 2012. — Pp. 105–109.

165. Advances in Device-to-Device Communications and Network Coding for IMT-Advanced / Afif Osseiran, Klaus Doppler, Cassio Ribeiro et al. // ICT-Mobile Summit 2009 Conference Proceedings. — 2009.
166. Performance impact of fading interference to device-to-device communication underlying cellular networks / Chia-Hao Yu, Klaus Doppler, Cássio Ribeiro, Olav Tirkkonen // Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on / IEEE. — 2009. — Pp. 858–862.
167. Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution / Negin Golrezaei, Andreas F Molisch, Alexandros G Dimakis, Giuseppe Caire // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 4. — Pp. 142–149.
168. *Direct LTE*. The Case for Device-to-Device Proximate Discovery // *Qualcomm Research, San Diego, CA, USA, Tech. Rep.* — 2013. — Vol. 92121.
169. Delivering Fairness and QoS Guarantees for LTE/Wi-Fi Coexistence Under LAA Operation / Massimiliano Maule, Dmitri Moltchanov, Pavel Kustarev et al. // *IEEE Access*. — 2018. — Vol. 6. — Pp. 7359–7373.
170. *Al-Kanj Lina, Dawy Zaher, Yaacoub Elias*. Energy-aware cooperative content distribution over wireless networks: Design alternatives and implementation aspects // *IEEE Communications Surveys & Tutorials*. — 2013. — Vol. 15, no. 4. — Pp. 1736–1760.
171. *Sankaran CB*. Data offloading techniques in 3GPP Rel-10 networks: A tutorial // *IEEE Communications Magazine*. — 2012. — Vol. 50, no. 6.
172. *Xing Bo, Seada Karim, Venkatasubramanian Nalini*. An experimental study on wi-fi ad-hoc mode for mobile device-to-device video delivery // INFOCOM Workshops 2009, IEEE / IEEE. — 2009. — Pp. 1–6.
173. WiFi Alliance, Wi-Fi Peer-to-Peer (P2P) Specifications, v1.2.

174. LTE for public safety networks / Tewfik Doumi, Mike F Dolan, Said Tatesh et al. // *IEEE Communications Magazine*. — 2013. — Vol. 51, no. 2. — Pp. 106–112.
175. Direct connection on the move: Characterization of user mobility in cellular-assisted D2D systems / Antonino Orsino, Dmitri Moltchanov, Margarita Gapeyenko et al. // *IEEE Vehicular Technology Magazine*. — 2016. — Vol. 11, no. 3. — Pp. 38–48.
176. Assisted handover based on device-to-device communications in 3GPP LTE systems / Antonino Orsino, Margarita Gapeyenko, Leonardo Militano et al. // Globecom Workshops (GC Wkshps), 2015 IEEE / IEEE. — 2015. — Pp. 1–6.
177. *Mathai A.-M.* An introduction to geometrical probability: distributional aspects with applications. — 1st edition. — CRC Press, 1999.
178. *Петров В.И., Молчанов Д.А., Кучерявый Е.А.* Оценка отношения сигнал/помеха в беспроводных сетях связи терагерцового диапазона частот // *Электросвязь*. — 2017. — no. 10. — Pp. 24–29.
179. Random triangle: A baseline model for interference analysis in heterogeneous networks / Andrey Samuylov, Dmitri Moltchanov, Yuliya Gaidamaka et al. // *IEEE Transactions on Vehicular Technology*. — 2016. — Vol. 65, no. 8. — Pp. 6778–6782.
180. Vibrational relaxation of pure liquid water / J. Lindnen, P. Vohringer, M. Pshenichnikov et al. // *Chem. Phys. Let.* — 2006. — Vol. 421. — Pp. 329–333.
181. *Mood A., Graybill F., Boes D.* Introduction to the theory of statistics. — 1st edition. — McGraw-Hill, 1974.
182. *Geim A., Novoselov K.* The rise of graphene // *Nature Materials*. — 2007. — Vol. 6. — Pp. 183–191.
183. Three node client relay system with packet retry limit / Sergey Andreev, Olga Galinina, Yevgeni Koucheryavy, Alexandra Lokhanova // Proc. of

- 16th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) / IEEE. — 2011. — Pp. 21–25.
184. Wang H, Chu X. Distance-constrained resource-sharing criteria for device-to-device communications underlaying cellular networks // *Electronics letters*. — 2012. — Vol. 48, no. 9. — Pp. 528–530.
185. Time-Dependent Energy and Resource Management in Mobility-Aware D2D-Empowered 5G Systems / Antonino Orsino, Andrey Samuylov, Dmitri Moltchanov et al. // *IEEE Wireless Communications*. — 2017. — Vol. 24, no. 4. — Pp. 14–22.
186. Dynamic power control mechanism for interference coordination of device-to-device communication in cellular networks / Jaheon Gu, Sueng Jae Bae, Bum-Gon Choi, Min Young Chung // In Proc. of Third International Conference on Ubiquitous and Future Networks (ICUFN). — 2011.
187. Xu Z., Dong X., Bornemann J. Design of a Reconfigurable MIMO System for THz Communications Based on Graphene Antennas // *IEEE Transactions on Terahertz Science and Technology*. — 2014. — Vol. 4, no. 5. — Pp. 609–617.
188. Li Qian Clara, Wu Geng, Hu Rose Qingyang. Analytical study on network spectrum efficiency of ultra dense networks // *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. — 2013. — Pp. 2764–2768. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6666617>.
189. Gotsis Antonis G, Stefanatos Stelios, Alexiou Angeliki. Spatial coordination strategies in future ultra-dense wireless networks // 2014 11th International Symposium on Wireless Communications Systems (ISWCS). — IEEE, 2014. — Pp. 801–807. <http://arxiv.org/abs/1405.2576><http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6933463>.

190. Cooperative Radio Resource Management in Heterogeneous Cloud Radio Access Networks / M. Gerasimenko, D. Moltchanov, R. Florea et al. // *IEEE Access*. — 2015. — Vol. 3. — Pp. 397–406.
191. *Moltchanov Dmitri, Koucheryavy Yevgeni*. Cross-layer modeling of wireless channels: An overview of basic principles // *Wireless personal communications*. — 2014. — Vol. 74, no. 1. — Pp. 23–44.
192. Evolved Universal Terrestrial Radio Access (E-UTRA). — 3GPP Technical Report (TR) 36.814-900, 2010.
193. On the Performance of Device-to-Device Underlay Communication with Simple Power Control / Chia-Hao Yu, Olav Tirkkonen, Klaus Doppler, Cassio Ribeiro // *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*. — 2009. — Pp. 1–5. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5073734>.
194. *Asadi Arash, Mancuso Vincenzo*. On the compound impact of opportunistic scheduling and D2D communications in cellular networks // Proc. of 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems / ACM. — 2013. — Pp. 279–288.
195. Social-aware D2D communications: Qualitative insights and quantitative analysis / Yong Li, Ting Wu, Pan Hui et al. // *IEEE Communications Magazine*. — 2014. — Vol. 52, no. 6. — Pp. 150–158.
196. 3GPP LTE-Assisted Wi-Fi-Direct: Trial Implementation of Live D2D Technology / Alexander Pyattaev, Jiri Hosek, Kerstin Johnsson et al. // *Etri Journal*. — 2015. — Vol. 37, no. 5. — Pp. 877–887.
197. Network-assisted D2D communications: Implementing a technology prototype for cellular traffic offloading / Alexander Pyattaev, Kerstin Johnsson, Adam Surak et al. // Proc. of Wireless Communications and Networking Conference (WCNC) / IEEE. — 2014. — Pp. 3266–3271.
198. *Zhuang J., Jalloul L., Novak R., Park J*. IEEE 802.16m Evaluation Methodology Document (EMD). — 2009.

199. *IEEE*. Std 802.11-2012: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. — 2012.
200. *ITU-R M.2135*. Guidelines for evaluation of radio interface technologies for IMT-Advanced // *ITU*. — 2009.
201. *Konstantinou K, Kang S, Tzaras C*. A Measurement-Based Model for Mobile-to-Mobile UMTS Links // Vehicular Technology Conference. — 2007. </home/headhunter/documents/papers/phy/AMeasurement-BasedModelforMobile-to-MobileUMTSLinks.pdf>.
202. *Gudmundson M*. Correlation model for shadow fading in mobile radio systems // *Electronics Letters*. — 1991. — Vol. 27, no. 23. — Pp. 2145–2146.
203. Direct Connection on the Move: Characterization of User Mobility in Cellular-Assisted D2D Systems / A. Orsino, D. Moltchanov, M. Gapeyenko et al. // *IEEE Vehicular Technology Magazine*. — 2016. — Vol. 11, no. 3. — Pp. 38–48.
204. A unifying perspective on proximity-based cellular-assisted mobile social networking / Sergey Andreev, Jiri Hosek, Thomas Olsson et al. // *IEEE Communications Magazine*. — 2016. — Vol. 54, no. 4. — Pp. 108–116.
205. Toward trusted, social-aware D2D connectivity: bridging across the technology and sociality realms / Aleksandr Ometov, Antonino Orsino, Leonardo Militano et al. // *IEEE Wireless Communications*. — 2016. — Vol. 23, no. 4. — Pp. 103–111.
206. *Whitehead Steve*. Adopting wireless machine-to-machine technology // *Computing and Control Engineering*. — 2004. — Vol. 15, no. 5. — Pp. 40–46.
207. Performance Evaluation of IEEE 802.11ah Actuators / Behnam Badihi, Luis Felipe Del Carpio, Parth Amin et al. // Proc. of IEEE 83rd Vehicular Technology Conference (VTC Spring), 2016 / IEEE. — 2016. — Pp. 1–5.

208. Cognitive machine-to-machine communications: visions and potentials for the smart grid / Yan Zhang, Rong Yu, Maziar Nekovee et al. // *IEEE network*. — 2012. — Vol. 26, no. 3.
209. LTE-direct vs. WiFi-direct for machine-type communications over LTE-A systems / Massimo Condoluci, Leonardo Militano, Antonino Orsino et al. // Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on / IEEE. — 2015. — Pp. 2298–2302.
210. IEEE 802.11ah: the WiFi approach for M2M communications / Toni Adame, Albert Bel, Boris Bellalta et al. // *IEEE Wireless Communications*. — 2014. — Vol. 21, no. 6. — Pp. 144–152.
211. A survey on IEEE 802.11 ah: An enabling networking technology for smart cities / Evgeny Khorov, Andrey Lyakhov, Alexander Krotov, Andrey Guschin // *Computer Communications*. — 2015. — Vol. 58. — Pp. 53–69.
212. *Rysavy Peter*. Meeting mobile demand with a combination of spectrum alternatives // *IEEE Wireless Communications*. — 2014. — Vol. 21, no. 2. — Pp. 6–7.
213. *Park Minyoung*. IEEE 802.11 ah: sub-1-GHz license-exempt operation for the internet of things // *IEEE Communications Magazine*. — 2015. — Vol. 53, no. 9. — Pp. 145–151.
214. On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios / Margarita Gapeyenko, Andrey Samuylov, Mikhail Gerashenko et al. // *IEEE Transactions on Vehicular Technology*. — 2017.
215. *Омётов А.Я., Кучерявый Е.А., Андреев С.Д.* О роли беспроводных технологий связи в развитии «Интернета вещей» // *Информационные технологии и телекоммуникации*. — 2014. — Vol. 3, no. 7. — Pp. 31–40.

216. Smart utility networks in TV white space / Chin-Sean Sum, Hiroshi Harada, Fumihide Kojima et al. // *IEEE Communications Magazine*. — 2011. — Vol. 49, no. 7.
217. *Aust Stefan, Prasad R Venkatesha, Niemegeers Ignas GMM*. A framework for massive access and radio resource management in urban WLANs // Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on / IEEE. — 2013. — Pp. 93–99.
218. Upper bound on capacity of 5G mmWave cellular with multi-connectivity capabilities / D Moltchanov, A Ometov, S Andreev, Y Koucheryavy // *Electronics Letters*. — 2018. — Vol. 54, no. 11. — Pp. 724–726.
219. *Aust Stefan, Ito Tetsuya*. Sub 1GHz wireless LAN deployment scenarios and design implications in rural areas // GLOBECOM Workshops (GC Wkshps), 2011 IEEE / IEEE. — 2011. — Pp. 1045–1049.
220. *Aust Stefan, Prasad R Venkatesha, Niemegeers Ignas GMM*. IEEE 802.11ah: Advantages in standards and further challenges for sub 1 GHz Wi-Fi // Communications (ICC), 2012 IEEE International Conference on / IEEE. — 2012. — Pp. 6885–6889.
221. *Aust Stefan, Prasad Venkatesha*. Analysis of the Performance Boundaries of Sub-1 GHz WLANs in the 920 MHz ISM-Band // Proc. of Tenth International Symposium on Wireless Communication Systems (ISWCS 2013). — 2013. — Pp. 1–5.
222. *Hazmi Ali, Rinne Jukka, Valkama Mikko*. Feasibility study of IEEE 802.11ah radio technology for IoT and M2M use cases // Proc. of IEEE Globecom Workshops / IEEE. — 2012. — Pp. 1687–1692.
223. Performance comparison between slotted IEEE 802.15.4 and IEEE 802.11ah in IoT based applications / Behnam Badihi Olyaei, Juho Pirkkanen, Orod Raeesi et al. // Proc. of IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) / IEEE. — 2013. — Pp. 332–337.

224. Capacity analysis of IEEE 802.11ah WLANs for M2M communications / Toni Adame, Albert Bel, Boris Bellalta et al. // Proc. of International Workshop on Multiple Access Communications / Springer. — 2013. — Pp. 139–155.
225. Исследование энергоэффективности технологии IEEE 802.11ah для передачи данных между автономными устройствами в контексте «Интернета вещей» / В.И. Петров, А.Я. Омётов, С.Д. Андреев, Е.А. Кучерявый // *Информационные технологии и телекоммуникации*. — 2016. — Vol. 4, no. 1. — Pp. 1–8.
226. *Sun Weiping, Choi Munhwan, Choi Sunghyun*. IEEE 802.11ah: A long range 802.11 WLAN at sub 1 GHz // *Journal of ICT Standardization*. — 2013. — Vol. 1, no. 1. — Pp. 83–108.
227. Advances in IEEE 802.11ah Standardization for Machine-Type Communications in Sub-1GHz WLAN / Yuan Zhou, Haiguang Wang, Shoukang Zheng, Zander Zhongding Lei // Proc. of IEEE Second Workshop on Telecommunication Standards. — 2013. — Pp. 1269–1273.
228. *Khorov Evgeny, Krotov Alexander, Lyakhov Andrey*. Modelling machine type communication in IEEE 802.11ah networks // Proc. of International Conference on Communication Workshop (ICCW) / IEEE. — 2015. — Pp. 1149–1154.
229. Advanced Power Control Techniques for Interference Mitigation in Dense 802.11 Networks / Oghenekome Oteri, Pengfei Xia, Frank Lasita, Robert Olesen // Proc. of 16th International Symposium on Wireless Personal Multimedia Communications (WPMC). — 2013. — Pp. 1–7.
230. Carrier Sense Multiple Access with Collision Resolution / Hyun Ho Choi, Jung Min Moon, In Ho Lee, Howon Lee // *IEEE Communications Letters*. — 2013. — Vol. 17, no. 6. — Pp. 1284–1287. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6459482.

231. CAS-based Channel Access Protocol for IEEE 802.11ah WLANs / Albert Bel, Toni Adame, Boris Bellalta et al. // Proc. of European Wireless Conference (EW). — 2014. — Pp. 1–6.
232. Performance Analysis of Grouping Strategy for Dense IEEE 802.11 Networks / Lei Zheng, Lin Cai, Jianping Pan, Minming Ni // Proc. of IEEE Global Communications Conference (GLOBECOM). — 2013. — Pp. 219–224. http://www.ece.uvic.ca/~zhengl/materials/GC13ZLPN_slides.pdf.
233. Performance analysis of group-synchronized DCF for dense IEEE 802.11 networks / Lei Zheng, Minming Ni, Lin Cai et al. // *IEEE Transactions on Wireless Communications*. — 2014. — Vol. 13, no. 11. — Pp. 6180–6192.
234. *Ghosh Monisha, Lasita Frank*. Puncturing of CRC Codes for IEEE 802.11ah // Proc. of IEEE 78th Vehicular Technology Conference (VTC Fall). — 2013. — Pp. 2–6. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6692382.
235. *Lin Hsiang-ho*. DeepSleep: IEEE 802.11 Enhancement for Communications // Proc. of Global Communications Conference (GLOBECOM). — 2012. — Pp. 5231–5236.
236. *Liu Ren Ping, Sutton Gordon, Collings Iain*. Power Save with Offset Listen Interval for IEEE 802.11ah Smart Grid Communications // Proc. of IEEE International Conference on Communications (ICC). — 2013. — Pp. 4488–4492. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6655274.
237. *Glaropoulos Ioannis, Mangold Stefan, Vukadinovic Vladimir*. Enhanced IEEE 802.11 power saving for multi-hop toy-to-toy communication // Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing / IEEE. — 2013. — Pp. 603–610.
238. The Impact of Interference from the Side Lanes on mmWave/THz Band V2V Communication Systems with Directional Antennas / Vitaly Petrov,

- Joonas Kokkonen, Dmitri Moltchanov et al. // *IEEE Transactions on Vehicular Technology*. — 2018.
239. IEEE 802.11ah Based M2M Networks Employing Virtual Grouping and Power Saving Methods / Kohei Ogawa, Yuki Sangenya, Masahiro Morikura et al. // Proc. of IEEE 78th Vehicular Technology Conference (VTC Fall). — 2013. — Pp. 1–5. http://search.ieice.org/bin/summary.php?id=e96-b_12_2976.
240. Liu Ren Ping, Sutton Gordon, Collings Iain. WLAN Power Save with Offset Listen Interval for Machine-to-Machine Communications // *IEEE Transactions on Wireless Communications*. — 2014. — Vol. 13, no. 5. — Pp. 2552–2562. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6776596.
241. Bianchi Giuseppe, Tinnirello Ilenia. Remarks on IEEE 802.11 DCF performance analysis // *IEEE Communications Letters*. — 2005. — Vol. 9, no. 8. — Pp. 765–767.
242. Analysis of the IEEE 802.11e EDCA under statistical traffic / Juki Wirawan Tantra, Chuan Heng Foh, Ilenia Tinnirello, Giuseppe Bianchi // IEEE International Conference on Communications / IEEE. — Vol. 2. — 2006. — Pp. 546–551.
243. Chen Churong, Law Choi Look. Throughput performance analysis and experimental evaluation of IEEE 802.11b radio link // Proc. of AFIPS, Fall Joint Comput. Conference of 6th International Conference on Information, Communications & Signal Processing / IEEE. — 2007. — Pp. 1–5.
244. Unsaturated throughput analysis of IEEE 802.11 in presence of non ideal transmission channel and capture effects / Fred Daneshgaran, Massimiliano Laddomada, Fabio Mesiti, Marina Mondin // *IEEE Transactions on Wireless Communications*. — 2008. — Vol. 7, no. 4. — Pp. 1276–1286.
245. Duty Cycle Challenges of IEEE 802.11ah Networks in M2M and IoT Applications / Ali Hazmi, Luis Felipe Del Carpio, Ahmet Goekceoglu et al. //

- Proc. of 22th European Wireless Conference European Wireless / VDE.
— 2016. — Pp. 1–7.
246. An efficient spatial group restricted access window scheme for IEEE 802.11 ah networks / Mengxi Dong, Zhanji Wu, Xiang Gao, Huan Zhao // Proc. of 2016 Sixth International Conference on Information Science and Technology (ICIST) / IEEE. — 2016. — Pp. 168–173.
247. Mobility-aware energy-quality trade-off for video delivery in dense heterogeneous networks / Pasquale Scopelliti, Giuseppe Araniti, Gabriel-Miro Muntean, Antonio Iera // Proc. of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) / IEEE. — 2016. — Pp. 1–6.
248. *Jung Byoung Hoon, Song Nah-Oak, Sung Dan Keun.* A network-assisted user-centric WiFi-offloading model for maximizing per-user throughput in a heterogeneous network // *IEEE Transactions on Vehicular Technology.* — 2014. — Vol. 63, no. 4. — Pp. 1940–1945.
249. Device-to-device communications for national security and public safety / Gabor Fodor, Stefan Parkvall, Stefano Sorrentino et al. // *IEEE Access.* — 2014. — Vol. 2. — Pp. 1510–1520.
250. *Bianchi Giuseppe.* IEEE 802.11-saturation throughput analysis // *IEEE Communications Letters.* — 1998. — Vol. 2, no. 12. — Pp. 318–320.
251. *Fodor Gabor, Sorrentino Stefano, Sultana Shabnam.* Network assisted device-to-device communications: use cases, design approaches, and performance aspects // *Smart Device to Smart Device Communication.* — Springer, 2014. — Pp. 135–163.
252. *Aust Stefan, Prasad Venkatesha, Niemegeers Ignas GMM.* IEEE 802.11ah: Advantages in standards and further challenges for sub 1 GHz Wi-Fi // Proc. of 2012 IEEE International Conference on Communications (ICC) / IEEE. — 2012. — Pp. 6885–6889.

253. Design, Analysis, and Hardware Emulation of a Novel Energy Conservation Scheme for Sensor Enhanced FiWi Networks (ECO-SFiWi) / Dung Pham Van, Bhaskar Prasad Rimal, Martin Maier, Luca Valcarenghi // *IEEE Journal on Selected Areas in Communications*. — 2016. — Vol. 34, no. 5. — Pp. 1645–1662.
254. *IEEE*. Proposed TGah Draft Amendment. — https://mentor.ieee.org/802.11/documents?is_group=00ah. — 2016. — August.
255. Continuous Probabilistic Count Queries in Wireless Sensor Networks / Anna Follmann, Mario Nascimento, Andreas Züfle et al. // *SSTD* / Springer. — 2011. — Pp. 279–296.
256. Toward Massive Ray-Based Simulations of mmWave Small Cells on Open Urban Maps / Dmitrii Solomitchii, Margarita Gapeyenko, Sebastian S Szyszkowicz et al. // *IEEE Antennas and Wireless Propagation Letters*. — 2017. — Vol. 16. — Pp. 1435–1438.
257. System-level analysis of IEEE 802.11 ah technology for unsaturated MTC traffic / Aleksandr Ometov, Nader Daneshfar, Ali Hazmi et al. // *International Journal of Sensor Networks*. — 2018. — Vol. 26, no. 4. — Pp. 269–282.
258. Анализ производительности беспроводной системы агрегации данных с состязанием для современных сенсорных сетей / А.Я. Омётов, С.Д. Андреев, А.М. Тюрликов, Е.А. Кучерявый // *Информатика и ее применения*. — 2016. — Vol. 3, no. 10. — Pp. 23–31.
259. Омётов А.Я., Щесняк А.С., Кучерявый Е.А. Аналитическая модель беспроводных сетей технологии IEEE 802.11ah // *Электросвязь*. — 2018. — no. 1. — Pp. 51–58.
260. Impact of machine-type communications on energy and delay performance of random access channel in LTE-advanced / Mikhail Gerasimenko, Vitaly Petrov, Olga Galinina et al. // *Transactions on Emerging Telecommunications Technologies*. — 2013. — Vol. 24, no. 4. — Pp. 366–377.

261. *IEEE*. TGah Functional Requirements and Evaluation Methodology Rev. 4. — <https://mentor.ieee.org/802.11/dcn/11/11-11-0905-04-00ah-tgah-functional-requirements-and-evaluation-methodology.doc>. — 2011. <https://mentor.ieee.org/802.11/dcn/11/11-11-0905-04-00ah-tgah-functional-requirements-and-evaluation-methodology.doc>.
262. *OMNeT++*. Online User Manual. — <http://www.omnetpp.org/doc/omnetpp/manual/usman.html>. — 2001-14.
263. Performance analysis of IoT-enabling IEEE 802.11ah technology and its RAW mechanism with non-cross slot boundary holding schemes / Muhammad Qutab-ud din, Ali Hazmi, Behnam Badihi et al. // Proc. of 16th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) / IEEE. — 2015. — Pp. 1–6.
264. Performance evaluation of IEEE 802.11ah and its restricted access window mechanism / Orod Raeesi, Juho Pirskanen, Ali Hazmi et al. // Proc. of International Conference on Communications Workshops (ICC) / IEEE. — 2014. — Pp. 460–466.
265. *Ефимушкина Т.В., Молчанов Д.А., Кучерявый Е.А.* Исследование модели управления доступом к каналам сети WIMAX // *T-Comm — Телекоммуникации и транспорт*. — 2011. — Vol. 5, no. 7. — Pp. 68–71.
266. Cooperative Access Class Barring for Machine-to-Machine Communications / S.-Y. Lien, T.-H. Liao, C.-Y. Kao, K.-C. Chen // *IEEE Transactions on Wireless Communications*. — 2012. — Vol. 11, no. 1. — Pp. 27–32.
267. Efficient small data access for machine-type communications in LTE / Sergey Andreev, Anna Larmo, Mikhail Gerasimenko et al. // 2013 IEEE International Conference on Communications (ICC). — IEEE, 2013. — Pp. 3569–3574. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6655105>.
268. Modeling contention-based M2M transmissions over 3GPP LTE cellular networks / Amir M Ahmadian, Olga Galinina, Sergey Andreev, Yev-

- geni Koucheryavy // Proc. of International Conference on Communications Workshops (ICC) / IEEE. — 2014. — Pp. 441–447.
269. *3GPP*. Evolved Universal Terrestrial Radio Access (EUTRA); Physical layer procedures, TS 36.213. — 2012.
270. LTE Radio Access Network (RAN) enhancements for diverse data applications. — 3GPP Technical Report (TR) 36.822, 2012. — September.
271. *3GPP*. Study on RAN Improvements for Machine-Type Communications. — 2011.
272. Ericsson, R2-096759. — Details of Latency Reduction Alternatives.
273. *3GPP*. Study on provision of low-cost MTC UEs based on LTE, TR 36.888. — 2012.
274. *3GPP*. Feasibility study for Further Advancements for E-UTRA (LTE-Advanced), TR 36.912. — 2011.
275. Reducing Energy Consumption of LTE Devices for Machine-to-Machine Communication / T. Tirronen, A. Larmo, J. Sachs et al. // 2nd IEEE IWM2M Workshop at GLOBECOM. — 2012.
276. Energy and Delay Analysis of LTE-Advanced RACH Performance under MTC Overload / M. Gerasimenko, V. Petrov, O. Galinina et al. // 2nd IEEE IWM2M Workshop at GLOBECOM. — 2012.
277. Understanding the IoT connectivity landscape: A contemporary M2M radio technology roadmap / S. Andreev, O. Galinina, A. Pyattaev et al. // *IEEE Communications Magazine*. — 2015. — September. — Vol. 53, no. 9. — Pp. 32–40.
278. *Bor-Yaliniz I., Yanikomeroğlu H.* The New Frontier in RAN Heterogeneity: Multi-Tier Drone-Cells // *IEEE Communications Magazine*. — 2016. — November. — Vol. 54, no. 11. — Pp. 48–55.

279. Technologies for efficient amateur drone detection in 5G millimeter-wave cellular infrastructure / Dmitrii Solomitckii, Margarita Gapeyenko, Vasili Semkin et al. // *IEEE Communications Magazine*. — 2018. — Vol. 56, no. 1. — Pp. 43–50.
280. Effects of heterogeneous mobility on D2D-and drone-assisted mission-critical MTC in 5G / Antonino Orsino, Aleksandr Ometov, Gabor Fodor et al. // *IEEE Communications Magazine*. — 2017. — Vol. 55, no. 2. — Pp. 79–87.
281. Vehicular cloud networking: architecture and design principles / E. Lee, E. K. Lee, M. Gerla, S. Y. Oh // *IEEE Communications Magazine*. — 2014. — February. — Vol. 52, no. 2. — Pp. 148–155.
282. Internet of Vehicles: From intelligent grid to autonomous cars and vehicular fogs / E. K. Lee, M. Gerla, G. Pau et al. // *International Journal of Distributed Sensor Networks*. — 2016. — September. — Vol. 2, no. 9. — Pp. 1–14.
283. ParticipAct: A Large-Scale Crowdsensing Platform / G. Cardone, A. Corradi, L. Foschini, R. Ianniello // *IEEE Transactions on Emerging Topics in Computing*. — 2016. — January. — Vol. 4, no. 1. — Pp. 21–32.
284. Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm / Bin Guo, Zhu Wang, Zhiwen Yu et al. // *ACM Computing Surveys*. — 2015. — September. — Vol. 48, no. 1. — Pp. 7:1–7:31.
285. *Moustafa Hassnaa, Zhang Yan*. Vehicular Networks: Techniques, Standards, and Applications. — 1st edition. — Boston, MA, USA: Auerbach Publications, 2009.
286. *Gozalvez J., Sepulcre M., Bauza R*. IEEE 802.11p vehicle to infrastructure communications in urban environments // *IEEE Communications Magazine*. — 2012. — May. — Vol. 50, no. 5. — Pp. 176–183.

287. *Nguyen-Minh Huong, Benslimane Abderrahim, Deng Der-Jiunn.* Reliable broadcasting using polling scheme based receiver for safety applications in vehicular networks // *Vehicular Communications*. — 2016. — April. — Vol. 4. — Pp. 1 – 14. <http://www.sciencedirect.com/science/article/pii/S2214209616000048>.
288. A Service-Based Approach to Traffic Sensor Data Integration and Analysis to Support Community-Wide Green Commute in China / Y. Han, G. Wang, J. Yu et al. // *IEEE Transactions on Intelligent Transportation Systems*. — 2016. — September. — Vol. 17, no. 9. — Pp. 2648–2657.
289. Social on the road: Enabling secure and efficient social networking on highways / T. H. Luan, R. Lu, X. Shen, F. Bai // *IEEE Wireless Communications*. — 2015. — February. — Vol. 22, no. 1. — Pp. 44–51.
290. MixGroup: Accumulative Pseudonym Exchanging for Location Privacy Enhancement in Vehicular Social Networks / R. Yu, J. Kang, X. Huang et al. // *IEEE Transactions on Dependable and Secure Computing*. — 2016. — January. — Vol. 13, no. 1. — Pp. 93–105.
291. On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology / J. Petajajarvi, K. Mikhaylov, A. Roivainen et al. // Proc. of International Conference on ITS Telecommunications (ITST). — 2015. — December.
292. *Petrov V., Andreev S., Koucheryavy Y.* An applicability assessment of IEEE 802.11 technology for machine-type communications // Proc. of the 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net). — 2012. — June. — Pp. 24–31.
293. Low Throughput Networks for the IoT: Lessons learned from industrial implementations / G. Margelis, R. Piechocki, D. Kaleshi, P. Thomas // Proc. of IEEE World Forum on Internet of Things (WF-IoT). — 2015. — December.

294. *Reynders B., Meert W., Pollin S.* Range and coexistence analysis of long range unlicensed communication // Proc. of International Conference on Telecommunications (ICT). — 2016. — May.
295. *Mikhaylov K., Petaejaevaervi J., Haenninen T.* Analysis of Capacity and Scalability of the LoRa Low Power Wide Area Network Technology // Proc. of European Wireless Conference (EW). — 2016. — May.
296. 3GPP, “Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE”, *3GPP TR 36.888/r12*, June 2013.
297. 3GPP, “Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)”, *3GPP TR 45.820/r13*, December 2015.
298. 3GPP, “Standardization of NB-IOT completed”, http://www.3gpp.org/news-events/3gpp-news/1785-nb_iot_complete, June 2016.
299. A Primer on 3GPP Narrowband Internet of Things (NB-IoT) / Y.-P. Eric Wang, Xingqin Lin, Ansuman Adhikary et al. // *CoRR*. — 2016. — Vol. abs/1606.04171. <http://arxiv.org/abs/1606.04171>.
300. *Кучерявый Е.А., Молчан С.А., Кондратьев В.В.* Принципы построения сенсоров и сенсорных сетей // *Электросвязь*. — 2006. — no. 6. — Pp. 10–15.
301. Vehicle-based relay assistance for opportunistic crowdsensing over narrowband IoT (NB-IoT) / Vitaly Petrov, Andrey Samuylov, Vyacheslav Begishev et al. // *IEEE Internet of Things Journal*. — 2017.
302. NB-IoT system for M2M communication / R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, A. Ghosh // Proc. of IEEE Wireless Communications and Networking Conference Workshops (WCNCW). — 2016. — April. — Pp. 428–432.
303. Maximum-likelihood detection for energy-efficient timing acquisition in NB-IoT / Harald Kroll, Matthias Korb, Benjamin Weber et al. // *CoRR*. — 2016. — Vol. abs/1608.02427.

304. *Lin X., Adhikary A., Wang Y. P. Eric.* Random Access Preamble Design and Detection for 3GPP Narrowband IoT Systems // *IEEE Wireless Communications Letters*. — 2016. — December. — Vol. 5, no. 6. — Pp. 640–643.
305. Properties of random direction models / Philippe Nain, Don Towsley, Benyuan Liu, Zhen Liu // Proc. of IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM). — 2005. — March.
306. *Молчанов Д.А., Кучерявый Е.А.* Приложения беспроводных сенсорных сетей // *Электросвязь*. — 2006. — no. 6. — Pp. 20–23.
307. *Iversen Villy B et al.* Teletraffic engineering handbook // *ITU-D SG*. — 2005. — Vol. 2. — P. 16.
308. *Alfa Attahiru.* Applied Discrete-Time Queues. — Springer, 2016.
309. *Moltchanov Dmitri.* Performance models for wireless channels // *Computer Science Review*. — 2010. — August. — Vol. 4, no. 3. — Pp. 153–184.
310. *Çil Eren Başar, Örmeci E Lerzan, Karaesmen Fikri.* Structural results on a batch acceptance problem for capacitated queues // *Mathematical Methods of Operations Research*. — 2007. — January. — Vol. 66, no. 2. — Pp. 263–274.
311. *Moltchanov Dmitri, Koucheryavy Yevgeni, Harju Jarmo.* Performance response of wireless channels for quantitatively different loss and arrival statistics // *Performance Evaluation*. — 2010. — January. — Vol. 67, no. 1. — Pp. 1–27.
312. Resource allocation and sharing for heterogeneous data collection over conventional 3GPP LTE and emerging NB-IoT technologies / Vyacheslav Begishev, Vitaly Petrov, Andrey Samuylov et al. // *Computer Communications*. — 2018. — Vol. 120. — Pp. 93–101.

313. *Ross S.* Introduction to probability models. — Academic Press, 2010.
314. *Cox F., Isham V.* Point processes. — Chapman and Hall, 1980.
315. *Moltchanov D.* Distance distributions in random networks // *Elsevier Ad Hoc Networks*. — 2012. — August. — Vol. 10, no. 6. — Pp. 1146–1166.
316. *WINTERsim SLS tool*, <http://winter-group.net/download/>.
317. Efficient small data access for machine-type communications in LTE / S. Andreev, A. Larmo, M. Gerasimenko et al. // Proc. of IEEE International Conference on Communications (ICC). — 2013. — June.
318. Impact of machine-type communications on energy and delay performance of random access channel in LTE-advanced / M. Gerasimenko, V. Petrov, O. Galinina et al. // *Transactions on Emerging Telecommunications Technologies*. — 2013. — Vol. 24, no. 4. — Pp. 366–377.

Список рисунков

1.1	Формализация системной топологии HetNet	37
1.2	Сравнение различных функций управления	44
1.3	Калибровка аналитической модели на базе анализа подключений (сверху) и функции CDF (снизу)	50
1.4	Функция CDF (аналитика)	53
1.5	Диаграмма пропускной способности (аналитика)	54
1.6	Взвешенная справедливая и средняя пропускная способность (слева), индекс Джейна (справа) для различных сценариев	55
1.7	Средняя пропускная способность в зависимости от справедливости	58
1.8	Диапазоны работы различных схем	59
1.9	Примеры развертывания сети H-CRAN	62
1.10	Сравнение эффективности для трех различных сценариев .	78
2.1	Архитектура гетерогенной сети 5G	83
2.2	Структура набора данных и результатов в предложенной методологии	85
2.3	Иллюстрация предложенной классификации для различных видов гетерогенных сетей HetNet	87
2.4	Предложенная классификация для различных типов гетерогенных сетей HetNet	88
2.5	Предложенная высокоуровневая архитектура и результатирующие потоки данных для сетей пятого поколения 5G с несколькими радиотехнологиями RAT для LTE/WiFi сетей	94
2.6	Пример топологии гетерогенной сети	100
2.7	Структура поддержки нескольких радиотехнологий доступа в мобильном устройстве	102
2.8	График функции Ламберта	114
2.9	Геометрия сети для случая двух RAT	129

2.10	Производительность системы для $K = 2$, ограничений нет . . .	130
2.11	Геометрия сети для случая трех RAT	132
2.12	Области передачи данных по разным каналам	133
2.13	Скорость передачи для $K = 3$	134
2.14	Мощность передачи $K = 3$	135
2.15	Энергоэффективность для $K = 3$	136
2.16	Производительность для $K = 2$ с ограничениями по максимальной мощности и минимальной скорости передачи бит	137
2.17	Сравнение производительности политик управления энергопотреблением: оптимальные и более простые подходы	139
2.18	Относительная потеря энергоэффективности: оптимальная схема управления мощностью и более простые подходы . . .	140
3.1	Модель трехузловой системы клиентской ретрансляции . . .	150
3.2	Пример функционирования системы клиентской ретрансляции	151
3.3	Архитектура для передачи трафика M2M с кооперацией . . .	159
3.4	Пример функционирования системы клиентской ретрансляции	162
3.5	Пример агрегации процессов поступления	163
3.6	Топология ретрансляционного облака	167
3.7	Пример временной диаграммы для облака ретрансляций . . .	170
3.8	Сравнение аналитического и имитационного моделирования для средней задержки пакета	177
3.9	Пачечное поступление заявок	178
3.10	Средняя задержка пакета в зависимости от вероятности коллизии	180
3.11	Средняя задержка пакета в зависимости от средней интенсивности поступления пакетов	181
3.12	Среднее количество потраченной на передачу энергии в зависимости от вероятности коллизии	182
3.13	Средняя задержка пакета в зависимости от числа источников	183
3.14	Пропускная способность в зависимости от количества узлов в облаке ретрансляций	184

3.15	Средняя задержка пакета в зависимости от интенсивности поступления пакетов λ_A для различных значений M	185
3.16	Энергоэффективность в зависимости от вероятности успешной передачи пакета	186
4.1	Модель рассматриваемой системы	194
4.2	Распределение показателей интерференции	200
4.3	Численные показатели SIR	201
4.4	Поддержка в создании D2D линии	207
4.5	Построение сети	215
4.6	Калибровка LTE: спектральная эффективность	216
4.7	Калибровка WiFi: спектральная эффективность 1×1 на квадратный км	217
4.8	Пример D2D сценария	220
4.9	Совокупная пропускная способность в зависимости от числа пользователей	225
4.10	Совокупная энергоэффективность в зависимости от числа пользователей	226
4.11	Совокупная пропускная способность ($N = 50$)	227
4.12	Совокупная энергоэффективность ($N = 50$)	228
4.13	Пример введения маршрута	232
4.14	Сравнение стандартизованной и исследуемой архитектур	234
4.15	Протокол управления соединением D2D	235
4.16	Сеть LTE Технологического Университета Брно	238
4.17	Переключение между интерфейсами в D2D	239
4.18	Измерение задержки сервера	241
5.1	Сценарий функционирования МТС	249
5.2	Пример работы алгоритма доступа BASIC	260
5.3	Модель генерации входного потока	262
5.4	Марковская модель для анализа работы системы по вложенным точкам	265
5.5	Упрощенная Марковская модель для основных переходов	265
5.6	Метод оценки потребления энергии	270

5.7	Пропускная способность системы при среднем времени генерации пакета 10 сек.)	273
5.8	Пропускная способность для разного количества окон RAW	274
5.9	Зависимость вероятности конфликта от входного потока . .	275
5.10	Пропускная способность для случаев близких к насыщению	277
5.11	Среднее время моделирования	278
5.12	Распределение ресурсных блоков для доступа данных к каналам	280
5.13	Примеры сигнализации для каналов PUSCH (слева) и PRACH (справа)	281
5.14	Пример временной диаграммы канала PUSCH	282
5.15	Предложенная сигнализация для LTE (COBALT), основанная на конкурентном доступе	285
5.16	Пример временной диаграммы COBALT	285
5.17	Потребление мощности для различных схем доступа	295
5.18	Функция распределения задержки для доступа данных . . .	296
6.1	Предлагаемые стратегии распределения ресурсов	306
6.2	Схема последовательного распределения ресурсов соты сети LTE	308
6.3	Разбиение пространства \mathcal{X} на макросостояния	311
6.4	Вероятность блокировки LTE в зависимости от предложенной нагрузки NB-IoT	314
6.5	Вероятность блокировки NB-IoT в зависимости от предложенной нагрузки LTE	314
6.6	Модель системы NB-IoT	318
6.7	Вероятность потери сообщения	320
6.8	Среднее время задержки	320
6.9	Базовая NB-IoT система	323
6.10	Улучшенная система с ретрансляцией	324
6.11	Иллюстрация задержки ожидания в системе	333
6.12	Подробное описание времени ожидания для сообщения в системе	335

6.13	Пример распределения уровней мощности для эксплуатируемого NB-IoT устройства	336
6.14	Иллюстрация рассматриваемой CeMO	340
6.15	Вероятность потери сообщения в базовом сценарии $R=10$ км	346
6.16	Вероятность потери сообщения в базовом сценарии $R=30$ км	347
6.17	Энергоэффективность устройств IoT в базовом сценарии $R=10$ км	347
6.18	Энергоэффективность устройств IoT в базовом сценарии $R=30$ км	348
6.19	Результат разделения ресурсов между кооператорами и БС	351
6.20	Зависимость вероятности потери сообщения от количества автомобилей в сети	352
6.21	Зависимость энергоэффективности от количества автомобилей в сети	353
6.22	Зависимость времени жизни батареи от количества автомобилей в сети	353

Список таблиц

1.1	Основные параметры и их значения	49
2.1	Эволюция поколений сотовых сетей в терминах услуг и производительности. Источник: GSMA Intelligence	80
2.2	Параметры аналитической модели	104
2.3	Параметры системы	131
3.1	Обозначения аналитической модели	164
3.2	Обозначения аналитической модели	168
4.1	Основные параметры для имитационного моделирования	218
4.2	Основные параметры моделирования	224
4.3	Интенсивность отклонения λ при различных уровнях мобильности.	229
4.4	Основные параметры для имитационного моделирования	241
5.1	Переходные состояния системы	266
5.2	Основные параметры для имитационного моделирования	271
5.3	Энергоэффективность системы	276
5.4	Основные параметры системы	287
5.5	Вероятности переходов $\pi_{ij} = \Pr\{N(t) = j N(t - 1) = i\}$	293
5.6	Сравнение в потребности ресурсных блоков RB	294
5.7	Сравнение средней задержки на доступе [моделирование : анализ]	297
6.1	Параметры системы для численного анализа	313
6.2	Обозначения, использованные в этой работе.	327
6.3	Обозначения используемые установкой моделирования.	346

Приложение А

Документы, подтверждающие внедрение основных результатов
диссертационной работы



Публичное акционерное общество междугородной
и международной электрической связи «Ростелеком»

ул. Гончарная, д. 30
г. Москва, Россия, 115172
тел.: (499) 999-82-83
факс: (499) 999-82-22

e-mail: Europe@rt.ru, web: www.rostelecom.ru

АКТ

г. Москва

О внедрении результатов диссертационной работы
Кучерявого Евгения Андреевича
на тему «Разработка и исследование комплекса моделей и методов распределения
ресурсов в беспроводных гетерогенных сетях связи» в проектах ПАО «Ростелеком»

Настоящим актом подтверждаем, что научные результаты диссертационной работы Кучерявого Евгения Андреевича «Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи», представленной на соискание ученой степени доктора технических наук, обладают актуальностью, представляют практический интерес и были внедрены в научно-исследовательской работе «Анализ результатов исследований Сектора стандартизации электросвязи Международного союза электросвязи (МСЭ-Т) в области сетей связи пятого поколения (5G/IMT-2020)», выполненной Санкт-Петербургским Государственным университетом телекоммуникаций им. проф. М.А. Бонч-Бруевича в интересах ПАО «Ростелеком».

Научные результаты диссертационной работы Е.А. Кучерявого, а именно:

- новая методология исследования распределения ресурсов в гетерогенных сетях связи, отличающаяся от известных комплексным анализом и синтезом различных технологий радиодоступа,
- классификация гетерогенных сетей, охватывающая их различные типы на системном уровне,
- модели и методы выгрузки трафика в гетерогенных сетях, использующих для выгрузки трафика сети D2D, построенные на технологии WiFi Direct с поддержкой со стороны сети LTE,

УТВЕРЖДАЮ



Президент

М.Э. Осеевский

» мая 2018 г.

использованы в разделе 10 ² «Модели трафика, параметры качества обслуживания и качества восприятия для 5G/IMT-2020» отчета по научно-исследовательской работе.

Председатель комиссии:

Руководитель
Представительства ПАО «Ростелеком»
в Международном союзе электросвязи

А.С. Бородин

Члены комиссии:

Начальник отдела международных связей
Департамента международного сотрудничества -
Заместитель Председателя 20-й Исследовательской комиссии
«Интернет вещей, умные города и поселения»
Сектора стандартизации электросвязи
Международного союза электросвязи

О.В. Миронников

Главный специалист
Проектного офиса «Индустриальный интернет»,
Эксперт SC41 JTC1 ISO/IEC

Ю.С. Наливайко



УТВЕРЖДАЮ:

Заместитель генерального директора
ПАО «ГИПРОСВЯЗЬ»

А.Б. Васильев

« 23 » *август* 2018 г.

АКТ

о внедрении результатов диссертационной работы
Кучерявого Евгения Андреевича на тему
«Разработка и исследование комплекса моделей и методов
распределения ресурсов в беспроводных гетерогенных сетях связи» в
ПАО «ГИПРОСВЯЗЬ»

Настоящим актом подтверждаем, что научные результаты диссертационной работы Кучерявого Евгения Андреевича «Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи», представленной на соискание ученой степени доктора технических наук, внедрены в ПАО «ГИПРОСВЯЗЬ» при разработке «Методики планирования сетей NB-IoT»

При разработке методики был использован следующий новый научный результат из диссертации Е.А.Кучерявого:

- модель и метод доставки данных от M2M устройств в

гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяющие для ряда сценариев как повысить вероятность доставки сообщений, так и повысить энергоэффективность передающих узлов.

Кроме того, при разработке методики были использованы результаты диссертации Е.А.Кучерявого в части модели для анализа показателей производительности системы NB-IoT и определения диапазонов нагрузок, при которых выполняются заданные параметры качества обслуживания устройств для сот радиусом 10 и 30 километров.

Председатель комиссии:

Начальник
информационно-аналитического
отдела



А.А.Иванов

Члены комиссии:

Ведущий специалист



Е.Б.Юдина

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ



Российский университет
дружбы народов (РУДН)

ФАКУЛЬТЕТ ФИЗИКО-МАТЕМАТИЧЕСКИХ И ЕСТЕСТВЕННЫХ НАУК

ул. Миклухо-Маклая, д. 6, Москва, Россия, 117198. Телефон: +7495 955 08 86, +7495 952 26 44
www.rudn.ru, sci@rudn.university

№ 23 от 18/11/18 г. 20 18
0801-18/142

УТВЕРЖДАЮ

Декан
факультета физико-математических
и естественных наук РУДН
профессор Воскресенский Л.Г.

А К Т

о реализации научных положений и результатов диссертационной работы Кучерявого Евгения Андреевича на тему «Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи» в научных исследованиях Российского университета дружбы народов

Составлен в том, что комиссия в составе: председателя комиссии Гайдамака Ю.В., д.ф.-м.н., доцента, и членов комиссии Севастьянова Л.А., д.ф.-м.н., профессора, и Гудковой И.А., к.ф.-м.н., доцента кафедры прикладной информатики и теории вероятностей, установила, что нижеперечисленные научные результаты диссертации Кучерявого Е.А. «Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи», представленной на соискание ученой степени доктора технических наук были внедрены в РУДН в рамках НИР Тема № 025201-0-000 «Опережающие исследования беспроводных сетей 5G сетей Интернета вещей».

1. Критерий оптимизации, для которого справедливое распределение ресурсов для гетерогенных сетей с учетом баланса между потоками трафика с разными требованиями к пропускной способности обеспечивается использованием модифицированного критерия max-min, основанного на доступной полосе пропускания, что позволяет достичь до 60% лучшего значения 5% процентиля производительности соты по сравнению с критериями, не учитывающими динамику изменений характеристик радиоканала.
2. Новый архитектурный элемент «Шлюз входа в сеть доступа» (AAGW), реализованный в соответствии со стандартами 3GPP, который позволяет взаимодействовать технологиям 3GPP и IEEE и обеспечивать значительное сокращение сквозных задержек вследствие того, что трафик не нуждается в транзите на удаленный шлюз 3GPP PDN, и существенное сокращение капитальных затрат на развертывание дополнительной инфраструктуры 3GPP.

3. Модель и метод доступа устройств M2M к ресурсам сети LTE на основе модифицированной процедуры COBALT, учитывающие особенности передачи данных малого объема, позволяющие уменьшить задержку пакетов по сравнению с методом PUSCH почти для 85% типов трафика и одновременно в 2-3 раза снизить энергопотребление по сравнению с методом PRACH.

Использование результатов диссертации Е.А. Кучерявого позволило эффективно провести исследования оптимизации механизмов распределения ресурсов в гетерогенных беспроводных сетях 5G, а также создать модельную распределенную сеть в Научном центре моделирования беспроводных сетей 5G Института прикладной математики и телекоммуникаций РУДН.

Председатель комиссии

Ю.В. Гайдамака

Члены комиссии:

Д.А. Севастьянов

И.А. Гудкова



Федеральное государственное автономное образовательное
учреждение высшего образования

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ



ул. Кирпичная, д. 33, Москва, Россия, 105187, тел./факс: 8 (495) 772 9569, e-mail: fbm@hse.ru, bm.hse.ru

Факультет бизнеса и менеджмента

СПРАВКА

о внедрении научных результатов диссертационной работы,
Кучерявого Евгения Андреевича
на тему: «Разработка и исследование комплекса моделей и
методов распределения ресурсов в беспроводных гетерогенных сетях связи».

Научные результаты, полученные в диссертации «Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи» использованы в материалах курса «Сбор, хранение и обработка данных в гетерогенных распределенных компьютерных сетях» (Big Data Collection, Storage&Processing in Heterogeneous Distributed Computer Networks») в разделах «Интернет вещей», «мобильные приложения», «беспроводные сети» магистерской программы «Системы больших данных» (Big Data Systems, <https://www.hse.ru/ma/bigdata>). В 2018 году программа вошла в Top-50 программ QS World University Rankings в области бизнес-аналитики (<https://www.topuniversities.com/>).

В материалах лекций, семинаров и проектной работе студентов используются следующие новые научные результаты, полученные Е.А. Кучерявым в диссертационной работе:

- новый тип сетей связи общего пользования - гетерогенный, объединяющий множество беспроводных технологий сбора данных;
- модель и метод облачной клиентской ретрансляции, учитывающие особенности трафика для устройств M2M, расположенных на границе соты, позволяющие увеличить производительность системы в 3 - 4 раза при уменьшении в несколько раз задержки пакетов;
- модель и метод доставки данных от M2M устройств в гетерогенных сетях 3GPP NB-IoT с возможностью клиентской ретрансляции, позволяющие для ряда сценариев повысить вероятность доставки сообщений и энергоэффективность передающих узлов.

Руководитель школы бизнес-информатики
факультета бизнеса и менеджмента
Национального исследовательского университета
«Высшая школа экономики»
д.т.н., профессор


С.В. Матузева

