

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича»

На правах рукописи

До Фук Хао

**РАЗРАБОТКА МОДЕЛЕЙ И МЕТОДОВ МИНИМИЗАЦИИ СЕТЕВОЙ
ЗАДЕРЖКИ НА ФРАГМЕНТЕ ГИБРИДНОЙ СЕТИ СВЯЗИ С
ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА**

2.2.15 – Системы, сети и устройства телекоммуникаций

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель
кандидат технических наук
Березкин Александр Александрович

Санкт-Петербург – 2025

ОГЛАВЛЕНИЕ

Введение.....	6
Глава 1. Введение в исследование снижение задержек в спутниковых каналах связи с применением технологий искусственного интеллекта	15
1.1 Спутниковые каналы связи: обзор.....	15
1.1.1 Определение и характеристики	15
1.1.2 Процесс передачи данных	16
1.1.3 Учитывание качества обслуживания (QoS).....	19
1.1.4 Важность задержки	20
1.2 Цели диссертации	21
1.3 Область и ограничения исследования	23
1.3.1 Область исследования.....	23
1.3.2 Ограничения исследования	24
1.4 Формулировка проблемы исследования	26
1.5 Структура диссертации	27
Глава 2. Анализ методов снижения задержек в спутниковых каналах связи .	29
2.1 Введение в снижение задержек в спутниковых каналах связи.....	29
2.1.1 Проблемы задержек в спутниковых каналах связи	29
2.1.2 Традиционные методы снижения задержек	30
2.2 Роль искусственного интеллекта в снижении задержки	33
2.2.1 Ограничения традиционных методов	33
2.2.2 Введение в технологии ИИ для спутниковых каналов связи	34
2.2.3 Потенциальные преимущества использования ИИ для снижения задержки.....	36
2.3 Существующие решения на основе ИИ для снижения задержки	37
2.4 Гибридные спутниковые сети: Фокус на снижение задержки	38
2.4.1 Определение и характеристики	38
2.4.2 Преимущества гибридных спутниковых сетей.....	40
2.4.3 Проблемы гибридных спутниковых сетей	41
2.4.4 Связанные исследования	42
2.5 Выводы главы 2	42

Глава 3. Математическое моделирование минимизации сетевой задержки в спутниковых каналах связи с использованием технологий искусственного интеллекта	44
3.1 Введение в математическое моделирование сетевой задержки с применением искусственного интеллекта.....	44
3.2 Особенности передачи данных в спутниковых каналах связи и их влияние на сетевую задержку	45
3.3 Методология управления передачей трафика в спутниковых каналах связи на основе алгоритмов ИИ.....	46
3.4 Обучение модели искусственного интеллекта для классификации сетевого трафика в спутниковых системах.....	50
3.5 Математическая модель приоритизации трафика в спутниковых каналах связи на основе системы М/М/1	51
3.6 Математический эксперимент по оценке средней задержки передачи пакетов в спутниковой сети	58
3.7 Математический эксперимент по определению коэффициента потери пакетов при различных параметрах нагрузки.....	65
3.8 Выводы главы 3	69
Глава 4. Методы минимизации сетевой задержки в гибридных спутниковых сетях с использованием искусственного интеллекта	71
4.1 Введение: Оптимизация распределения ресурсов в многолучевых спутниковых системах.....	71
4.2 Математическая формулировка задачи оптимизации	73
4.2.1 Математическая модель канала спутниковой связи.....	74
4.2.2 Оптимизация освещения зон покрытия спутниковыми лучами (VIP) .	76
4.2.3 Постановка задачи минимизации задержек в спутниковых сетях.....	76
4.2.4 Очередная модель обслуживания трафика в спутниковой сети	78
4.2.5 Формулирование целевой функции и ограничений	80
4.3 Разработка алгоритмических решений на основе методов ИИ	81
4.3.1 Оптимизация распределения ресурсов с помощью NSGA-II.....	82
4.3.2 Применение эволюционного алгоритма SPEA2 для минимизации задержек	84
4.4 Экспериментальные исследования и анализ результатов	85
4.4.1 Описание сценариев симуляции и параметров моделирования.....	85

4.4.2	Оценка производительности предложенных методов и анализ эффективности	86
4.5	Практическая реализация методов минимизации задержки: тестирование на симуляторе	92
4.5.1	Методологии приоритизации трафика с использованием ИИ	92
4.5.2	Классификация сетевого трафика с применением глубокой нейронной сети (DNN)	94
4.5.3	Оптимальное выделение полосы пропускания в сети GEO-спутников	96
4.5.4	Результаты симуляции задержек в очередях на базе SimPy	99
4.6	Выводы главы 4	104
Глава 5.	Сжатие данных без потерь в гибридных спутниковых сетях на основе st-gnn	106
5.1	Введение: Оптимизация распределения ресурсов в многолучевых спутниковых системах	106
5.2	Обзор методов сжатия и подходов к снижению задержек в гибридных спутниковых сетях	107
5.3	Теоретические основы пространственно-временного сжатия данных	109
5.3.1	Архитектура гибридных спутниково-наземных сетей	109
5.3.2	Модель канала связи в спутниковых системах	110
5.3.3	Основы пространственно-временных графовых нейронных сетей	112
5.3.4	Подходы к сжатию данных без потерь в сетевом трафике	114
5.3.5	Использование рекуррентных нейронных сетей в задаче сжатия	116
5.4	Методология предложенного метода сжатия	121
5.4.1	Алгоритм сжатия данных с использованием ИИ	121
5.4.2	ST-GNN как предсказательная модель для сжатия	122
5.4.3	Процесс кодирования и декодирования данных	124
5.4.4	Разработка декомпрессора на основе ИИ	126
5.5	Экспериментальные исследования и анализ результатов	127
5.5.1	Оценка эффективности предложенного метода (Эксперимент 1)	128
5.5.2	Сравнение с существующими подходами (Эксперимент 2)	130
5.6	Результаты метода на реальном оборудовании GEO-модема	135
5.6.1	Описание системы и методика сбора и анализа экспериментальных данных	135

5.6.2	Параметры и конфигурация GEO-модема	138
5.6.3	Полученные результаты и сравнение с симуляционными исследованиями.....	142
5.6.4	Практические выводы и потенциал внедрения.....	147
5.7	Выводы главы 5	147
	Заключение.....	149
	Список сокращений и условных обозначений	152
	Список литературы	154
	Приложение А.	167
	Приложение Б.	168

ВВЕДЕНИЕ

Актуальность темы исследования.

Современное развитие телекоммуникационных технологий требует постоянного совершенствования сетевых решений, особенно в условиях стремительного роста объёмов передаваемых данных и расширения охвата сетей связи. Одной из ключевых задач становится обеспечение надёжного и качественного взаимодействия между пользователями и различными сервисами, независимо от их географического положения. Спутниковые системы связи играют важную роль в этом процессе, позволяя охватить отдалённые и труднодоступные регионы. Однако существенной проблемой остаются сетевые задержки в спутниковых каналах, что снижает качество предоставляемых услуг. Поэтому разработка моделей и методов минимизации сетевых задержек с использованием технологий искусственного интеллекта становится критически важной задачей для развития телекоммуникационной инфраструктуры.

Гибридные спутниковые сети используются для обеспечения надёжного и всепогодного покрытия на глобальном уровне, что особенно важно для удалённых и труднодоступных регионов, где традиционные наземные сети либо отсутствуют, либо имеют ограниченные возможности. Объединение спутниковых и наземных технологий позволяет создавать системы с большей пропускной способностью, меньшими задержками и улучшенной надёжностью связи. Такие сети сочетают преимущества спутниковой связи, обеспечивающей широкий охват, и наземных сетей, которые предоставляют более высокую скорость и стабильность на локальных уровнях, что делает их критически важными для современных телекоммуникационных систем.

Актуальность исследования "Разработка моделей и методов минимизации сетевой задержки на фрагменте гибридной сети связи с использованием технологий искусственного интеллекта" обусловлена быстрым развитием спутниковых систем и ростом требований к качеству обслуживания в гибридных сетях связи. Стратегические планы развития отрасли связи Российской Федерации до 2035 года подчёркивают необходимость создания гибридных орбитально-

наземных сетей, способных обеспечивать надёжную связь в любых условиях, включая труднодоступные регионы, такие как Арктика.

Одной из основных проблем при создании таких сетей являются большие сетевые задержки, которые негативно влияют на качество предоставляемых услуг, включая передачу данных, видеосвязь и другие критически важные сервисы. Взаимодействие многоуровневых спутниковых группировок с наземными сегментами осложняется изменчивостью параметров каналов связи, где задержки, джиттер и стабильность соединения становятся ключевыми показателями качества обслуживания.

Применение методов искусственного интеллекта для минимизации сетевых задержек открывает новые подходы к повышению эффективности и надёжности спутниковых каналов связи. Эти технологии позволяют адаптировать сеть в реальном времени к изменяющимся условиям передачи данных, что существенно улучшает качество обслуживания.

Таким образом, исследования в данной области имеют важное значение как для развития научно-технического потенциала телекоммуникационной отрасли, так и для успешной реализации стратегических задач страны, направленных на создание современных и надёжных систем связи.

Степень разработанности темы.

Анализ научной литературы, как отечественной, так и зарубежной, показывает, что вопросы минимизации сетевых задержек в спутниковых каналах связи уже привлекали внимание учёных, однако исследования в этой области не были проведены в полном объёме и комплексно. В частности, работа Nie (2021) посвящена моделированию и анализу характеристик сетевых задержек в спутниковых каналах при различных условиях, таких как высоты орбит, типы сигналов и погодные условия.

В работе акцентируется внимание на таких факторах, как доплеровский сдвиг, многолучевое распространение, помехи и другие, которые влияют на задержки сигнала. Другие исследования (например, Zong, 2023; Deng, 2021; Хохлов, 2021; Кулиева, 2023) сосредоточены на оптимизации методов передачи

данных для повышения эффективности использования пропускной способности спутниковых каналов и совершенствовании технологий управления потоками данных с целью снижения задержек и потерь пакетов.

В последние годы наблюдается рост интереса к применению машинного обучения и нейронных сетей для улучшения управления трафиком и снижения задержек в спутниковых каналах связи. Например, в работе Рачесо (2020) представлены алгоритмы машинного обучения, которые позволяют прогнозировать поступление пакетов на приёмный терминал на основе характеристик потока кадров.

Несмотря на существующие методы уменьшения сетевых задержек, их эффективность остаётся недостаточной в условиях растущей нагрузки на коммуникационные сети. Использование технологий искусственного интеллекта открывает возможности для создания более точных и адаптивных моделей, способных быстро реагировать на изменения в сети и оптимизировать передачу данных. Однако комплексные решения, объединяющие методы искусственного интеллекта и современные алгоритмы управления потоками данных, остаются недостаточно разработанными, что подчёркивает необходимость дальнейших исследований в этой области.

Научная задача диссертационной работы заключается в разрешении научного противоречия между растущими требованиями к минимизации сетевой задержки для обеспечения QoS в гибридных спутниково-наземных сетях и ограниченной эффективностью существующих, преимущественно реактивных, методов управления трафиком в условиях высокой динамики и стохастичности спутниковых каналов. Для разрешения этого противоречия необходимо разработать и исследовать новый комплексный подход, основанный на проактивных моделях искусственного интеллекта, способных адаптивно управлять сетевыми ресурсами в реальном времени.

Объект исследования. спутниковый канал связи фрагмента гибридной орбитально-наземной сети связи.

Предмет исследования. методы и алгоритмы уменьшения сетевой задержки.

Цель и задачи исследования. Цель исследования: Разработка моделей и методов минимизации сетевой задержки в спутниковых каналах связи с использованием технологий искусственного интеллекта для повышения эффективности и надежности передачи данных в гибридных спутниково-земных системах.

Задачи исследования:

- Провести анализ существующих подходов к снижению сетевой задержки в спутниковых и гибридных сетях, выявить их ограничения и определить перспективные направления для применения технологий искусственного интеллекта.
- Разработать модель прогнозирования трафика и динамической приоритизации пакетов на основе нейросетевых методов для сокращения задержки в очереди.
- Разработать методы многокритериальной оптимизации распределения сетевых ресурсов в многолучевых спутниковых системах с использованием гибридного подхода, сочетающего графовые нейронные сети и эволюционные алгоритмы.
- Разработать и исследовать метод сжатия данных без потерь на основе пространственно-временных графовых нейронных сетей (ST-GNN) как средства снижения задержки передачи.
- Провести сравнительный анализ эффективности предложенных моделей и методов посредством имитационного моделирования и экспериментальных исследований, оценивая ключевые показатели качества обслуживания, такие как средняя задержка и коэффициент потерь пакетов.

Научная новизна. Научная новизна полученных результатов состоит в следующем:

1. Разработанная модель искусственного интеллекта для прогнозирования сетевого трафика и определения приоритетов использует

современные подходы на основе теории систем очередей, что позволяет значительно улучшить показатели задержки и потерь пакетов в спутниковых сетях.

2. Классификация трафика в зависимости от требований приложений к скорости передачи данных является важным элементом, позволяющим более эффективно управлять сетевыми ресурсами. В исследовании предлагается применение эволюционных алгоритмов NSGA-II и SPEA2 для решения задач оптимизации распределения ресурсов, что ведет к значительному снижению сетевой задержки и повышению качества обслуживания в гибридных спутниковых системах.

3. Уникальный метод сжатия данных без потерь, использующий ST-GNN и рекуррентные нейронные сети (RNN), демонстрирует возможность значительного уменьшения объема передаваемых данных без ущерба для качества, что является критически важным для работы спутниковых систем в условиях ограниченной пропускной способности.

Теоретическая и практическая значимость исследования.

Теоретическая значимость исследования:

– Расширение теории систем очередей: Данное исследование развивает и применяет теорию систем очередей для оптимизации спутниковых сетей, что позволяет значительно улучшить задержку и уменьшить потерю пакетов в спутниковых коммуникациях через прогнозирование и классификацию сетевого трафика.

– Исследование представляет собой вклад в развитие теории управления гибридными спутниковыми сетями путем внедрения графовых нейронных сетей (GNN) для классификации трафика и применения эволюционных методов для оптимизации распределения ресурсов. Разработанные модели открывают новые возможности для теоретического обоснования и практического применения методов снижения сетевой задержки и управления ресурсами в сложных системах связи.

– Интеграция технологий ИИ в теорию сжатия данных: Исследование предлагает метод сжатия данных без потерь на основе пространственно-временных графовых нейронных сетей (ST-GNN), открывая новые направления в обработке и оптимизации передачи данных по спутниковым каналам, одновременно минимизируя сетевую задержку при обеспечении качества информации.

Практическая значимость исследования

– Использование искусственного интеллекта и графовых нейронных сетей для эффективного прогнозирования сетевого трафика и управления приоритетами, что обеспечивает более эффективное распределение ресурсов и уменьшает потери пакетов

– Предложенные методы, включающие использование GNN для классификации трафика и эволюционные алгоритмы для оптимизации распределения пропускной способности, частотных спектров и энергопотребления, позволяют существенно снизить сетевую задержку, улучшить качество обслуживания и повысить эффективность работы спутниковых систем. Эти результаты могут быть полезны для проектирования и эксплуатации современных систем связи.

– Внедрение метода сжатия данных без потерь на основе ST-GNN, что позволяет сократить объем передаваемых данных и тем самым ускорить процесс передачи в условиях ограниченной пропускной способности спутниковых сетей, сохраняя высокое качество информации.

Методология и методы исследования.

– Применение методов машинного обучения для создания модели, которая предсказывает сетевой трафик и определяет приоритеты для каждого типа трафика, что позволяет сократить задержку и потерю пакетов в спутниковых сетях. Модель основана на теории систем очередей и включает в себя как теоретические, так и практические аспекты, подтверждаемые симуляциями и численными экспериментами

– Методологическая основа исследования включает использование графовых нейронных сетей (GNN) для классификации трафика в соответствии с требованиями приложений, а также применение эволюционных алгоритмов NSGA-II и SPEA2 для оптимизации распределения сетевых ресурсов. Экспериментальные исследования проводились на симуляционных моделях гибридных спутниковых сетей, что позволило оценить эффективность предложенных подходов в снижении сетевой задержки и улучшении качества обслуживания.

– Разработка метода, который комбинирует пространственно-временные графовые нейронные сети (ST-GNN) с рекуррентными нейронными сетями (RNN) для сжатия данных без потерь. Этот подход снижает объем передаваемых данных, сохраняя при этом качество информации и минимизируя сетевую задержку, что позволяет улучшить работу современных гибридных спутниковых систем связи.

Основные положения, выносимые на защиту.

1. Модель прогнозирования и динамической приоритизации трафика на основе многослойного персептрона (MLP) для снижения задержки и потерь пакетов в спутниковом сегменте гибридной орбитально-наземной сети.

2. Методы оптимизации распределения сетевых ресурсов на основе графовых нейронных сетей (GNN) и эволюционных алгоритмов (NSGA-II/SPEA2) для минимизации задержки в спутниковой сегменте гибридной орбитально-наземной сети связи.

3. Метод сжатия данных без потерь на основе пространственно-временной графовой нейронной сети (ST-GNN) для снижения сетевой задержки в спутниковом канале связи.

Степень достоверности и апробации результатов. Достоверность полученных автором научных и практических результатов диссертационной работы подтверждается корректным использованием математического и имитационного моделирования, соответствием расчетов с результатами экспериментальных исследований, проведенных лично автором. Основные

теоретические и практические результаты работы, в том числе разработанные имитационные модели, использованы в учебном процессе кафедры Сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича при чтении лекций и проведении практических занятий и лабораторных работ.

Апробация результатов исследования. Основные результаты диссертационной работы докладывались и обсуждались на 27st International Conference on Distributed Computer and Communication Networks, DCCN 2024 (Москва, 2024); 8 и 9-я международные конференции «Информация, управление и коммуникационные технологии» (ICCT) (Владикавказ, 2024; Гомель, 2025); The 25th и 26th International Conference on Advanced Communication Technology, ICACT 2023 и 2024 (Корея, 2023, 2024); The 1st и 2nd International Conference on Advanced Computing & Next-Generation Communication (ICACNGC 2022, 2023); на 79-й конференциях Санкт-Петербургского отделения Общероссийской общественной организации «Российское научно-техническое общество радиотехники, электроники и связи им. А. С. Попова» (СПб НТОРЭС) (Санкт-Петербург, 2024), Международной научно-технической и научно-методической конференции «Актуальные проблемы инфотелекоммуникаций в науке и образовании» СПбГУТ (Санкт-Петербург, 2023, 2024); Подготовка профессиональных кадров в магистратуре для цифровой экономики (ПКМ, 2023);

Публикация по теме диссертации. Основные результаты диссертации изложены в 20 опубликованных работах, в том числе в 6 работах, опубликованных в журналах из перечня ВАК Министерства образования и науки Российской Федерации; в 6 работах, опубликованных в трудах, индексируемых Scopus и Web of Science, и 8 публикациях в других изданиях и материалах конференций.

Личный вклад автора. Результаты теоретических и экспериментальных исследований получены автором самостоятельно. В работах, опубликованных в соавторстве, соискателю принадлежит основная роль при постановке и решении задач, а также обобщении полученных результатов.

Соответствие специальности. Диссертационная работа соответствует пунктам п. 1, 2, 3, 18, 19 паспорта специальности 2.2.15 – Системы, сети и устройства телекоммуникаций.

Структура и объем диссертации. Диссертация состоит из введения, 5 глав, выводов, списка используемых источников и двух приложений. Полный объем диссертации составляет 168 страниц, из них к основной части относится 102 страниц. Работа содержит 35 рисунка, 15 таблиц и список из 110 литературных источников.

Благодарность. Я хотел бы выразить свою глубокую благодарность моему научному руководителю за его неоценимую помощь и поддержку, которые позволили мне завершить эту диссертацию. Также выражаю благодарность всем преподавателям кафедры за их постоянную поддержку и содействие. Особую признательность хочу выразить Ле Чан Дык за их руководство и терпение на протяжении всего периода моей работы над кандидатской диссертацией. Искренне благодарю своих родителей за их неизменную поддержку на протяжении всего моего пути. И, конечно, отдельные слова благодарности моей жене, которая проявляла огромную терпимость и заботу, чтобы я мог сосредоточиться на завершении работы.

ГЛАВА 1. ВВЕДЕНИЕ В ИССЛЕДОВАНИЕ СНИЖЕНИЕ ЗАДЕРЖЕК В СПУТНИКОВЫХ КАНАЛАХ СВЯЗИ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

1.1 Спутниковые каналы связи: обзор

1.1.1 Определение и характеристики

Спутниковые каналы связи [1] — это системы, которые обеспечивают двустороннюю передачу данных между спутниками, находящимися на орбите, и наземными станциями. Также они могут осуществлять связь между спутниками в космосе. Основное назначение спутниковых каналов связи — это обеспечение связи на больших расстояниях, включая труднодоступные регионы, где наземные сети отсутствуют или имеют ограничения. Таблица 1 показывает характеристики спутниковых каналов связи.

Основные характеристики спутниковых каналов связи включают:

- Глобальное покрытие: Спутники могут охватывать большие территории Земли, обеспечивая связь на международном уровне и в удалённых районах. Геостационарные спутники (ГЕО) [2] предлагают непрерывное покрытие определённых регионов, а спутники на низких орбитах (ЛЕО) [3] и средних орбитах (МЕО) [4] обеспечивают более оперативное покрытие за счёт меньшего расстояния до Земли.
- Задержка передачи [5]: Одним из важных факторов является задержка сигнала. В системах на основе ГЕО задержка может достигать 500 миллисекунд в оба конца из-за расстояния до спутника. ЛЕО и МЕО системы обладают меньшими задержками из-за их более низкой орбиты.
- Пропускная способность: Пропускная способность спутниковых систем зависит от используемого частотного диапазона [6] (например, Ku, Ka, C диапазоны). Высокочастотные диапазоны могут обеспечивать более высокие

скорости передачи данных, но они подвержены атмосферным явлениям, таким как дождь и облачность, что может снижать качество сигнала.

– Надёжность и качество сигнала: Спутниковые каналы связи требуют сложных технологий для минимизации потерь сигнала и интерференции. Современные методы управления и обработки сигнала позволяют улучшить надёжность и качество передачи данных [7].

– Интеграция с наземными сетями [8]: Спутниковые системы часто работают в сочетании с наземными коммуникационными системами, что позволяет обеспечить комплексное глобальное покрытие и оптимизировать маршруты передачи данных.

Таблица 1. Характеристики спутниковых каналов связи

Характеристика	Описание
Глобальное покрытие	Спутники охватывают обширные территории Земли, обеспечивая связь на международном уровне и в труднодоступных районах.
Задержка передачи	Задержка сигнала зависит от орбиты спутника. В системах ГЕО задержка составляет около 500 миллисекунд, а в системах ЛЕО и МЕО задержка значительно меньше.
Пропускная способность	Зависит от используемого частотного диапазона (Ku, Ka, C и другие). Высокочастотные диапазоны обеспечивают высокую скорость передачи данных.
Надёжность и качество сигнала	Сигналы могут ослабляться из-за погодных условий, таких как дождь или облачность. Современные технологии позволяют улучшить качество и надёжность связи.
Интеграция с наземными сетями	Спутниковые системы часто работают в сочетании с наземными сетями для обеспечения глобального покрытия и оптимизации передачи данных.

1.1.2 Процесс передачи данных

Процесс передачи данных [9] через спутниковый канал связи включает несколько ключевых этапов, каждый из которых играет важную роль в обеспечении эффективного и надёжного обмена информацией между пользователями и системами. Основные этапы этого процесса следующие:

- Подготовка данных: На первом этапе данные, которые необходимо передать, подготавливаются к отправке. Это включает в себя их упаковку в пакеты и кодирование для обеспечения защиты от ошибок. На этом этапе также могут применяться различные методы сжатия данных [10] для уменьшения объема передаваемой информации и увеличения скорости передачи.

- Передача данных на спутник: Подготовленные данные передаются с наземной станции к спутнику. Это может осуществляться с помощью различных типов антенн и передающих устройств, в зависимости от частотного диапазона и требований к пропускной способности. На этом этапе важно обеспечить надежность передачи и минимизировать потери сигнала, используя технологии модуляции и кодирования.

- Передача через спутник: Спутник получает данные от наземной станции и передает их на другой спутник или непосредственно на другую наземную станцию. Спутники могут функционировать в разных орбитальных уровнях (ГЕО, МЕО, ЛЕО), что влияет на задержку передачи и качество сигнала. Спутники также могут использовать репитерные функции для усиления и повторной передачи сигнала.

- Приём данных на наземной станции: Данные, переданные спутником, принимаются другой наземной станцией. На этом этапе данные декодируются и проверяются на наличие ошибок. Возможно применение методов коррекции ошибок для исправления любых искажений, возникших в процессе передачи.

- Обработка и использование данных: После успешного получения и декодирования данных они передаются конечным пользователям или системам для дальнейшей обработки и использования. В этом этапе данные могут быть сохранены, проанализированы или использованы для принятия решений.

Рисунок 1: Схематическое представление процесса передачи данных через спутниковые каналы связи, включая этапы отправки информации с наземных устройств, обработки на земной станции, передачи на спутник, ретрансляции и

последующего приема на другой земной станции с последующей доставкой данных конечным пользователям.



Рисунок 1. Как передавать данные через спутниковые каналы связи

Часть 1: Процесс передачи данных (Uplink) [11]: Исходные данные, которые обычно создаются различными устройствами, такими как компьютеры, телефоны и камеры, сначала обрабатываются на земной станции перед передачей на спутник. На этом этапе данные кодируются (encoder) для обеспечения безопасности и точности передачи. Затем они модулируются (modulator) для преобразования в аналоговый сигнал, подходящий для передачи по радиоволнам. Этот аналоговый сигнал усиливается (amplifier) перед тем, как быть переданным антенной земной станции в космос. Радиоволна, несущая данные, направляется к спутнику, который благодаря точно рассчитанной орбите и положению принимает сигнал.

Часть 2: Процесс приема данных (Downlink) [12]: После получения сигнала от земной станции спутник усиливает сигнал для обеспечения его качества перед повторной передачей на другую земную станцию или на ту же станцию. На приемной станции сигнал принимается антенной и затем дополнительно усиливается для устранения шумов и потерь. После усиления сигнал снова преобразуется в цифровой вид (demodulator) и декодируется (decoder), чтобы восстановить исходные данные. В конце концов, данные передаются на конечные

устройства, такие как компьютеры или телефоны, для использования пользователем.

1.1.3 Учет качества обслуживания (QoS)

В спутниковых каналах связи качество обслуживания (QoS) [13] играет критически важную роль, поскольку оно напрямую влияет на эффективность и надежность передачи данных. Учет QoS включает несколько ключевых аспектов:

- Управление пропускной способностью: Спутниковые каналы связи должны обеспечивать достаточную пропускную способность для удовлетворения требований различных приложений и услуг. Эффективное управление пропускной способностью [14] позволяет распределять ресурсы между пользователями и приложениями, обеспечивая необходимую скорость передачи данных и минимизируя задержки. Это особенно важно для приложений, чувствительных к времени, таких как видеоконференции и потоковое видео.

- Минимизация задержек: Одной из главных проблем в спутниковых каналах связи является высокая задержка, особенно в системах геостационарных спутников. Для достижения высокого уровня QoS необходимо минимизировать задержки передачи данных, что может быть достигнуто с помощью технологий искусственного интеллекта и оптимизации маршрутизации [15]. Применение алгоритмов для управления сетевыми потоками и предсказания задержек помогает улучшить общее качество обслуживания.

- Обеспечение надежности и доступности: Для обеспечения стабильности и надежности связи важно учитывать факторы, такие как погода и технические сбои. Технологии обеспечения QoS должны включать методы для обнаружения и устранения неисправностей, резервные пути передачи данных и автоматическое переключение на альтернативные каналы при сбоях. Это гарантирует, что связь останется доступной даже в условиях неблагоприятных погодных условий или других проблем.

– Качество сигнала и коррекция ошибок [16]: В спутниковых каналах связи качество сигнала может значительно варьироваться в зависимости от внешних условий. Для поддержания высокого уровня QoS важно использовать технологии для коррекции ошибок и улучшения качества сигнала. Это может включать в себя использование продвинутых методов модуляции и кодирования, а также применение адаптивных алгоритмов для регулирования параметров связи.

Таким образом, учет аспектов качества обслуживания в спутниковых каналах связи требует комплексного подхода к управлению пропускной способностью, минимизации задержек, обеспечению надежности и поддержанию высокого качества сигнала. Современные технологии и методы, включая искусственный интеллект, играют ключевую роль в оптимизации QoS и обеспечении эффективной передачи данных через спутниковые каналы.

1.1.4 Важность задержки

Влияние на пользовательский опыт [17]: Задержка в спутниковых каналах связи оказывает значительное влияние на качество пользовательского опыта. В приложениях, чувствительных к времени, таких как видеозвонки, онлайн-игры и потоковое видео, высокая задержка может вызывать задержки и задержки в передаче данных, что приводит к ухудшению качества связи и неудовлетворенности пользователей. Оптимизация задержки критична для обеспечения бесперебойной и качественной коммуникации.

Эффективность передачи данных [18]: Задержка напрямую влияет на эффективность передачи данных в спутниковых системах. Высокая задержка может замедлить передачу информации, увеличивая время, необходимое для обмена данными между пользователями или системами. Это может негативно сказаться на таких приложениях, как мониторинг в реальном времени и автоматизация, где требуется быстрая обработка и передача данных.

Роль в критически важных приложениях [19]: В критически важных приложениях, таких как управление транспортными системами и аварийное

реагирование, задержка может иметь серьезные последствия. Быстрое принятие решений и оперативная передача информации зависят от минимизации задержек. Использование технологий искусственного интеллекта для оптимизации и снижения задержки в таких сценариях помогает повысить надежность и безопасность систем.

Таким образом, управление и минимизация задержки в спутниковых каналах связи являются ключевыми для обеспечения высокого качества обслуживания, эффективной передачи данных и надёжности критически важных приложений.

1.2 Цели диссертации

Быстрый рост глобальных требований к коммуникациям [20] и появление приложений, чувствительных к пропускной способности и задержке, усилили необходимость в эффективных и надежных спутниковых системах связи. Традиционные методы снижения задержки в спутниковых каналах имеют ограничения в адаптации к динамическим условиям сети и удовлетворении строгих требований современных приложений. В этом контексте интеграция технологий искусственного интеллекта (ИИ) представляет собой перспективное направление для повышения производительности спутниковых коммуникационных сетей. Основные цели данной диссертации следующие:

- Разработка моделей прогнозирования и приоритизации трафика на основе ИИ: Одной из ключевых проблем в спутниковых коммуникациях является управление изменяющимся и непредсказуемым характером трафика данных. Для решения этой проблемы в диссертации предполагается разработать передовые модели на основе ИИ, способные точно прогнозировать трафик в спутниковых сетях. С использованием алгоритмов машинного обучения, в частности методов глубокого обучения, эти модели будут анализировать исторические и текущие данные для прогнозирования будущих нагрузок на трафик. Такая прогнозирующая способность позволит динамически приоритизировать пакеты

данных в зависимости от таких факторов, как срочность, важность и требования к качеству обслуживания. Реализация этих моделей позволит более эффективно использовать доступную пропускную способность и ресурсы, снижая заторы и минимизируя задержки. Исследование сосредоточится на выборе подходящих архитектур ИИ, методах обучения и методах валидации для достижения высокой точности прогнозирования и оперативности при различных условиях сети.

– Оптимизация распределения ресурсов в гибридных спутниковых сетях: Гибридные спутниковые сети, объединяющие различные типы спутниковых систем (например, ГЕО, МЕО, ЛЕО) и интегрирующиеся с наземными сетями, предлагают расширенное покрытие и ёмкость. Однако оптимизация распределения ресурсов в таких сложных и гетерогенных средах представляет собой значительные вызовы. В данной диссертации предполагается разработать методы оптимизации с использованием ИИ-техник, в частности генетических алгоритмов, для эффективного распределения ресурсов, таких как пропускная способность, мощность и временные слоты. Предложенные методы будут учитывать несколько факторов, включая топологию сети, шаблоны спроса пользователей и требования к QoS, чтобы определить оптимальные стратегии распределения ресурсов, которые минимизируют задержки в сети. Исследование будет включать проектирование структуры генетического алгоритма, определение соответствующих функций приспособленности и проведение обширных симуляций для оценки улучшений производительности по сравнению с существующими схемами распределения ресурсов.

– Реализация методов сжатия без потерь на основе ИИ: Сжатие данных является важной техникой для уменьшения объема данных, передаваемых через каналы с ограниченной пропускной способностью, тем самым уменьшая время передачи и задержки. В диссертации предполагается реализация инновационных методов сжатия без потерь на основе ИИ, специально адаптированных для спутниковых систем связи. В частности, исследование сосредоточится на разработке модели пространственно-временной графовой нейронной сети (ST-

GNN), которая эффективно захватывает и использует пространственные и временные корреляции в данных. Модель сжатия на основе ST-GNN будет нацелена на достижение высоких коэффициентов сжатия без потери целостности данных, что делает её подходящей для приложений, где критична точность данных. Исследование будет включать проектирование архитектуры сети, обучение модели на соответствующих наборах данных и оценку её производительности в терминах эффективности сжатия и вычислительных затрат по сравнению с традиционными методами сжатия.

Достигнув этих целей, диссертация направлена на значительный вклад в область спутниковых коммуникаций, предлагая решения на основе ИИ, которые эффективно снижают задержки сети, улучшают эффективность передачи данных и повышают общий пользовательский опыт в гибридных спутниковых сетевых средах.

1.3 Область и ограничения исследования

1.3.1 Область исследования

Данная диссертация сосредоточена на исследовании и разработке решений на основе искусственного интеллекта (ИИ), направленных на минимизацию сетевой задержки в гибридных спутниковых коммуникационных сетях. Исследование охватывает следующие ключевые области:

- Прогнозирование и Приоритизация Трафика [21, 22]: Исследование будет углублено в проектирование и внедрение моделей машинного обучения для прогнозирования паттернов трафика в спутниковых сетях. Анализируя исторические и текущие данные, эти модели будут прогнозировать загруженность трафика и позволять динамическую приоритизацию пакетов данных. В рамках исследования будут рассмотрены различные алгоритмы машинного обучения, включая, но не ограничиваясь, нейронные сети, машины опорных векторов и

методы обучения с подкреплением, для определения наиболее эффективных подходов к точному и своевременному прогнозированию трафика.

– Оптимизация Распределения Ресурсов [23]: Исследование будет направлено на применение генетических алгоритмов и других техник эволюционных вычислений для оптимизации распределения ресурсов, таких как пропускная способность, мощность и частотный спектр в гибридных спутниковых сетях. Исследование учтет гетерогенную природу гибридных сетей, которые объединяют различные спутниковые системы и наземную инфраструктуру, и будет стремиться разработать алгоритмы, которые адаптивно распределяют ресурсы на основе текущих условий сети и требований QoS для минимизации задержек и максимизации пропускной способности.

– Сжатие без Потерь: Диссертация также сосредоточится на разработке модели сжатия без потерь на основе ИИ с использованием пространственно-временных графовых нейронных сетей (ST-GNN) [24]. Модель будет направлена на эффективное сжатие данных путем захвата как пространственных, так и временных зависимостей в потоках данных, передаваемых по спутниковым каналам. Исследование будет включать проектирование архитектуры ST-GNN, обучение и валидацию модели на соответствующих наборах данных, а также сравнение ее производительности с традиционными методами сжатия в терминах коэффициента сжатия, скорости и использования ресурсов.

1.3.2 Ограничения исследования

Хотя исследование направлено на предоставление комплексных решений для снижения задержек в спутниковых коммуникационных сетях, признаются определенные ограничения:

– Оценка на основе симуляций: Эффективность и результативность предложенных моделей и методов будет оцениваться в основном через симуляции. Из-за практических ограничений, таких как высокая стоимость и

сложность реальных развертываний спутниковых сетей, исследование будет опираться на смоделированные среды, которые приближены к реальным условиям. Хотя симуляции предоставляют ценные данные, они могут не охватывать все нюансы и непредсказуемости реальных сетевых операций, что может повлиять на обобщаемость результатов.

– Ограниченная доступность данных: Разработка и обучение моделей ИИ требуют обширных и качественных наборов данных. Однако получение данных из реальных спутниковых сетей может быть сложным из-за конфиденциальности, проблем безопасности и ограниченного доступа. В результате, исследование может использовать синтетические или общедоступные наборы данных, которые могут не полностью отражать разнообразие и сложность реальных паттернов трафика и условий сети. Это ограничение может повлиять на точность и надежность моделей ИИ.

– Фокус на гибридных спутниковых сетях: Исследование специально нацелено на гибридные спутниковые сети, которые интегрируют несколько спутниковых систем и наземные компоненты. Хотя такой фокус позволяет провести глубокий анализ и разработать целевые решения, полученные результаты и разработанные методы могут не быть непосредственно применимы к другим типам спутниковых сетей, таким как независимые системы ГЕО или ЛЕО. Для адаптации и проверки предложенных решений в различных контекстах сетей потребуются дальнейшие исследования.

– Признавая эти ограничения, исследование нацелено на предоставление реалистичной и прозрачной оценки предложенных подходов на основе ИИ для снижения задержек в спутниковых коммуникационных каналах. Будущая работа может устранить эти ограничения, включив реальное тестирование, расширив сбор данных и адаптируя решения к более широкому спектру конфигураций сетей.

1.4 Формулировка проблемы исследования

Эволюция спутниковых технологий открыла новые возможности для глобальной связи, особенно в удаленных и труднодоступных районах. Однако, несмотря на прогресс, спутниковые каналы связи сталкиваются с рядом ограничений, таких как задержка и ограниченная пропускная способность, которые отрицательно влияют на их эффективность. В условиях увеличивающейся потребности в приложениях реального времени, таких как телемедицина, видеоконференции и критически важные системы управления, эти проблемы становятся все более заметными.

Основные сложности связаны с задержками распространения сигнала между земными станциями и спутниками, что особенно критично для геостационарных спутников, где расстояния составляют десятки тысяч километров. Ограниченная пропускная способность спутниковых каналов дополнительно усугубляет ситуацию, влияя на качество связи, что особенно остро ощущается при неблагоприятных атмосферных условиях и внешних помехах.

Современные решения, направленные на снижение сетевых задержек, такие как оптимизация протоколов или использование низкоорбитальных спутников, хоть и показывают улучшения, не всегда способны эффективно адаптироваться к динамичным условиям сети, особенно в случае гибридных спутниковых систем. Эти методы часто оказываются недостаточно гибкими для интеграции с новыми требованиями, которые предъявляют современные сети и приложения, такие как 5G и IoT.

Таким образом, очевидно, что для эффективного снижения сетевых задержек в гибридных спутниковых сетях требуется инновационный подход, который выходит за рамки традиционных решений. Искусственный интеллект (ИИ) предлагает перспективный способ решения этих проблем, позволяя улучшить управление ресурсами, прогнозирование трафика и адаптацию к изменяющимся условиям сети. Это исследование направлено на разработку и внедрение моделей и методов на основе ИИ для минимизации задержек и

повышения производительности спутниковых сетей, чтобы соответствовать требованиям современных приложений.

1.5 Структура диссертации

Диссертация организована таким образом, чтобы систематически решать исследовательские задачи и представить результаты в логичной и последовательной форме. Структура работы включает следующие главы:

Глава 1: Введение. В этой главе вводится тема исследования, формулируются цели, определяются объем и ограничения диссертации. Она задает контекст для последующих глав, описывая ключевые вызовы и предлагаемые решения для снижения задержек в спутниковых каналах связи с использованием технологий искусственного интеллекта.

Глава 2: Анализ методов снижения задержек в спутниковых каналах связи. Эта глава представляет собой всесторонний обзор существующих методов для уменьшения задержек в спутниковых каналах связи. Рассматривается роль искусственного интеллекта в этих методах, а также специфические проблемы и возможности, связанные с использованием гибридных спутниковых сетей.

Глава 3: Разработка моделей для минимизации сетевой задержки с использованием технологий искусственного интеллекта. В этой главе подробно рассматривается разработка моделей прогнозирования трафика и приоритизации на основе ИИ. Включены математическая формулировка этих моделей, процесс их обучения и оценка их эффективности в снижении задержек в спутниковых каналах связи.

Глава 4: Разработка методов минимизации сетевой задержки в гибридных спутниковых сетях с использованием технологий искусственного интеллекта. Эта глава сосредоточена на разработке и применении методов распределения ресурсов для гибридных спутниковых сетей. Обсуждаются использование генетических алгоритмов и других ИИ-техник для оптимизации распределения

ресурсов, а также оценивается их эффективность в минимизации сетевой задержки.

Глава 5: Сжатие данных без потерь на основе ST-GNN для снижения задержек в гибридных спутниковых сетях с использованием технологий искусственного интеллекта. В этой главе представлена реализация и оценка модели сжатия данных без потерь на основе ST-GNN. Рассматривается, как данная модель может использоваться для уменьшения объема данных, передаваемых через спутниковые каналы, что способствует снижению задержек и улучшению эффективности связи.

ГЛАВА 2. АНАЛИЗ МЕТОДОВ СНИЖЕНИЯ ЗАДЕРЖЕК В СПУТНИКОВЫХ КАНАЛАХ СВЯЗИ

2.1 Введение в снижение задержек в спутниковых каналах связи

2.1.1 Проблемы задержек в спутниковых каналах связи

Задержка в спутниковых каналах связи, или временной интервал между передачей и приёмом данных, является ключевой проблемой, влияющей на эффективность связи. Основные причины задержек можно выделить в следующие категории:

- Длительные задержки распространения сигнала [25]: Расстояния между земными станциями и спутниками на орбите играют важную роль в увеличении времени передачи сигнала. Это явление неизбежно и представляет собой серьёзную проблему для приложений, требующих мгновенной реакции, таких как видеоконференции, онлайн-игры и удалённые операции.

- Ограничения пропускной способности [26]: Спутниковая пропускная способность является ограниченным ресурсом, зависящим от частотного спектра, выделенного международными организациями, и физических возможностей спутников. Когда нагрузка на канал высока, данные вынуждены ставиться в очередь, что увеличивает задержку. Это особенно актуально для приложений, требующих больших объёмов данных, таких как передача мультимедиа или научные исследования.

- Динамические условия канала [27]: Множество внешних факторов, таких как погодные условия (дождь, снег, атмосферные возмущения), солнечная активность и движение спутников, могут ухудшить качество сигнала. Это приводит к затуханию и увеличению количества ошибок при передаче данных, что требует дополнительных повторных попыток передачи и, соответственно, увеличивает задержки. Спутники на негеостационарных орбитах также вносят

дополнительную вариативность в качестве связи из-за их движения относительно Земли.

Задержки в спутниковых каналах негативно сказываются на ряде приложений. В реальном времени они могут вызвать эхо, задержки откликов и общее ухудшение качества взаимодействия. Для удалённого управления и мониторинга в критически важных отраслях, таких как оборона, аэрокосмическая промышленность или высокоточная автоматизация, задержки могут привести к нарушению процессов и значительным операционным сбоям. В области высокочастотной торговли задержка напрямую влияет на скорость выполнения сделок, что может оказать значительное влияние на конкурентоспособность торговых систем.

2.1.2 Традиционные методы снижения задержек

Традиционные методы снижения задержек в спутниковых каналах связи сосредоточены на оптимизации различных аспектов процесса связи. К этим методам относятся.

а) Эффективные методы кодирования

Эффективные методы кодирования [28] играют ключевую роль в минимизации задержки в спутниковой связи. Эти методы включают:

- Оценка на Основе Симуляций: Эффективность и результативность предложенных моделей и методов будет оцениваться в основном через симуляции. Из-за практических ограничений, таких как высокая стоимость и сложность реальных развертываний спутниковых сетей, исследование будет опираться на смоделированные среды, которые приближены к реальным условиям. Хотя симуляции предоставляют ценные данные, они могут не охватывать все нюансы и непредсказуемости реальных сетевых операций, что может повлиять на обобщаемость результатов.

– Коды исправления ошибок (ЕСС): Методы ЕСС, такие как корректирующие коды с прямым восстановлением (FEC), добавляют избыточные данные к передаваемому сигналу, позволяя приемнику обнаруживать и исправлять ошибки без необходимости повторных передач. Уменьшая потребность в повторной передаче данных, ЕСС помогает снизить задержку. Однако существует компромисс между объемом добавленной избыточности и общей скоростью передачи данных, что требует тщательного управления для оптимизации производительности.

– Сжатие данных: Сжатие данных уменьшает объем передаваемых данных, что ускоряет передачу и снижает задержку. Особое значение имеют методы без потерь, которые сохраняют исходные данные без потери информации, что важно для приложений, где требуется высокая целостность данных. Однако алгоритмы сжатия должны быть эффективными, чтобы избежать добавления задержек, связанных с их обработкой.

– Модуляционные схемы: Продвинутое схемы модуляции, такие как квадратурная амплитудная модуляция (QAM) и ортогональное частотное разделение (OFDM), могут увеличить скорость передачи данных, упаковывая больше информации в каждый передаваемый символ. Эти технологии повышают эффективность передачи данных и помогают уменьшить задержку, особенно в условиях ограниченной полосы пропускания.

б) Увеличение полосы пропускания канала

Увеличение доступной полосы пропускания для спутниковой связи — это еще один способ снижения задержки:

– Более высокие частотные диапазоны [29]: Работа в более высоких частотных диапазонах, таких как Ка-диапазон (26,5–40 ГГц) или Ку-диапазон (12–18 ГГц), предоставляет больше полосы пропускания по сравнению с традиционными частотами С-диапазона (4–8 ГГц). Это позволяет повысить скорость передачи данных, что сокращает время, необходимое для передачи больших объемов данных. Однако более высокие частоты более подвержены

атмосферным воздействиям, таким как дождевые затухания, что может повлиять на надежность связи и задержку.

- Методы мультиплексирования: Методы временного (TDM), частотного (FDM) и кодового (CDMA) мультиплексирования используются для максимального использования доступной полосы пропускания, позволяя нескольким сигналам использовать один и тот же канал связи. Эти методы увеличивают общую пропускную способность данных, уменьшая время ожидания для пакетов данных и, следовательно, снижая задержку.

- Проектирование спутниковых созвездий: Развертывание созвездий спутников, таких как спутники на низкой околоземной орбите (LEO), может снизить задержку за счет уменьшения расстояния между спутниками и наземными станциями. Спутники LEO находятся ближе к Земле, что сокращает задержки распространения по сравнению с геостационарными спутниками. Однако проектирование и поддержание созвездия спутников требует значительных технических и финансовых усилий.

в) Интеллектуальные методы маршрутизации

Интеллектуальные методы маршрутизации [30] используются для оптимизации маршрута, по которому данные передаются через сеть, что минимизирует задержки:

- Маршрутизация по кратчайшему пути: Алгоритмы, такие как алгоритмы Дейкстры и Беллмана-Форда, используются для нахождения кратчайшего пути между источником и приемником, сокращая общее время передачи. В спутниковых сетях это включает динамический выбор наиболее эффективного маршрута через спутниковые узлы и наземные станции.

- Адаптивная маршрутизация: Адаптивные алгоритмы маршрутизации корректируют маршруты в реальном времени в зависимости от текущих условий сети, таких как перегрузка или сбои на линиях связи. Избегая перегруженных маршрутов и динамически выбирая пути с меньшими задержками, эти алгоритмы помогают минимизировать задержку.

- Балансировка нагрузки: Методы балансировки нагрузки распределяют трафик данных равномерно по нескольким путям или каналам, предотвращая перегрузку одного маршрута. Это обеспечивает более стабильное время передачи данных и помогает избежать всплесков задержки, которые могут возникать при перегрузке отдельных маршрутов.

г) Оптимизация кэширования

Кэширование [31] — это еще один метод снижения задержки, особенно в сетях доставки контента (CDN):

- Кэширование на краю сети: Размещение кэшей данных ближе к конечным пользователям, на краю сети, уменьшает расстояние, которое должны пройти данные, что снижает задержку. В спутниковой связи кэширование на краю сети включает хранение часто запрашиваемых данных на наземных станциях или промежуточных спутниковых узлах, что уменьшает необходимость извлечения данных с удаленных серверов.

- Сети доставки контента (CDN): CDN используют распределенную сеть серверов для хранения копий контента в нескольких местах. Обслуживая контент с ближайшего доступного сервера, CDN значительно снижают задержку, особенно для больших файлов, таких как видео или обновления программного обеспечения.

2.2 Роль искусственного интеллекта в снижении задержки

2.2.1 Ограничения традиционных методов

Хотя традиционные методы до определённой степени эффективны в решении проблемы задержки, они сталкиваются с рядом ограничений, особенно когда речь идёт о всё более сложных и динамичных спутниковых сетях связи. Таблица 2 демонстрирует основные ограничения традиционных подходов к уменьшению задержки в спутниковых каналах связи.

Одним из главных недостатков является статический характер традиционных методов, основанных на заранее заданных параметрах и конфигурациях. Такие методы не обладают гибкостью для оперативного реагирования на изменения в трафике или условиях каналов связи. В результате фиксированные маршруты и распределение пропускной способности могут приводить к неэффективной работе и увеличению задержки, особенно в условиях динамически изменяющихся сетей.

Таблица 2. Ограничения традиционных методов уменьшения задержки в спутниковых каналах связи.

Ограничения традиционных методов	Описание
Статические подходы	Многие традиционные методы зависят от заранее определённых параметров и статических конфигураций, которые могут быть неэффективны при изменении условий сети. Например, фиксированные маршруты или распределение пропускной способности могут не реагировать на внезапные изменения трафика или состояния канала, что приводит к снижению производительности и увеличению задержки.
Сложность и масштабируемость	По мере увеличения размеров и сложности спутниковых сетей традиционные методы могут испытывать трудности с масштабируемостью. Управление множеством переменных, таких как распределение частот, мощности и коррекция ошибок, становится всё более сложным и часто требует ручного вмешательства или простых эвристик, которые могут быть недостаточны для крупномасштабных систем.
Неспособность прогнозировать и адаптироваться	Традиционные методы, как правило, не способны предсказывать будущие условия сети или адаптироваться в режиме реального времени. Это приводит к неэффективности, например, недоиспользованию пропускной способности в периоды низкого трафика или перегрузке в пиковые моменты, что увеличивает задержку.

Кроме того, по мере роста размеров спутниковых сетей, традиционные методы сталкиваются с проблемами масштабируемости и усложнения управления. С увеличением числа спутников и наземных станций управление такими параметрами, как распределение частот, мощности и коррекция ошибок,

становится всё более сложным. Это часто требует ручного вмешательства или использования простых эвристик, что не всегда достаточно для крупных систем. Дополнительно, традиционные методы не способны прогнозировать изменения условий сети и адаптироваться в режиме реального времени, что увеличивает вероятность перегрузки и, как следствие, задержки в периоды пиковой нагрузки.

2.2.2 Введение в технологии ИИ для спутниковых каналов связи

Искусственный интеллект (ИИ) предлагает перспективные решения для преодоления ограничений традиционных методов снижения задержки. Наиболее актуальными технологиями ИИ в спутниковых каналах связи являются следующие:

- Машинное обучение (ML) [32]: Алгоритмы машинного обучения способны анализировать большие объёмы исторических и актуальных данных для выявления закономерностей и построения прогнозов. В спутниковой связи ML может использоваться для прогнозирования трафика, предсказания качества каналов и оптимизации распределения ресурсов. Обучаясь на основе прошлых данных, модели ML могут принимать обоснованные решения, которые помогают сократить задержки и улучшить общую производительность сети.

- Глубокое обучение (DL) [33]: Являясь подмножеством машинного обучения, глубокое обучение использует нейронные сети с множеством слоёв для моделирования сложных взаимосвязей в данных. DL может применяться к таким задачам, как обработка сигналов, динамическое распределение ресурсов и оптимизация маршрутизации. Способность DL-моделей обрабатывать большие объёмы данных и выявлять сложные паттерны делает их идеальными для управления сложностью спутниковых сетей.

- Обучение с подкреплением (RL) [34]: В обучении с подкреплением модели обучаются принимать решения на основе проб и ошибок с целью максимизации целевого вознаграждения. В спутниковой связи RL может

применяться для адаптивной маршрутизации и планирования, где модель учится выбирать оптимальные маршруты или графики для минимизации задержек в условиях изменяющейся среды. Способность RL обучаться на основе постоянного взаимодействия с сетью делает его особенно подходящим для динамических и непредсказуемых сред.

2.2.3 Потенциальные преимущества использования ИИ для снижения задержки

Интеграция искусственного интеллекта (ИИ) в системы спутниковой связи предоставляет несколько потенциальных преимуществ для снижения задержки:

- Адаптивное управление трафиком: ИИ может анализировать данные о трафике в реальном времени, чтобы прогнозировать congestions и динамически приоритизировать пакеты данных в зависимости от их срочности. Например, ИИ может отдавать приоритет экстренной связи или приложениям, требующим реального времени, по сравнению с менее срочными данными, что гарантирует, что критическая информация передается с минимальной задержкой.
- Интеллектуальная маршрутизация и планирование: Алгоритмы ИИ могут оптимизировать маршруты передачи данных, постоянно оценивая условия сети, такие как качество связи и нагрузка на трафик. Выбирая наиболее эффективные маршруты и планируя передачу данных в оптимальные моменты, ИИ может значительно сократить задержку и улучшить согласованность доставки данных.
- Динамическое распределение ресурсов [35, 36]: ИИ может использоваться для более эффективного распределения ресурсов, таких как пропускная способность и мощность, в зависимости от реального спроса и условий сети. Например, ИИ может динамически регулировать распределение пропускной способности, чтобы обеспечить достаточное количество ресурсов для

приложений с высоким приоритетом в периоды пикового использования, что снижает задержки и улучшает общую производительность сети.

2.3 Существующие решения на основе ИИ для снижения задержки

Таблица 3 представляет собой обзор существующих методов на основе искусственного интеллекта, используемых для минимизации задержек в спутниковых каналах связи.

Таблица 3. Существующие решения на основе ИИ для снижения задержки

Название метода	Тип ИИ (ML, DL, RL, Heuristic)	Описание	Ссылка на статью
Глубокие нейронные сети (DNN)	DL	Глубокие нейронные сети используют большие объемы данных для предсказания задержек и их минимизации.	[37]
Рекуррентные нейронные сети (RNN, LSTM, GRU)	DL	Рекуррентные нейронные сети применяются для предсказания временных рядов задержек в спутниковых сетях.	[38]
Обучение с подкреплением (RL, DQN, Q-Learning)	RL	Алгоритмы обучения с подкреплением динамически настраивают параметры сети, обучаясь минимизировать задержки.	[39]
Генетические алгоритмы (GA)	Heuristic	Генетические алгоритмы используют эволюционные подходы для поиска оптимальных решений по маршрутизации.	[40]
Алгоритм роя частиц (PSO)	Heuristic	Оптимизирует маршруты и управление ресурсами для минимизации задержек путем имитации поведения частиц.	[41]
Гибридные методы (RL + DL)	RL + DL	Комбинация методов глубокого обучения и обучения с подкреплением для достижения максимальной эффективности в минимизации задержек.	[42]

В таблице перечислены различные подходы, включая машинное обучение (ML), глубокое обучение (DL), обучение с подкреплением (RL) и эвристические методы, такие как генетические алгоритмы (GA) и алгоритм роя частиц (PSO). Каждый из этих методов направлен на решение проблемы оптимизации сетевых задержек с помощью современных ИИ-технологий.

Глубокие нейронные сети (DNN) и рекуррентные нейронные сети (RNN, LSTM, GRU) относятся к методам глубокого обучения (DL). Они используются для анализа и предсказания задержек, основываясь на большом объеме данных. DNN эффективно выявляют сложные взаимосвязи между входными параметрами и задержками, а RNN особенно полезны для временных рядов данных, таких как задержки в сети. Эти методы позволяют не только прогнозировать поведение сети, но и вносить коррективы для уменьшения задержек.

Эвристические методы, такие как генетические алгоритмы и алгоритм роя частиц, представляют собой оптимизационные подходы, которые помогают находить оптимальные маршруты для передачи данных и управлять ресурсами сети. В то же время методы обучения с подкреплением, такие как Q-Learning и Deep Q-Networks (DQN), используют взаимодействие с сетью для улучшения стратегии управления. Гибридные подходы, комбинирующие обучение с подкреплением и глубокое обучение, позволяют достигать еще более высоких результатов, сочетая преимущества обоих подходов.

2.4 Гибридные спутниковые сети: Фокус на снижение задержки

2.4.1 Определение и характеристики

Рисунок 2 представляет собой архитектуру гибридной спутниковой сети [43], которая включает в себя различные уровни платформ, обеспечивающих связь, с целью минимизации сетевой задержки.

Гибридные спутниковые сети объединяют различные типы спутников (например, геостационарные, спутники средней околоземной орбиты и спутники

низкой околоземной орбиты) с наземной коммуникационной инфраструктурой для обеспечения широкого охвата и улучшения производительности. Эти сети используют сильные стороны каждого компонента:

- Платформы на земле: Дроны (UAV) на уровне городских или сельских районов могут использоваться для обеспечения локальной связи и передачи данных.
- Воздушные платформы: Стратосферные платформы высокой высоты (HAPS), находящиеся на высоте до 100 км над землей, играют важную роль в расширении покрытия сети и снижении задержки в областях с трудным доступом.

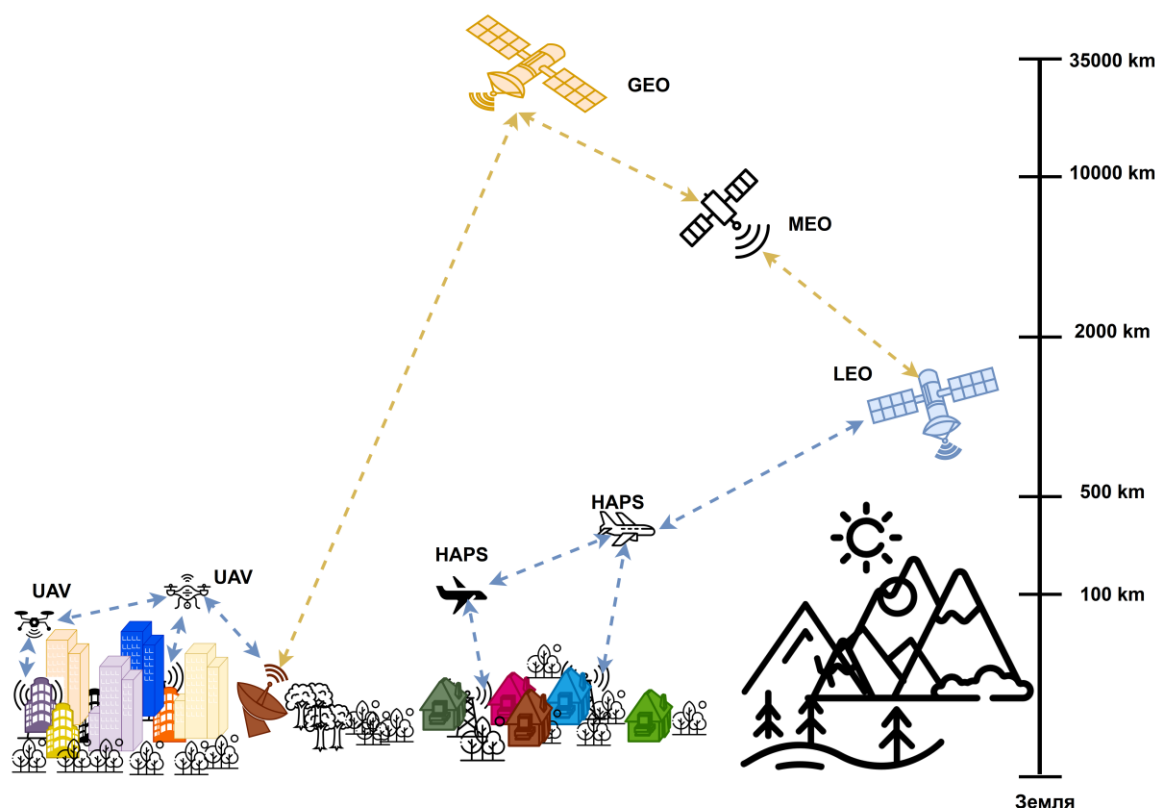


Рисунок 2. Архитектура гибридных спутниковых сетей

- Геостационарные спутники (GEO): Размещенные на фиксированной высоте над экватором, GEO спутники обеспечивают непрерывное покрытие больших территорий, что делает их идеальными для вещания и дальнодействующей связи. Однако их высокая орбита приводит к значительным задержкам.

- Спутники средней околоземной орбиты (МЕО): Спутники МЕО находятся на высотах от 2,000 до 35,786 километров, предлагая компромисс между покрытием спутников GEO и меньшими задержками спутников LEO. Спутники МЕО часто используются для навигации и передачи данных.

- Спутники низкой околоземной орбиты (LEO): Спутники LEO, находящиеся на высотах ниже 2,000 километров, обеспечивают связь с низкой задержкой благодаря близости к Земле. Однако их ограниченная зона покрытия требует создания созвездия спутников для обеспечения непрерывного обслуживания.

Гибридная сеть использует комбинацию всех этих платформ для обеспечения связи с минимальной задержкой. Спутники на низкой и средней орбите, а также стратосферные платформы помогают ускорить передачу данных, в то время как геостационарные спутники обеспечивают более широкое покрытие.

2.4.2 Преимущества гибридных спутниковых сетей

Гибридные спутниковые сети предоставляют несколько преимуществ, которые могут способствовать снижению задержек:

- Увеличение охвата: Комбинируя различные типы спутников, гибридные сети могут обеспечить глобальное покрытие, включая удаленные или недостаточно обслуживаемые районы, где наземные сети недоступны. Это широкое покрытие гарантирует, что пользователи могут поддерживать связь с низкой задержкой независимо от их местоположения.

- Улучшенная надежность: Избыточность, присущая гибридным сетям, с несколькими спутниками и наземными узлами, повышает общую надежность коммуникационной системы. В случае сбоя одного из компонентов (например, спутника или наземной станции) сеть может перенаправить трафик через

альтернативные маршруты, поддерживая непрерывность обслуживания и минимизируя задержки.

– Потенциал для снижения задержек: Спутники LEO в гибридных сетях обеспечивают связь с низкой задержкой, особенно для приложений в реальном времени. Динамическое маршрутизирование трафика через наиболее низкозадерживающие пути позволяет гибридным сетям оптимизировать производительность и снизить задержки.

2.4.3 Проблемы гибридных спутниковых сетей

Несмотря на свои преимущества, гибридные спутниковые сети сталкиваются с несколькими проблемами:

– Проблемы совместимости: Интеграция различных типов спутников и наземных сетей требует тщательной координации для обеспечения совместимости. Различия в протоколах, форматах данных и архитектуре сетей могут привести к проблемам совместимости, что может увеличить задержки или снизить общую эффективность сети.

– Управление задержками: Управление задержками в гибридной сети, где различные компоненты имеют различные задержки, является сложной задачей. Сеть должна динамически выбирать оптимальный путь для передачи данных, балансируя компромиссы между задержкой, пропускной способностью и надежностью.

– Оптимизация распределения ресурсов: Эффективное распределение ресурсов, таких как пропускная способность, мощность и вычислительные мощности, между различными компонентами сети представляет собой вызов. Неоптимальное распределение ресурсов может привести к перегрузке, увеличению задержек или недоиспользованию возможностей сети.

2.4.4 Связанные исследования

Существующие исследования по гибридным спутниковым сетям изучили различные аспекты снижения задержки:

- Оптимизация созвездий спутников: Исследования рассматривали оптимальный дизайн созвездий спутников для минимизации задержек, особенно для спутников LEO. Исследования сосредоточены на определении лучших орбитальных параметров, интервалов между спутниками и стратегий переключения для снижения задержек связи.

- Межспутниковые ссылки (ISL): Исследования межспутниковых ссылок, которые позволяют прямую связь между спутниками, подчеркнули их потенциал для снижения задержек за счет более прямого маршрутизирования данных. Исследования изучили дизайн и оптимизацию сетей ISL в рамках гибридных систем.

- Оптимизация на основе ИИ: Несколько исследований предложили методы оптимизации на основе ИИ для распределения ресурсов и маршрутизации в гибридных сетях. Эти методы показали обещающие результаты в улучшении производительности в условиях динамичной и непредсказуемой среды.

2.5 Выводы главы 2

1. Глава 2 представила всесторонний анализ методов снижения задержек в спутниковых каналах связи, освещая как традиционные подходы, так и возникающую роль ИИ. Традиционные методы, хотя и эффективны в определённых сценариях, сталкиваются с ограничениями в адаптации к сложной и динамичной природе современных спутниковых сетей. Технологии ИИ предлагают значительный потенциал для преодоления этих вызовов, предоставляя более адаптивные, интеллектуальные и эффективные решения для снижения задержек.

2. В главе также были представлены гибридные спутниковые сети как перспективный подход к улучшению производительности по снижению задержек. Хотя гибридные сети предлагают несколько преимуществ, таких как расширенное покрытие и надёжность, они также представляют собой уникальные проблемы, требующие дальнейшего исследования. Эта диссертация нацелена на решение этих проблем путём разработки новых моделей и методов на основе ИИ, которые улучшат производительность по снижению задержек в гибридных спутниковых сетях, способствуя продолжающейся эволюции технологий спутниковой связи.

ГЛАВА 3. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МИНИМИЗАЦИИ СЕТЕВОЙ ЗАДЕРЖКИ В СПУТНИКОВЫХ КАНАЛАХ СВЯЗИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

3.1 Введение в математическое моделирование сетевой задержки с применением искусственного интеллекта

В условиях стремительного роста объемов передаваемых данных и повышения требований к качеству обслуживания (QoS) в современных спутниковых системах связи задача минимизации сетевой задержки становится все более актуальной [44]. Спутниковые системы, будучи основным компонентом глобальной связи, сталкиваются с серьезными вызовами по поддержанию стабильной и низкой задержки при передаче данных, особенно в условиях ограниченных сетевых ресурсов [45].

В данном разделе рассматриваются ключевые подходы к разработке моделей минимизации сетевой задержки, основанные на использовании технологий искусственного интеллекта (ИИ). Эти методы, включая глубокие нейронные сети и алгоритмы машинного обучения, направлены на решение задач прогнозирования и оптимизации сетевых параметров, что позволяет повысить адаптивность и эффективность передачи данных в спутниковых сетях [46]. Применение ИИ открывает новые возможности для адаптивного управления трафиком, что способствует снижению задержек и улучшению QoS даже при высоких нагрузках на сеть [47].

Предлагаемые модели включают использование многослойных перцептронов (multi-layered perceptrons, MLP) для прогнозирования сетевых параметров и разработки схем приоритезации трафика. Это позволяет эффективно управлять различными типами данных и минимизировать задержки за счет оптимального распределения ресурсов [48]. Данная методология также включает схемы динамической расстановки приоритетов трафика на основе анализа

показателей QoS, что делает возможным систематическое распределение данных в очереди по степени важности для минимизации задержек в передаче [49].

Таким образом, внедрение и развитие моделей на основе ИИ предоставляет значительные преимущества для модернизации спутниковых систем связи, что позволяет им соответствовать требованиям сетей будущего, таких как сети 6G, обеспечивая оптимальную производительность и минимальные задержки.

3.2 Особенности передачи данных в спутниковых каналах связи и их влияние на сетевую задержку

Сети спутниковой связи играют ключевую роль в глобальных телекоммуникациях, требуя эффективного управления передачей данных для оптимального использования ограниченных ресурсов. В условиях постоянного роста объема передаваемых данных и необходимости поддержания высокого качества обслуживания (QoS), одним из важных факторов является применение инновационных стратегий, таких как искусственный интеллект (ИИ), для классификации и приоритезации трафика.

Результаты исследований в области передачи данных через спутниковые каналы связи показывают, что использование маршрутизации на основе политик (policy-based routing, PBR) эффективно для управления трафиком, позволяя уменьшить воздействие низкоприоритетных действий пользователей, таких как просмотр веб-страниц или загрузка файлов, на более важные данные [50]. Это способствует повышению QoS и качества восприятия в спутниковых системах связи.

Исследования также подчеркивают важность точной классификации трафика для распределения ресурсов. В работе [51] предложены методы глубокого анализа пакетов (deep packet inspection, DPI) и использования сверточных нейронных сетей для надежной классификации трафика в условиях ограниченных частотно-временных ресурсов спутниковых каналов.

Дополнительно, гибридные архитектуры управления трафиком с применением мультиагентных подходов, таких как Multi-Agent Deep Deterministic Policy Gradient (MADDPG), демонстрируют высокую эффективность в оптимизации управления передачей данных [52]. Эти методы позволяют получать оптимальные стратегии для передачи данных, решая проблемы, связанные с распределенным управлением трафиком в спутниковых сетях.

При этом, применение протоколов маршрутизации, таких как CGR (contact graph routing), также требует дальнейшего совершенствования для удовлетворения требований по задержке и устойчивости к сбоям в спутниковой связи [53]. Эти протоколы направлены на минимизацию задержек и улучшение QoS для различных типов трафика.

Предложенное исследование восполняет пробел в существующих методиках, предлагая всеобъемлющую стратегию классификации и приоритезации трафика. Использование ИИ позволяет прогнозировать параметры трафика и присваивать приоритеты передаваемым данным, что приводит к значительному улучшению QoS и сокращению сетевых задержек. Методика направлена на модернизацию спутниковых систем связи, делая их более адаптивными и эффективными для передачи данных в условиях растущего трафика.

3.3 Методология управления передачей трафика в спутниковых каналах связи на основе алгоритмов ИИ

На рисунок. 3 показаны ключевые этапы первоначального процесса внедрения этого исследования. Сетевой трафик от конечных пользователей проходит через участок, называемый «Определение типа трафика на основе ИИ». На этом участке определяется тип трафика и оценивается его уровень приоритета. Затем трафик сортируется в приоритетной очереди, что оптимизирует распределение ресурсов и повышает QoS на участке спутниковой связи.

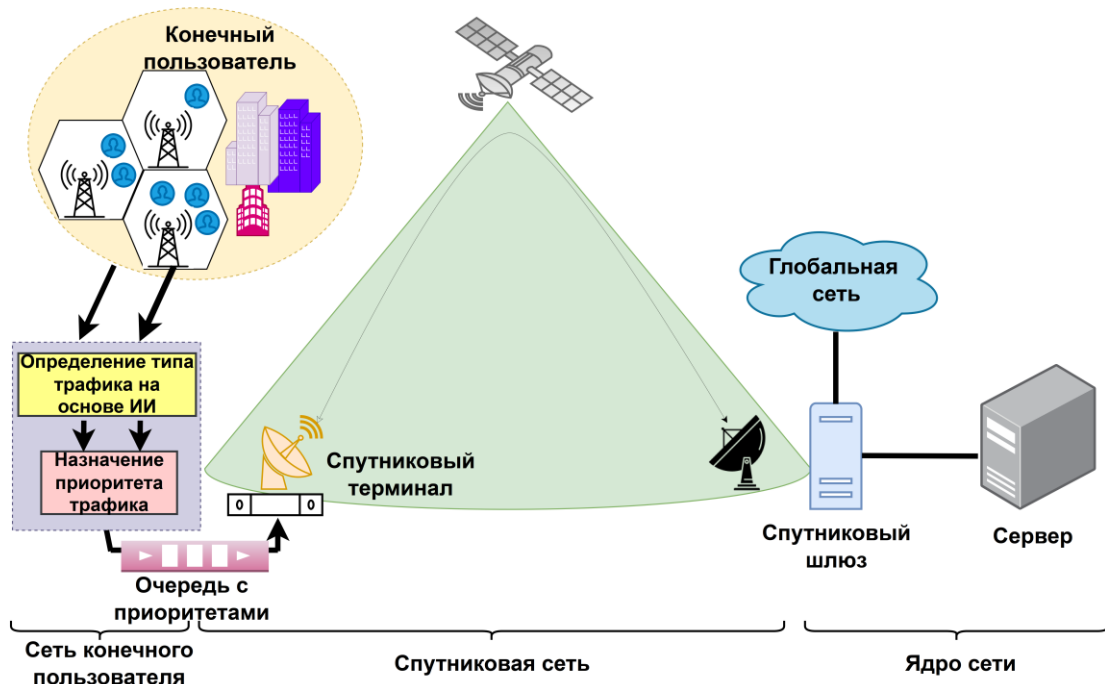


Рисунок 3. Предлагаемая организация гибридной спутниковой сети связи

На рисунок. 4 показан процесс определения типа трафика и назначение ему приоритета для отдельных конечных пользователей.

Разработанная ИИ модель, предназначенная для точного определения типа трафика и присвоения ему соответствующего уровня приоритета, основана на классификации четырех хорошо известных типов трафика из набора данных [54] для повышения QoS при передаче по спутниковой сети. Для этих задач используется нейросетевой алгоритм MLP.

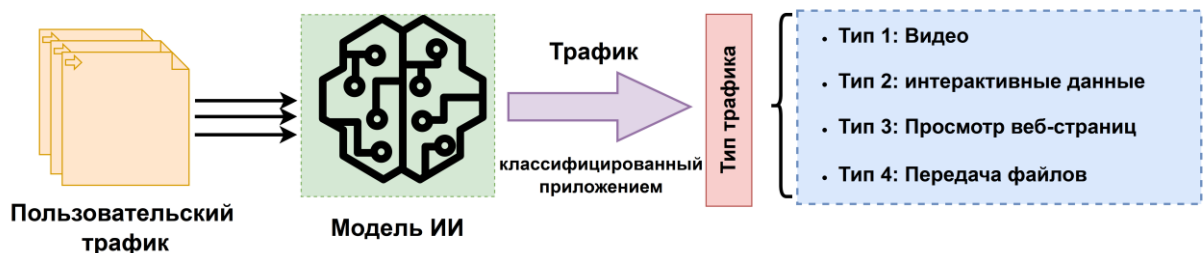


Рисунок 4. Нейросетевая модель определения типа и назначения приоритетов сетевому трафику

Набор данных о сетевом трафике содержит файлы формата PCAP, содержащие извлеченные параметры переданных пакетов, которые классифицированы по четырем типам передаваемого трафика: файлы большого объема, потоковое видео, веб-страницы и данные приложений реального времени. На этапе обучения модели ИИ будет использоваться информация из набора

данных [54] с использованием модели MLP для определения типа трафика [55]. MLP – это искусственная нейронная сеть с входным, скрытым и выходным слоями, которая позволяет изучать сложные взаимосвязи между входными данными. Математическое представление каждого слоя описывается следующим уравнением:

$$O_i^{(l)} = \varphi(u_i^{(l)}) = \varphi\left(\sum_{j=1}^{n_l} O_j^{(l-1)} w_{j,i}^{(l)} + w_{0,i}^{(l)}\right), 1 \leq l \leq L \quad (1)$$

где $\varphi(\cdot)$ - функция активации, которая играет решающую роль в определении выходных данных нейронной сети, обычно она представляет собой нелинейную касательную гиперболическую функцию для скрытых слоев, в то время как для выходного слоя используется линейная функция; L - число слоев нейронной сети, не являющихся входными; l - индекс реального слоя; $O_i^{(l)}$ - выходные данные нейрона i в слое l ; $w_{j,1}^{(l)}$ - веса, связанные со связями между нейронами в соседних слоях; $w_{0,i}^{(l)}$ - коэффициент, который соответствует смещению для нейрона i в слое l . Конечный результат работы нейронной сети обозначается $O^{(L)} = y$, где L представляет индекс конечного слоя, а n_L - его длину.

Приоритет трафика имеет решающее значение в конфигурациях QoS для управления сетью, управления различными типами трафика в зависимости от их конкретных потребностей [56]. Классификация типов трафика и присвоение уровней приоритета позволяет сетевым администраторам внедрять политики QoS, гарантируя получение критически важными приложениями необходимых сетевых ресурсов при эффективной обработке менее важного трафика. Это необходимо для поддержания производительности и надежности сети, особенно в сценариях с ограниченной пропускной способностью или перегрузкой сети спутниковой связи.

Уровни приоритета соответствуют следующим типам трафика: сетевое управление с уровнем приоритета 1 для основных протоколов управления сетью, трафик критически значимых коммуникаций с ультрамалыми задержками

(тактильный интернет, дополненная реальность, дискретная автоматика и т.п.) с приоритетом 2, трафик мультимедиаконференц-связи с приоритетом 3 для голосового и видеотрафика с низкой задержкой, потоковое вещание (приоритет 4) для передачи аудио и видео в реальном времени, критический трафик (приоритет 5) критически важных приложений, приоритетный трафик (приоритет 6) для важных, но менее критичных приложений, и фоновый режим (приоритет 7) для задач с более низким приоритетом (передача файлов и т.п.) [57].

На рисунок. 5 наглядно представлено распределение уровней приоритета трафика в сети, которое является важным аспектом оптимизации для обеспечения эффективности спутниковой связи. После сортировки трафика следующим ключевым шагом является обеспечение приоритетной передачи через сети спутниковой связи.



Рисунок 5. Присвоение уровней приоритета для сетевого трафика

В нашем исследовании модель была обучена четырем типам трафика, выдавая уровни приоритета передаваемым пакетам на основе обучающего набора данных, взятых из работы. Трафик приложения для передачи видео классифицируется как потоковое вещание с уровнем приоритета 4, что подчеркивает важность стабильной потоковой передачи видео. Критически важному трафику приложений, связанному с интерактивными данными, присвоен приоритет 5, что подчеркивает его жизненно важную роль в сценариях, требующих своевременного и надежного обмена данными. Трафик пользовательской передачи данных является фоновым с приоритетом 7, что не отменяет его роль для эффективной передачи больших массивов данных по спутниковым каналам связи. Трафик, генерируемый при просмотре веб-страниц, которому присваивается приоритет 6, имеет решающее значение для обеспечения

интерактивного взаимодействия с пользователем, особенно в приложениях, требующих отображения информации в режиме реального времени.

3.4 Обучение модели искусственного интеллекта для классификации сетевого трафика в спутниковых системах

На экспериментальном этапе для создания надежной и адаптируемой модели ИИ была использована реализация MLP с применением фреймворка Keras, который осуществляет вычисления в соответствии с формулой (1). Модель MLP обучалась на различных обучающих наборах данных о сетевом трафике, в которые был добавлен контрольный набор данных.

Для численной оценки качества нейросетевых алгоритмов модели используем простой классификатор — точность, который соответствует доле пакетов, по которым разработанный классификатор принял правильное решение, от общего числа пакетов в контрольных обучающих данных. Точность выражается в виде безразмерного коэффициента величиной от 0 до 1, где значение $>0,9$ принимается за приемлемый уровень точности.

Рисунок. 6 иллюстрирует ключевой аспект результатов 150 экспериментов, отображая значения точности обучаемой модели MLP в зависимости от того, сколько итераций происходило во время обучения модели. В каждом эксперименте количество итераций обучения модели равнялось порядковому номеру эксперимента.

Примечательно, что для экспериментов с 83 и более итерациями точность тестирования на контрольном наборе данных превышает 90% (за исключением точечного результата неудачного эксперимента при обучении модели со 129 итерациями, где была показана точность 0,87), что подчеркивает работоспособность модели для достаточно точного распознавания типа передаваемых пакетов через моделируемый спутниковый канал передачи данных. Однако присутствие даже одного неудачного эксперимента говорит о том, что для

использования в реальных системах классификации трафика набор обучающих данных следует увеличить, что требует дальнейшего исследования.

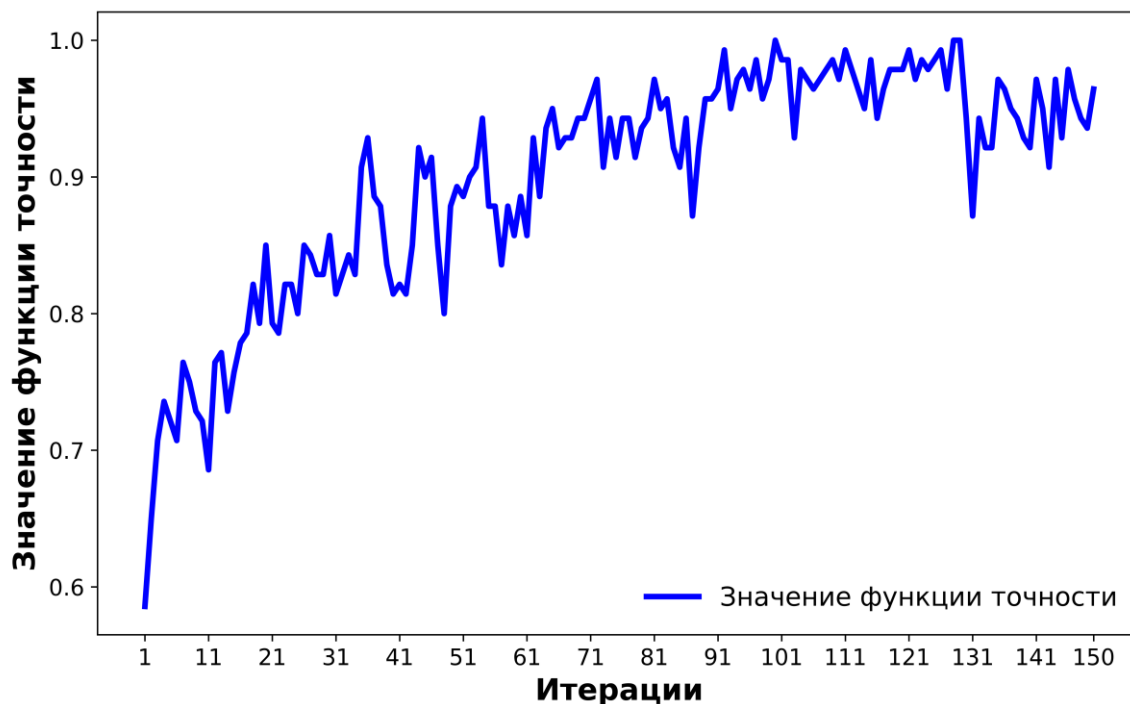


Рисунок 6. Классификатор точности для обучающих данных модели на основе ИИ

Полученная точность предложенной модели в целом соответствует верному распознаванию типа трафика с точностью более 90% и последующему присвоению ему уровня приоритета при обучении модели не менее чем с 87 итерациями на используемом наборе данных для обучения и контроля.

3.5 Математическая модель приоритизации трафика в спутниковых каналах связи на основе системы М/М/1

Центральной идеей предложенного подхода является переход от статической к динамической, интеллектуальной приоритизации трафика. Для этого была разработана двухуровневая модель: на верхнем уровне модель искусственного интеллекта на основе MLP выполняет проактивное прогнозирование, а на нижнем уровне аналитическая формула использует этот прогноз для расчета конкретного значения приоритета.

Для решения задачи снижения сетевой задержки предложена комплексная модель, основанная на взаимодействии трех ключевых компонентов: интеллектуального модуля прогнозирования, аналитической формулы расчета приоритета и теоретического аппарата для оценки эффективности. Такой подход позволяет перейти от реактивного к проактивному управлению трафиком.

В математических представлениях будем использовать следующие обозначения: T_i – i -й тип трафика, передаваемый через сети спутниковой связи, где $i = 1, 2, \dots, m$; p_j – j -й уровень приоритета, где $j = 1, 2, \dots, n$; x_i – входные параметры, соответствующие типу трафика T_i ; y_i – прогнозируемый тип трафика x_i , полученный из модели MLP; r_i – эффективная скорость передачи пакетов трафика типа T_i ; l_i – допустимая задержка для трафика типа T_i ; u_i – уровень срочности (параметр, который принимает целочисленное значение от 0 до i в порядке убывания приоритета для трафика типа T_i); B – доступная полоса пропускания сети (пакеты/с); D – возникшая задержка при передаче в сети; L – коэффициент потери пакетов (%), который определяет долю пакетов, отбрасываемых сетью во время передачи.

В модели MLP функция прогнозирования представлена в виде

$$y_i = f_{\text{MLP}}(x_i) \quad (2)$$

где f_{MLP} – лимит модуля MLP, который определяет тип трафика на основе входных характеристик x_i . Функции отображения f_1, f_2 и f_3 определяются следующим образом:

$$\begin{aligned} f_1(r_i, B) &= \begin{cases} \frac{r_i}{B}, & \text{если } r_i \leq r_{\text{th}}; \\ 1 + \xi \frac{(r_i - r_{\text{th}})}{(B - r_{\text{th}})}, & \text{если } r_i > r_{\text{th}}; \end{cases} \\ f_2(d_i, D) &= \frac{1}{1 + e^{\alpha(D - d_i)}}; \\ f_3(l_i, L) &= 1 - e^{-\beta(L - l_i)}; \end{aligned} \quad (3)$$

где r_{th} – выделяемая полоса пропускания, доступная для передачи трафика; $f_1(r_i, B)$ моделирует требования к скорости передачи данных и соотношение доступной полосы пропускания; $f_2(d_i, D)$ представляет экспоненциальную

зависимость между допустимой задержкой и текущей задержкой сети; $f_3(l_i, L)$ моделирует зависимость допустимых потерь к текущему уровню потери пакетов. Параметры α, β и ξ должны быть тщательно выбраны на основе конкретных характеристик передаваемого трафика в каналах спутниковой связи и желаемого поведения системы определения приоритета передаваемого типа трафика [58].

Значение α определяет экспоненциальный спад допустимой задержки передачи пакетов при увеличении сетевой задержки, при этом более высокие значения α приводят к более выраженному снижению. Выбор α зависит от нормируемой величины задержки для конкретного сетевого приложения, которое и генерирует передаваемый трафик.

Значение β определяет экспоненциальное снижение допустимой потери пакетов при увеличении уровня потерь. Подобно параметру α , более высокие значения β увеличивают скорость снижения допустимых потерь пакетов. Оптимизация β зависит от устойчивости приложения к потере пакетов.

Значение ξ формирует нелинейную зависимость между требуемой скоростью передачи данных и доступной полосой пропускания. Выбор ξ зависит от конкретных требований приложения к скорости передачи данных при ограниченной полосе пропускания канала связи [58].

Таким образом, предложенная система использует модель ИИ (MLP) для получения интеллектуального прогноза о характере трафика u_i . Этот прогноз затем интегрируется как ключевой компонент в аналитическую формулу для расчета динамического уровня приоритета p_i .

Уровень приоритета для каждого типа трафика рассчитывается по следующей аналитической формуле:

$$p_i = w_1 f_1(r_i, B) + w_2 f_2(d_i, D) + w_3 f_3(l_i, L) + w_4 u_i + w_5 y_i \quad (4)$$

где веса $w_1, \dots, w_5$ соответствуют относительному влиянию каждого компонента (требований к скорости передачи данных, величине задержки, величине потерь, уровню срочности и определяемому типу трафика) при определении уровня приоритета для передаваемого трафика.

Данная формула представляет собой структуру для принятия решений, объединяющую как традиционные сетевые параметры (скорость, задержка), так и интеллектуальный компонент. Ключевую роль в этой формуле играет компонент y_i – прогнозируемый тип трафика, который генерируется предварительно обученной моделью MLP. Без этого компонента, формула p_i была бы лишь статическим набором правил. Именно компонент y_i придает системе способность адаптироваться к изменяющемуся в реальном времени характеру трафика, превращая её из реактивной в проактивную систему. Таким образом, ИИ не заменяет формулу, а обеспечивает её интеллектуальным входом, без которого динамическая приоритизация невозможна.

Средняя задержка D_{avg} вычисляется как средняя задержка передачи пакетов, вычисляемая по формуле:

$$D_{avg} = \frac{1}{N} \sum_{i=1}^N D_i \quad (5)$$

где D_i – задержка i -го пакета; N – общее количество пакетов.

При оценке эффективности предлагаемого решения для приоритизации трафика [59, 60] были рассмотрены следующие модели:

- «Первым пришел – первым обслужен» (First Come First Served, FCFS) – обеспечивает приоритизацию пакетов на основе порядка их поступления без учета типа или приоритета трафика (эта модель выполняет роль базового алгоритма для сравнения).
- Очередь с приоритетами (Priority Queueing, PQ) – использует статическое присвоение уровня приоритета всем входящим пакетам.
- Взвешенный алгоритм равномерного обслуживания (Weighted Fair Queueing, WFQ) – использует пропорциональные веса для разных типов трафика (его можно предложить для сравнения моделей по уровню адаптивности).
- Произвольное раннее обнаружение (Random Early Detection, RED) – это модель с подходом к управлению очередью путем упреждающего уничтожения пакетов для предотвращения перегрузки, основываясь на статистической вероятности перегрузки канала передачи данных.

– Циклический алгоритм (Round Robin, RR) – обрабатывает пакеты в циклическом порядке, предоставляя отправную точку для сравнения моделей по влиянию приоритизации трафика, исходя из параметров сети связи.

Теперь сравним значения средней задержки D_{avg} , получаемые с помощью предлагаемой методики приоритизации трафика и с помощью методов, перечисленных выше.

Для предложенной методики приоритизации и доставки трафика примем для i -го типа трафика скорость обслуживания, равной μ_i . Как только определяется уровень приоритета p_i для каждого типа трафика, предложенная схема сможет использовать эту информацию для корректировки скорости обслуживания μ_i для каждого типа трафика i . Скорость обслуживания μ_i представляет собой скорость (пакеты/с), с которой терминал спутниковой связи обрабатывает пакеты трафика типа i . Типам трафика с более высоким уровнем приоритета [61] (более высокий p_i) может быть присвоена более высокая скорость обслуживания μ_i , что означает, что наземная станция обрабатывает и передает эти пакеты быстрее, ведь трафик более высокого приоритета должен испытывать меньшие задержки, чтобы соответствовать требованиям QoS. И наоборот, типам трафика с более низкими уровнями приоритета (более низкий p_i) может быть назначена более низкая скорость обслуживания μ_i , что приводит к более низкой скорости обработки и потенциально более высокой задержке для этих пакетов. Скорость обслуживания μ_i в предложенной схеме регулируется на основе формулы

$$\mu_i = B(1 + p_i AF) \quad (6)$$

где B - доступная полоса пропускания в сети (пакетов/с); AF (Adjustment Factor) - коэффициент, используемый при приоритизации трафика для изменения скорости обслуживания исходя из уровня приоритета.

Для теоретического анализа и наглядной демонстрации эффективности предложенного механизма приоритизации воспользуемся классическим аппаратом теории массового обслуживания. Важно подчеркнуть, что модель М/М/1 (и далее М/М/1/К) используется в данной работе не для точного

моделирования сложной и динамичной спутниковой сети, а исключительно как аналитический инструмент. Он позволяет формально доказать, что увеличение скорости обслуживания μ_i , управляемое предлагаемым механизмом, принципиально ведет к снижению задержки и потерь.

После определения скорости передачи μ_i на основе уровня приоритета p_i (6) можно рассчитать задержку D_i , испытываемую каждым пакетом трафика типа i , используя формулу модели массового обслуживания M/M/1:

$$D_i = \frac{1}{\mu_i - \lambda} \quad (7)$$

где λ – интенсивность входного потока пакетов (параметр нагрузки); μ_i – скорость обслуживания для трафика типа i . Более высокая скорость обслуживания μ_i приводит к меньшей задержке D_i , в то время как более низкая скорость обслуживания приводит к более высокой задержке.

Пусть $\mu_{\text{proposed},i}$ – скорость обслуживания для трафика типа i в предложенной модели, $\mu_{\text{other},i}$ – скорость обслуживания для трафика типа i в схеме обслуживания FCFS, а $D_{\text{proposed},i}$ и $D_{\text{other},i}$ – задержки для трафика типа i в предложенной модели и модели FCFS соответственно.

Учитывая, что более высокая скорость обслуживания приводит к меньшей задержке, имеем

$$\mu_{\text{proposed},i} > \mu_{\text{other},i} \Rightarrow D_{\text{proposed},i} < D_{\text{other},i} \quad (8)$$

Поэтому если задержки $D_{\text{proposed},i}$ для отдельных типов трафика в целом ниже, чем $D_{\text{other},i}$ то средняя задержка $D_{\text{avg, proposed}}$ предлагаемой схемы также будет ниже $D_{\text{avg, other}}$:

$$\sum_{i=1}^N D_{\text{proposed},i} < \sum_{i=1}^N D_{\text{other},i} \Rightarrow D_{\text{avg, proposed}} < D_{\text{avg, other}} \quad (9)$$

Пусть $\rho_{\text{proposed},i}$ и $\mu_{\text{proposed},i}$ – соответственно коэффициент загрузки и скорость обслуживания для трафика типа i в предложенной модели; $\rho_{\text{other},i}$ и $\mu_{\text{other},i}$ – коэффициент загрузки и скорость обслуживания для любого другого типа трафика, кроме i . Коэффициент загрузки канала ρ представляет собой долю времени, в течение которой передающее оборудование занято обслуживанием

передачи пакетов. Когда система работает в режиме, близком к полной загрузке ($\rho \rightarrow 1$), очередь пакетов будет расти, увеличивая вероятность их потери из-за переполнения буфера. В этой ситуации для более ясного сравнения используется модель М/М/1/К, которая эквивалентна модели массового обслуживания М/М/1, за исключением того, что очередь конечна и имеет длину $(K - 1)$, где K - емкость буфера очереди. Коэффициент загрузки для трафика типа i , обозначенный как ρ_i , в очереди М/М/1/К определяется формулой:

$$\rho_i = \frac{\lambda}{\mu_i} \quad (10)$$

где λ - скорость поступления пакетов.

Учитывая, что более высокая скорость обслуживания μ_i приводит к более низкому коэффициенту загрузки ρ_i , имеем:

$$\mu_{\text{proposed},i} > \mu_{\text{other},i} \Rightarrow \rho_{\text{proposed},i} < \rho_{\text{other},i} \quad (11)$$

В очереди М/М/1/К коэффициент потери пакетов L_{rate} определяется следующей формулой:

$$L_{\text{rate}} = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}} \quad (12)$$

Более низкий коэффициент загрузки подразумевает более низкий коэффициент потери пакетов:

$$\rho_{\text{proposed},i} < \rho_{\text{other},i} \Rightarrow L_{\text{rate, proposed},i} < L_{\text{rate, other},i} \quad (13)$$

Средний коэффициент потери пакетов $L_{\text{rate, avg}}$ для каждого типа трафика рассчитывается следующим образом:

$$L_{\text{rate, avg}} = \frac{1}{N} \sum_{i=1}^N L_{\text{rate},i} \quad (14)$$

Из формулы (14) следует, что если коэффициент потери пакетов $L_{\text{rate, proposed},i}$ для отдельных типов трафика в целом будет ниже, чем $L_{\text{rate, other},i}$ то средний коэффициент потери пакетов $L_{\text{rate, avg, proposed}}$ для конкретного типа трафика также будет ниже, чем $L_{\text{rate, avg, other}}$:

$$\sum_{i=1}^N L_{\text{rate, proposed},i} < \sum_{i=1}^N L_{\text{rate, other},i} \quad (15)$$

$$\Rightarrow L_{\text{rate, ave, proposed}} < L_{\text{rate, ave, other}} \quad (16)$$

Используя неравенства (15) и (16) и учитывая взаимосвязь между скоростью обслуживания, коэффициентом загрузки и коэффициентом потери пакетов, можно математически вывести, что предложенная модель обладает потенциалом для достижения более низкого среднего коэффициента потери пакетов по сравнению с базовой моделью, если средний коэффициент потери пакетов для каждого типа трафика будет ниже среднего коэффициента потери пакетов для базовой модели FCFS.

3.6 Математический эксперимент по оценке средней задержки передачи пакетов в спутниковой сети

Сравним средние задержки, полученные с помощью предлагаемой методики приоритизации трафика и базового метода FCFS, описанного в предыдущем разделе.

В предлагаемой методике используются конкретные параметры для содержательного анализа: скорость поступления λ составляет 5 пакетов /с, базовая скорость обслуживания $\mu_{\text{base-rate}}$ зафиксирована на уровне 12 пакетов /с, а корректирующий коэффициент AF выбран равным 0,2 как рекомендуется для высоконагруженных спутниковых каналов связи [58]. Все эти параметры включены в экспериментальную реализацию модели.

Стабильность системы имеет решающее значение для оценки времени задержки и расчета коэффициента потери пакетов, поскольку для поддержания сбалансированной работы операционной среды требуется скорость обслуживания, превышающая скорость поступления пакетов. Постараемся оценить эффективность и надежность предлагаемых схем в соответствии с установленными критериями путем проведения экспериментов.

В предлагаемой модели важна приоритизация типов трафика. Категоризация трафика позволяет присваивать значения приоритета p_i для адаптации реагирования модели. Мы создали иерархию из четырех уровней

приоритетов трафика, где большие значения указывают на более высокие уровни приоритета для соответствующих тегов трафика, упрощая практическую реализацию модели.

Эта стратегия определения приоритетов присваивает различные уровни приоритета различным типам трафика. Самому распространенному трафику передачи файлов присваивается значение приоритета $p_{\text{bulk-data}}$, равное 1, что указывает на более низкий приоритет. Трафику просмотра веб-страниц присваивается значение приоритета $p_{\text{web-browsing}}$, равное 2, что означает более высокий приоритет, чем для трафика передачи файлов. Трафику передачи интерактивных данных управления и мониторинга присвоено значение приоритета $p_{\text{interactive-data}}$, равное 3. Видеотрафику присваивается наивысший приоритет $p_{\text{video-traffic}}$ со значением 4.

Для предлагаемой модели рассчитывается скорость обслуживания μ_i для каждого типа трафика (передача файлов, просмотр веб-страниц, передача интерактивных данных и видеотрафик) по формулам

$$\mu_{\text{bulk-data}} = \mu_{\text{base-rate}} + AF \cdot p_{\text{bulk-data}} = 12,2 \text{ пакетов/с} \quad (17)$$

$$\mu_{\text{web-browsing}} = \mu_{\text{base-rate}} + AF \cdot p_{\text{web-browsing}} = 12,4 \text{ пакетов/с} \quad (18)$$

$$\mu_{\text{interactive-data}} = \mu_{\text{base-rate}} + AF \cdot p_{\text{interactive-data}} = 12,6 \text{ пакетов/с} \quad (19)$$

$$\mu_{\text{video-traffic}} = \mu_{\text{base-rate}} + AF \cdot p_{\text{video-traffic}} = 12,8 \text{ пакетов/с} \quad (20)$$

Для метода FCFS не назначается приоритет каждому типу трафика, поэтому $p_{\text{no-assign-priority}} = 0$. В этом случае значение скорости обслуживания рассчитывается следующим образом:

$$\mu_{\text{baseline}} = \mu_{\text{base-rate}} + AF \cdot p_{\text{no-assign-priority}} = 12 \text{ пакетов/с} \quad (21)$$

Для базового метода FCFS среднее время задержки, измеряемое в секундах, рассчитывается следующим образом:

$$D_{\text{avg-baseline}} = \frac{1}{(\mu_{\text{baseline}} - \lambda)} = 0,1429 \text{ с} \quad (22)$$

Для предложенной методики будем рассчитывать среднее время задержки следующим образом:

$$D_{\text{arg-bulk-data}} = \frac{1}{(\mu_{\text{bulk-data}} - \lambda)} = 0,1389 \text{ с} \quad (23)$$

$$D_{\text{arg-web-browsing}} = \frac{1}{(\mu_{\text{web-browsing}} - \lambda)} = 0,1351 \text{ с} \quad (24)$$

$$D_{\text{avg-interactive-data}} = \frac{1}{(\mu_{\text{interactive-data}} - \lambda)} = 0,1316 \text{ с} \quad (25)$$

$$D_{\text{arg-video-traffic}} = \frac{1}{(\mu_{\text{video-traffic}} - \lambda)} = 0,1282 \text{ с} \quad (26)$$

Результаты расчетов указывают на видимую тенденцию: определение приоритетов для разных типов трафика снижает среднюю задержку. Для трафика передачи файлов средняя задержка снижается на 2,8%, в то время как для трафика с высоким приоритетом, таким как видеоданные, задержки сокращаются уже значительно - на 10,3%. Трафик просмотра веб-страниц и трафик интерактивных данных показывают снижение задержки на 5,5% и 7,9% соответственно. Таким образом, присваивание приоритетов разным типам трафика может использоваться для оптимизации производительности спутниковой системы связи и повышения эффективности передачи данных.

Анализ предложенной модели обслуживания трафика на основе ИИ можно интегрировать с математическим моделированием для всестороннего сравнения результатов. Можно использовать популярную модель массового обслуживания М/М/1, скорость поступления λ – 5 пакетов/с и фиксированную базовую скорость обслуживания $\mu_{\text{base-rate}}$ 12 пакетов/с. В экспериментальном моделировании передавалось 20000 пакетов, а значения скоростей обслуживания, как для базового метода FCFS, так и для предлагаемой модели определялись с помощью выражений (17)-(20).

На рисунок 7 показано сравнение среднего времени задержки, полученного в результате расчетов по математическим формулам (17)-(20) и нейросетевого моделирования для различных сценариев передачи трафика. В базовом случае использования метода FCFS среднее время задержки по формуле (22) составляет 0,1429 с, что точно отражает среднее время задержки, полученное при моделировании с использованием модели М/М/1 и формулы (7), равное 0,1434 с. Введение назначения приоритетов для различных типов трафика

приводит к минимальным отклонениям. Для низкоприоритетного трафика передачи файлов среднее время задержки при вычислении по формуле (22) и при моделировании составляет 0,1389 с и 0,1390 с соответственно.

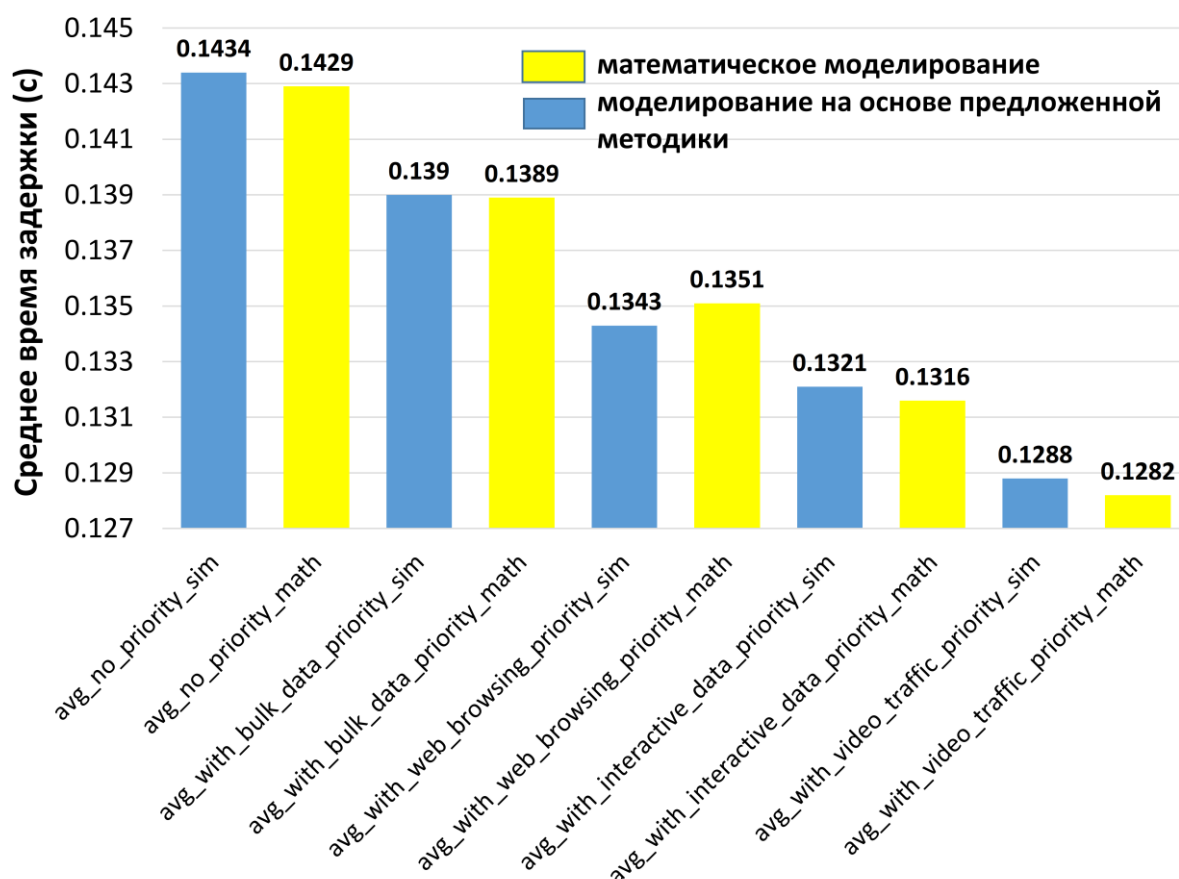


Рисунок 7. Расхождение вычисленного среднего времени задержки передачи пакетов между математическим и имитационным моделированием

Аналогично для трафика просмотра веб-страниц время составляет 0,1351 с и 0,1343с; для интерактивного трафика телеметрии – 0,1316 с и 0,1321 с соответственно. Даже для видеотрафика с присвоенным максимальным уровнем приоритета среднее время задержки показывает незначительные различия: 0,1282 с при расчете по формулам (17)-(20) и 0,1288 с при моделировании. Незначительное отклонение, составляющее 0,05%, подчеркивает точность моделирования, подтверждая его эффективность при прогнозировании и оценке среднего времени задержки для различных сценариев передачи данных.

На рисунок. 8 сравнивается среднее время задержки, полученное при моделировании с установкой приоритетности передаваемых данных и в сценариях без присвоения приоритета.

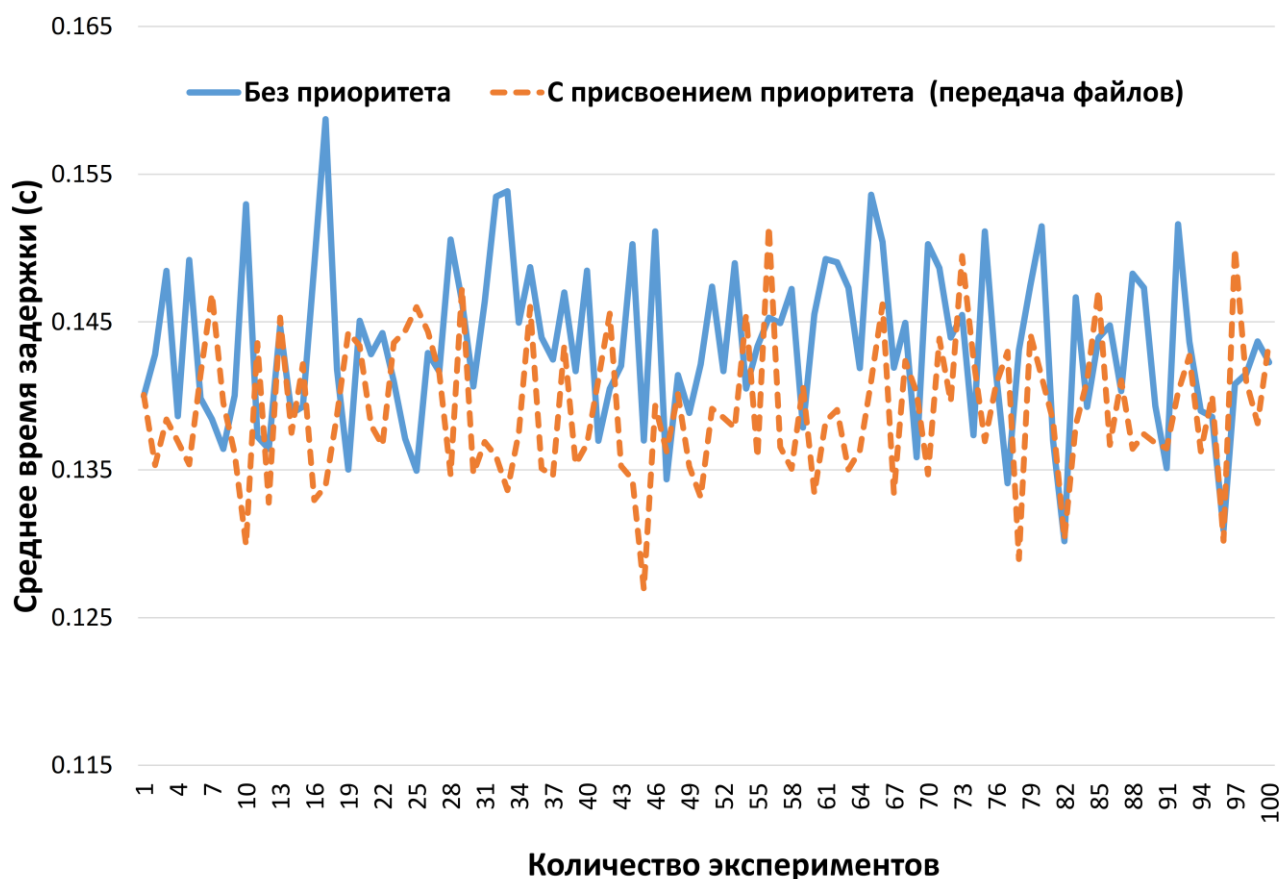


Рисунок 8. Среднее время задержки пакетов при передаче файлов

Данные получены в результате 100 экспериментов повторения процесса моделирования, они свидетельствуют о значительном снижении средней задержки при установке приоритета передаваемым данным по сравнению со сценариями без присвоения приоритета трафику. Имитационный эксперимент демонстрирует снижение средней задержки до 3,03%, говоря о положительном влиянии расстановки приоритетов для трафика на производительность даже при передаче массового низкоприоритетного трафика в исследуемой системе. Сравнение результатов моделирования с математически рассчитанным средним временем задержки выявляет минимальное несоответствие, что указывает на согласованность процесса моделирования и эксперимента, основанного на предложенной методике классификации трафика.

На рисунок. 9 показано сравнение среднего времени задержки, подчеркивающее влияние приоритизации согласно предложенной модели на передачу трафика просмотра веб-страниц по сравнению со сценариями без

приоритизации трафика. Представлены результаты 100 проведенных экспериментов.

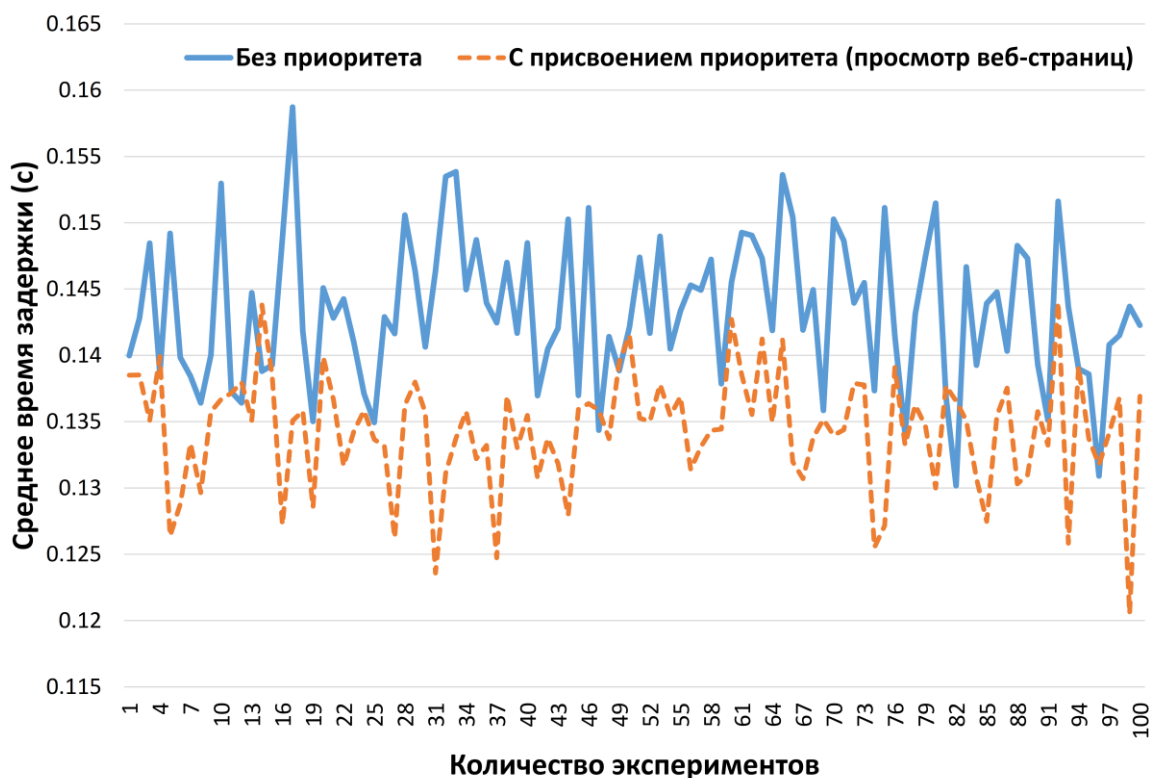


Рисунок 9. Среднее время задержки пакетов при просмотре веб-страниц

Определение приоритетности трафика просмотра веб-страниц сокращает среднюю задержку по сравнению со сценариями без определения приоритетов. В отличие от низкоприоритетного трафика передачи файлов, трафик просмотра веб-страниц демонстрирует более существенное улучшение — средняя задержка при моделировании снижается в среднем на 5,75% по формуле (5).

На рисунок. 10 представлено сравнение среднего времени задержки, подчеркивающее влияние приоритизации трафика интерактивных данных по сравнению со сценариями без его приоритизации. Результаты показывают существенное сокращение средней задержки при приоритизации трафика интерактивных данных — на 7,86%, что превосходит улучшение, наблюдаемое при доставке низкоприоритетного трафика передачи файлов.

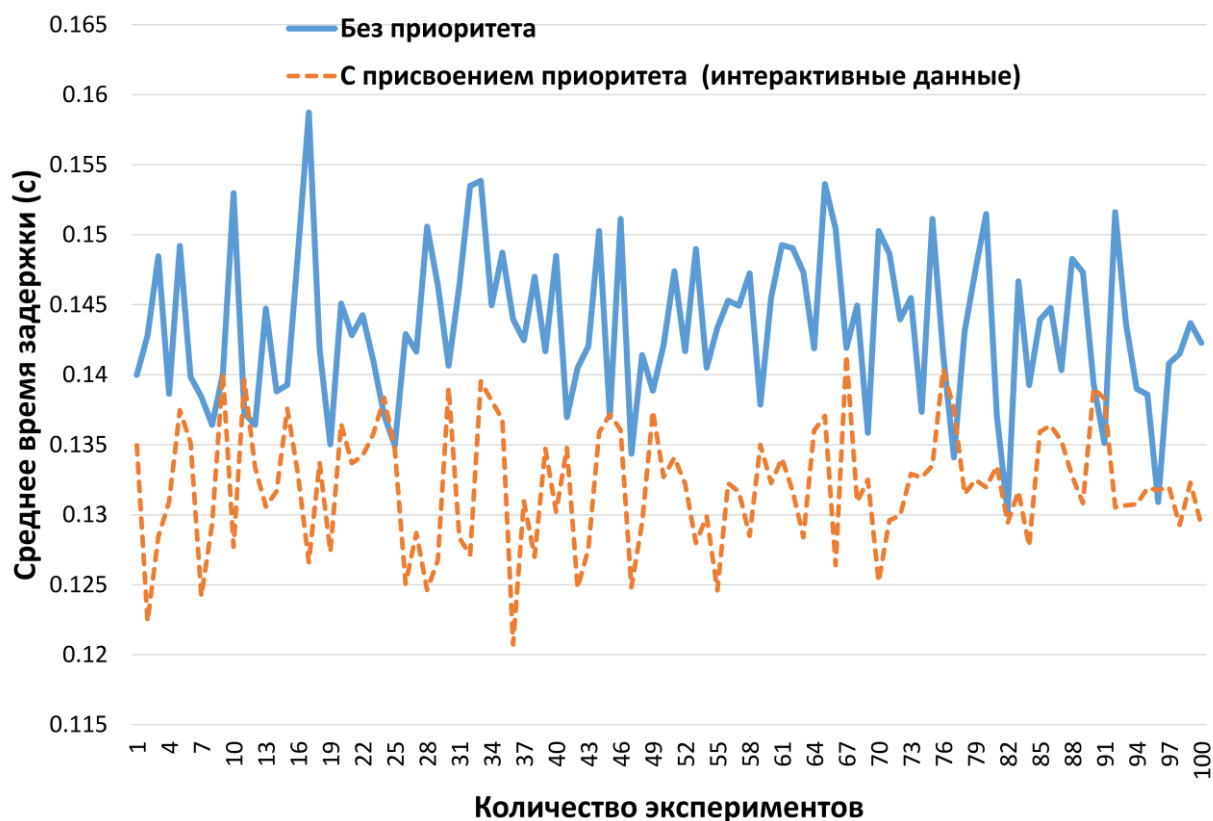


Рисунок 10. Среднее время задержки пакетов при передаче интерактивных данных

На рисунок. 11 сравнивается среднее время задержки при приоритизации видеотрафика по сравнению с передачей данных без приоритизации трафика. Результаты указывают на значительное сокращение средней задержки при приоритизации видеотрафика — на 10,18%, что подчеркивает существенное влияние предложенной методики приоритизации трафика на снижение задержки передачи высокоприоритетного видеотрафика пакетов при высокой загруженности канала. При этом средняя задержка низкоприоритетного трафика, пусть и менее значительно, но тоже снижается — на 3,03%.

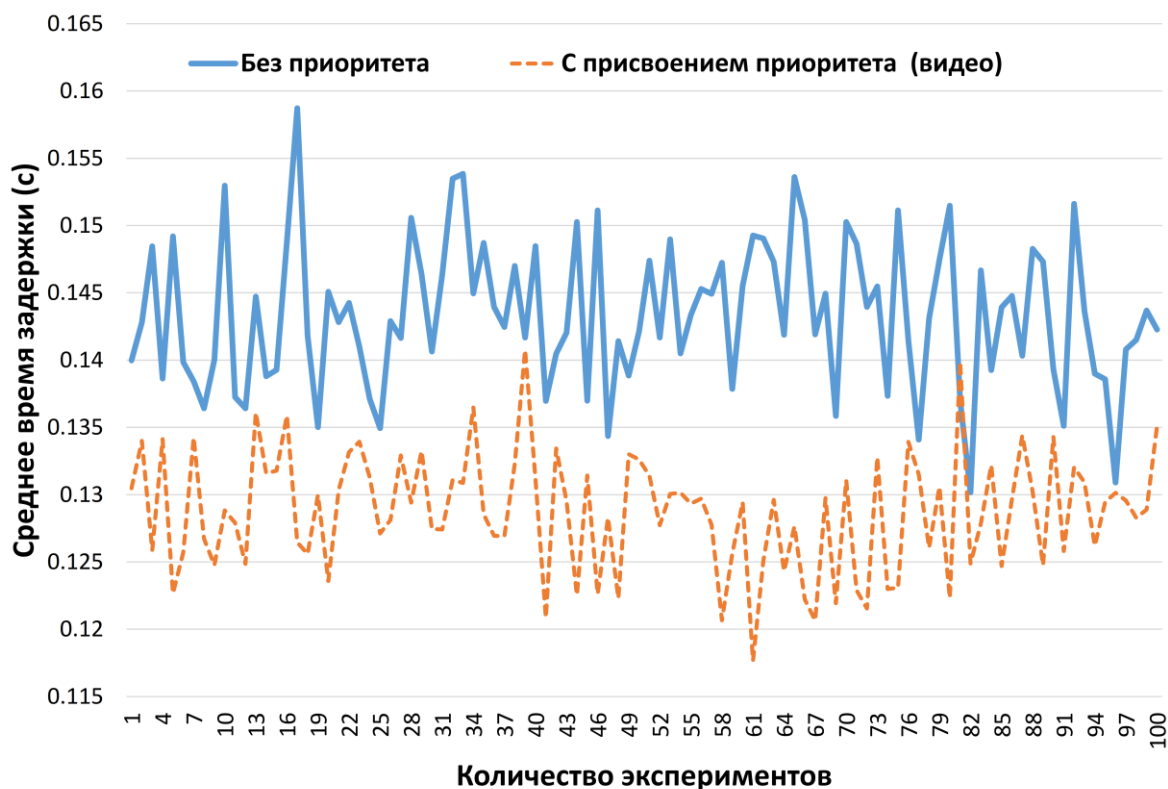


Рисунок 11. Среднее время задержки пакетов при передаче видеотрафика

3.7 Математический эксперимент по определению коэффициента потери пакетов при различных параметрах нагрузки

Проведём расчет и сопоставление коэффициентов потери пакетов между предложенной методикой приоритизации трафика и базовым методом обслуживания FCFS, применительно модели M/M/1/K.

Значения коэффициента загрузки канала для различных типов трафика вычисляются с использованием формул (10) и (17)-(20). Для наименее приоритетного трафика передачи файлов, например, $\rho_{\text{bulk-data}} = \lambda / \mu_{\text{bulk-data}} = 0,4098$. Аналогично для трафика просмотра веб-страниц, интерактивного трафика телеметрии и наиболее высокоприоритетного видеотрафика значения коэффициентов загрузки составляют: $\rho_{\text{web-browsing}} = 0,4032$; $\rho_{\text{interactive-data}} = 0,3968$; $\rho_{\text{video-traffic}} = 0,3906$. Эти значения являются ключевыми показателями для оценки эффективности сети применительно к предложенной схеме присвоения

приоритета передаваемым пакетам. В базовых схемах без приоритета для конкретного трафика базовое значение коэффициента загрузки из формулы (10) составляет $\rho_{\text{baseline}} = \lambda / \mu_{\text{baseline}} = 0,4167$.

Для базовых методов приоритет каждому типу трафика не назначается и коэффициент потери пакетов рассчитывается по формуле (12):

$$L_{\text{rate-avg-baseline}} = \frac{(1 - \rho_{\text{baseline}}) \rho_{\text{baseline}}^K}{1 - \rho_{\text{baseline}}^{K+1}} = 0,0178 \quad (27)$$

Для предложенной методики коэффициент потери пакетов рассчитывается для каждого типа трафика. Для расчета емкость буфера очереди K принята равной 4. Коэффициент потери пакетов наименее приоритетного трафика передачи файлов $L_{\text{rate-avg-bulk-data}}$, трафика просмотра веб-страниц $L_{\text{rate-avg-web-browsing}}$, интерактивного трафика телеметрии $L_{\text{rate-avg-interactive-data}}$ и наиболее высокоприоритетного видеотрафика $L_{\text{rate-avg-video-traffic}}$ согласно (12) будут равны:

$$L_{\text{rate-avg-bulk-data}} = \frac{(1 - \rho_{\text{bulk-data}}) \rho_{\text{bulk-data}}^K}{1 - \rho_{\text{bulk-data}}^{K+1}} = 0,0168 \quad (28)$$

$$L_{\text{rate-avg-web-browsing}} = \frac{(1 - \rho_{\text{web-browsing}}) \rho_{\text{web-browsing}}^K}{1 - \rho_{\text{web-browsing}}^{K+1}} = 0,0159 \quad (29)$$

$$L_{\text{rate-avg-interactive-data}} = \frac{(1 - \rho_{\text{interactive-data}}) \rho_{\text{interactive-data}}^K}{1 - \rho_{\text{interactive-data}}^{K+1}} = 0,0151 \quad (30)$$

$$L_{\text{rate-avg-video-traffic}} = \frac{(1 - \rho_{\text{video-traffic}}) \rho_{\text{video-traffic}}^K}{1 - \rho_{\text{video-traffic}}^{K+1}} = 0,0143 \quad (31)$$

Присвоение уровня приоритета различным типам трафика существенно снижает коэффициент потери пакетов. Важно отметить, что приоритизация даже для низкоприоритетного трафика передачи файлов приводит к снижению коэффициента потери пакетов на 5,62% , улучшая надежность передачи данных, что следует из (27) и (28). Для трафика с высоким приоритетом, таким как видеоданные, коэффициент потери пакетов значительно сокращается: исходя из (27) и (31) - на 19,66%, что подчеркивает эффективность приоритизации в обеспечении целостности видеоданных. Это преимущество распространяется и на другие типы трафика: с заметными уменьшениями для веббраузинга (на 10,67%) и интерактивных данных телеметрии (на 15,17%) (см. формулы (29) и

(30)). В итоге, исходя из неравенств (15) и (16), присваивание приоритета разным типам трафика минимизирует среднюю величину потери пакетов во время передачи данных для всех типов трафика. На рисунке 12 представлено сравнение коэффициентов потери пакетов для математического расчета (math) и предложенного моделирования (sim).

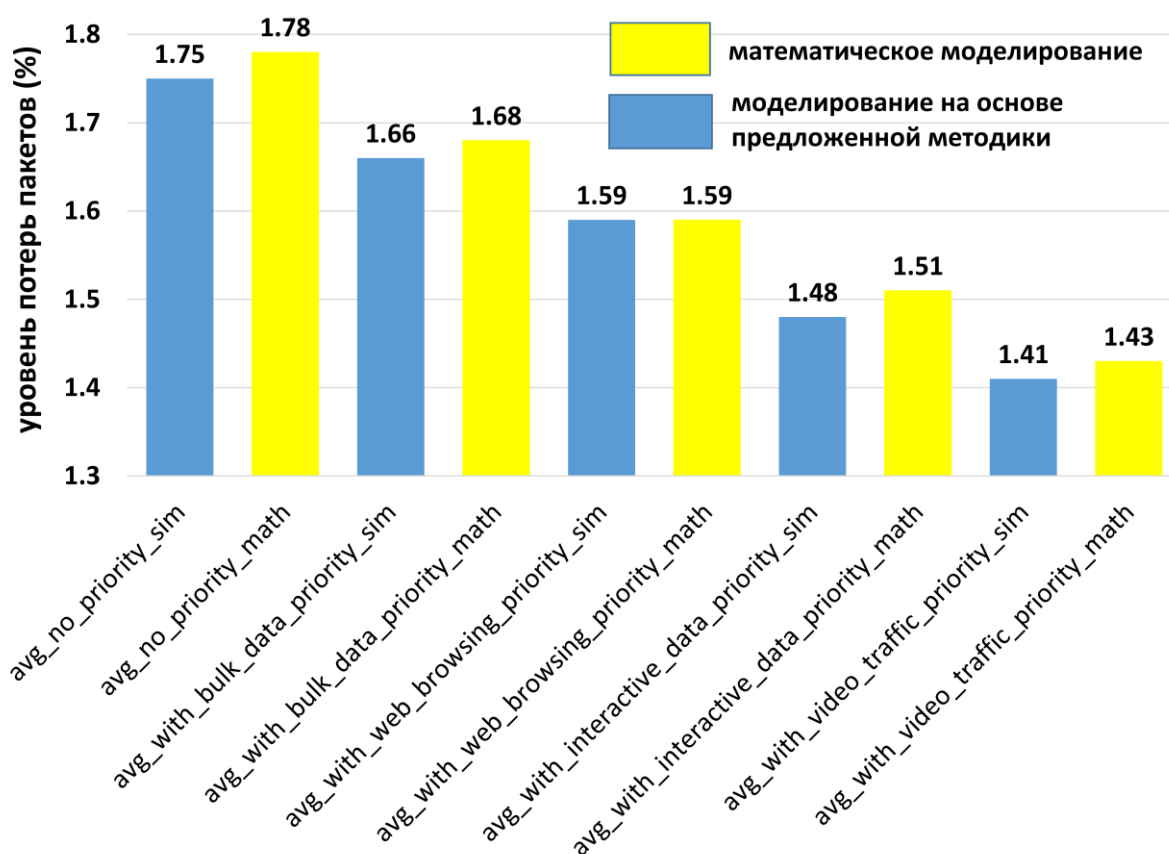


Рисунок 12. Расхождение вычисленного коэффициента потери пакетов между математическим и имитационным моделированием

Анализ проводится для пяти типов трафика: без приоритета (no_priority), передачи файлов (bulk_data), веб-браузинга (web_browsing), интерактивной телеметрии (interactive_data) и видеотрафика с наивысшим приоритетом (video_traffic).

В случае, когда трафику не присваивается приоритет, результаты симуляции показывают коэффициент потери пакетов 1,75%, что близко к вычисленному расчетному значению, согласно (27), равному 1,78%. Это минимальное отклонение подтверждает точность предложенной методики при

сопоставлении с результатами математического расчета и подтверждает её применимость в оценке уровня потери пакетов.

Назначение приоритетов для каждого типа трафика приводит к незначительным различиям между коэффициентами потери пакетов, рассчитанными математически и полученными при моделировании. Для трафика передачи файлов, трафика просмотра веб-страниц, интерактивного трафика телеметрии и видеотрафика расхождения составляют от 0,01% до 0,03%, что подчеркивает точность предложенной имитационной модели для оценки уровня потери пакетов.

Таблица 4 иллюстрирует коэффициенты потери пакетов для различных значений емкости буфера очереди K для случая без присвоения пакетам приоритета (no_priority), для наименее приоритетного трафика передачи файлов (bulk_data), трафика просмотра веб-страниц (web_browsing), интерактивного трафика телеметрии (interactive_data) и наиболее высокоприоритетного видеотрафика (video_traffic). Эти данные подтверждают наличие четкой тенденции: более высокие значения K соответствуют более низким уровням потери пакетов. Это согласуется с ожидаемым результатом, поскольку более высокая пропускная способность буфера K уменьшает вероятность потери пакетов. Обратная зависимость подчеркивает важность выбора адекватного размера емкости буфера передатчика при проектировании сети для повышения общей производительности и минимизации потерь данных.

В модели $M/M/1/K$ с увеличением K уровень потерь пакетов уменьшается. Основываясь на экспериментальных наблюдениях, приведенных в таблице, можно сделать заключение, что уровни потерь пакетов, полученных согласно формуле (12) (math) и в процессе моделирования (sim) по предложенной методике, практически эквивалентны. Относительно небольшое расхождение между значениями коэффициента потери пакетов, полученными путем моделирования и в результате математических расчетов, повышает доверие к

точности модели [8], подтверждая учет факторов, влияющих на величину потерь сетевых пакетов.

Таблица 4. Коэффициент потери пакетов $L_{rate,avg}$ в зависимости от размера емкости буфера κ

Тип моделирования и трафика	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
no_priority_sim	29,1%	10,7%	4,19%	1,75%	0,75%
no_priority_math	29,4%	10,9%	4,35%	1,78%	0,74%
with_bulk_data_priority_sim	28,7%	10,5%	4,03%	1,66%	0,68%
with_bulk_data_priority_math	29,1%	10,6%	4,18%	1,68%	0,68%
with_web_browsing_priority_sim	28,3%	10,3%	3,92%	1,59%	0,67%
with_web_browsing_priority_math	28,7%	10,4%	4,02%	1,59%	0,64%
with_interactive_data_priority_sim	27,9%	10,0%	3,74%	1,48%	0,59%
with_interactive_data_priority_math	28,4%	10,1%	3,86%	1,51%	0,59%
with_video_traffic_priority_sim	27,6%	9,74%	3,58%	1,41%	0,55%
with_video_traffic_priority_math	28,1%	9,88%	3,71%	1,43%	0,56%

3.8 Выводы главы 3

1. Предложенная методика минимизации сетевой задержки с использованием технологий искусственного интеллекта показала высокую эффективность в снижении времени задержки и уменьшении потерь пакетов при передаче данных через спутниковые каналы связи. Применение искусственного интеллекта для классификации трафика и определения приоритетов позволяет значительно повысить качество обслуживания (QoS), особенно в условиях ограниченных сетевых ресурсов. Это доказывает, что предложенный подход превосходит традиционные методы управления трафиком, такие как FCFS,

демонстрируя лучшие результаты в оптимизации работы спутниковых систем связи.

2. В ходе экспериментов было показано, что предложенная модель не только обеспечивает более точные результаты по минимизации задержек, но и успешно проходит проверку на соответствие между математическими расчетами и результатами имитационного моделирования. Это подтверждает надежность методики и её применимость для реальных сценариев спутниковой связи. В дальнейшем развитие этой методики и её интеграция в сети нового поколения, такие как 6G, откроют новые возможности для повышения эффективности и адаптивности спутниковых систем связи.

ГЛАВА 4. МЕТОДЫ МИНИМИЗАЦИИ СЕТЕВОЙ ЗАДЕРЖКИ В ГИБРИДНЫХ СПУТНИКОВЫХ СЕТЯХ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

4.1 Введение: Оптимизация распределения ресурсов в многолучевых спутниковых системах

Спутниковая связь является примером систем с ограниченными ресурсами, которые сталкиваются с ограничениями в спектре, мощности спутников и объёме хранилищ. Эта проблема усугубляется ростом инфраструктур спутниковых сетей на низкой околоземной орбите (LEO). Максимизация эффективности использования спектра, объёма хранилищ и распределения мощности спутников при сохранении качества обслуживания (QoS) является критически важной.

Многолучевая антенна (МБА) [62] предлагает решение проблемы ограниченности ресурсов спутниковой связи, используя несколько узконаправленных лучей в отличие от традиционных методов с одним глобальным лучом. Однако повышенная гибкость систем спутников с многолучевой антенной (MBS) [63] также вносит сложности в управление динамическим распределением ресурсов (DRA).

В [64] был изучен эффект динамических каналов на нисходящую линию спутниковой связи, управляемую посредством контроля мощности с использованием теории управления очередями. Контроль мощности рассматривался как задача распределения серверных ресурсов. Кроме того, в [65] был предложен метод решения проблем распределения мощности в системах MBS. Однако этот подход упрощает операции, игнорируя межлучевую интерференцию.

Управление расписанием лучей, известное как Beam Hopping (BH), играет важную роль в динамическом распределении ресурсов (DRA) во временной области. Путём выборочной активации лучей в разные моменты времени BH обеспечивает широкое покрытие при использовании меньшего числа каналов

передачи. Метод, предложенный в [66], вводит подход управления ВН на основе $\max\text{-min}$ контроля скорости, адаптированный к требованиям пользовательского трафика данных, что обеспечивает эффективное динамическое управление ВН.

Кроме того, в [67] был использован статистический подход к долгосрочным задержкам для максимизации использования пропускной способности системы. Однако такие подходы сталкиваются с трудностями при одновременном учёте различных требований к качеству обслуживания (QoS), таких как задержка в реальном времени, справедливость распределения лучей и пропускная способность системы. Для решения этой проблемы было предложено несколько методов многокритериальной оптимизации. Например, в [68] был представлен двухступенчатый метод для оптимизации распределения мощности и временных слотов для ВН.

Генетический алгоритм (GA) [69] — это эвристический метод оптимизации, вдохновлённый естественным отбором и генетикой. Он итеративно уточняет популяцию потенциальных решений с использованием таких операций, как отбор, кроссовер и мутация. Были разработаны варианты GA для решения специфических задач в различных областях. Их актуальность для распределения ресурсов в системах MBS обусловлена сложной и динамичной природой среды спутниковой связи.

Основные вклады этого исследования можно резюмировать следующим образом:

- Введение комплексной модели, специально разработанной для распределения ресурсов в многолучевых спутниковых системах
- Проведение сравнительного анализа различных вариантов GA для распределения ресурсов с целью оценки их эффективности в достижении оптимального использования ресурсов и удовлетворения пользователей.

Это исследование развивает методы управления ресурсами спутников, предлагая подход на основе генетического алгоритма (GA) для динамического распределения ресурсов в системах MBS. Подход направлен на оптимизацию

нескольких целей, включая минимизацию задержек связи, потерь пакетов и потребления энергии, при этом учитывая динамические требования трафика в спутниковых коммуникациях [49].

4.2 Математическая формулировка задачи оптимизации

Спутниковая связь является примером систем с ограниченными ресурсами, которые сталкиваются с ограничениями в спектре, мощности спутников и объёме хранилищ. Эта проблема усугубляется ростом инфраструктур спутниковых сетей на низкой околоземной орбите (LEO). Максимизация эффективности использования спектра, объёма хранилищ и распределения мощности спутников при сохранении качества обслуживания (QoS) является критически важной.

В этом разделе рассматриваются тонкости моделирования проблемы распределения ресурсов в системах MBS. Этот подход основан на исследовании Яксинь Хуан и др. [70]. Архитектурное изображение рассматриваемой системы представлено на рисунке 13.

В раздел подробно рассматривается модель канала связи, охватывающая скорость передачи данных и качество сигнала. Далее формулируется задача оптимизации с целью минимизации задержек связи, потерь пакетов и потребления энергии, обеспечивая эффективное использование ресурсов в рамках системных ограничений.

Распределение ресурсов систематически оптимизируется через детальное моделирование компонентов системы MBS и разбиение задачи на части. Исследуются распространение сигнала, уровни мощности и источники шума, а также паттерн освещения лучей (BIP), влияющий на распределение ресурсов. Оптимизация снижает задержки, потери пакетов и потребление энергии, соблюдая ограничения по ресурсам. Модели очередей отслеживают динамику потока данных, а целевая функция оценивает стратегии распределения, интегрируя ключевые метрики.

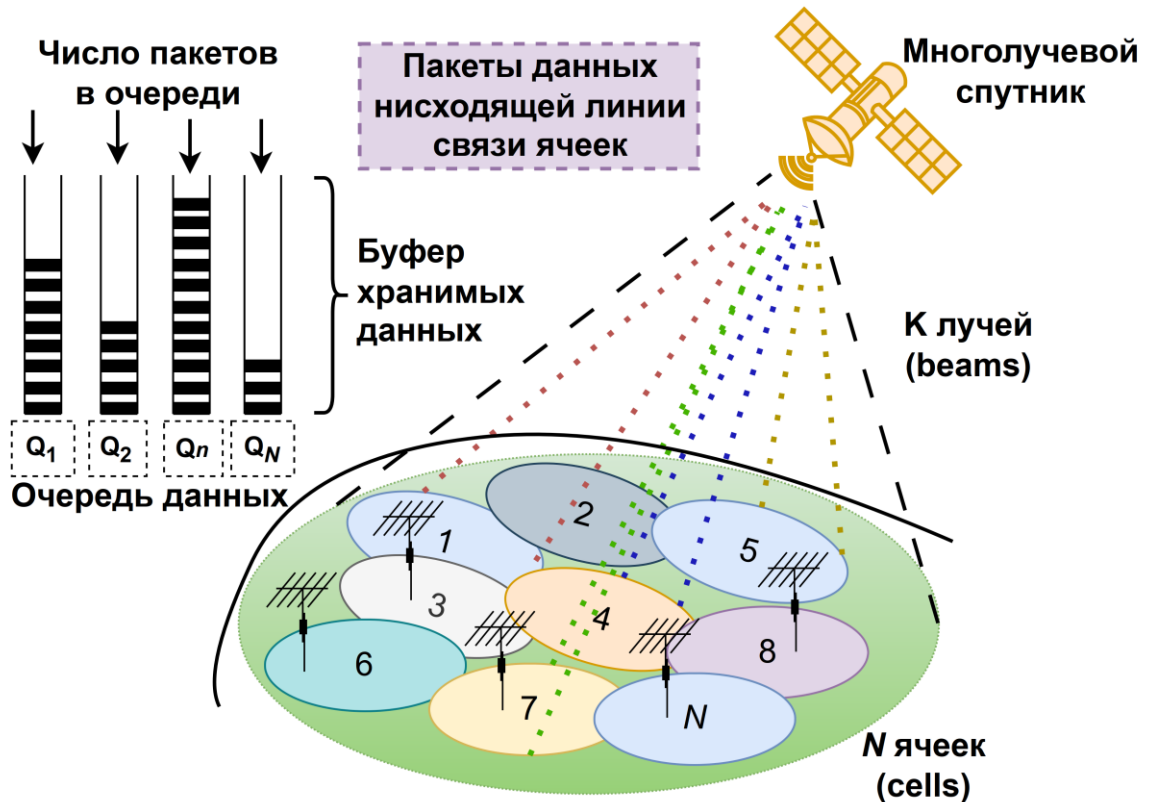


Рисунок 13. Многолучевая система

4.2.1 Математическая модель канала спутниковой связи

а) Увеличение полосы пропускания канала

Этот анализ включает изучение факторов, влияющих на распространение сигнала, уровни мощности и источники шума, которые определяют достигаемые скорости передачи данных и влияние решений по распределению ресурсов [36].

Уравнение бюджета канала [71] является основополагающим для оценки производительности спутниковой связи. Оно вычисляет мощность принятого сигнала на наземном терминале, учитывая усиления и потери вдоль пути передачи.

$$P_r = \frac{P_t G_t G_r}{L_{fs} L_m} \quad (32)$$

где P_r : Приемная мощность, P_t : Мощность передатчика, G_t : Усиление антенны передатчика, G_r : Усиление антенны приемника, L_{fs} : Потери на свободном пространстве, L_m : Различные потери (атмосферное затухание, потери

на оборудовании и др.). Усиление антенн передатчика и приемника рассчитывается по формуле:

$$G = \eta \left(\frac{\pi D}{\lambda} \right)^2 \quad (33)$$

где η : эффективность антенны, D : диаметр антенны, λ : длина волны сигнала. Потери в свободном пространстве, которые представляют собой затухание сигнала из-за расстояния, определяются выражением $L_{fs} = \left(\frac{4\pi d}{\lambda} \right)^2$, где d : расстояние распространения сигнала.

б) Отношение сигнал/шум (SNR) и скорость передачи данных

Достижимая скорость передачи данных в канале связи напрямую связана с SNR. В спутниковой связи мощность шума (N_p) рассчитывается как $N_p = N_0 B^n$, где N_0 : Спектральная плотность мощности шума, B^n : Выделенная полоса частот для ячейки n .

Спектральная плотность мощности шума дополнительно определяется как $N_0 = k_b T_{sys}$, где k_b — постоянная Больцмана [72], T_{sys} — Температура шума системы (включая вклад шума антенны, атмосферы и приемника).

Учитывая помехи от других ячеек (I_{cci}), отношение сигнал/помехи плюс шум (SINR) для ячейки n определяется как:

$$\text{SINR}^n = \frac{P_r^n}{N_p + I_{cci}^n} \quad (34)$$

Наконец, применяя теорему о емкости Шеннона [73], можно определить теоретически максимальную скорость передачи данных (R^n), достижимую в ячейке n :

$$R^n = B^n \log_2 (1 + \text{SINR}^n) \quad (35)$$

Эта модель коммуникационного канала предоставляет основную структуру для понимания того, как решения по распределению ресурсов, такие как распределение мощности и пропускной способности, непосредственно влияют на скорости передачи данных и общую производительность многоручевой спутниковой системы.

4.2.2 Оптимизация освещения зон покрытия спутниковыми лучами (VIP)

VIP определяет конкретный набор ячеек, освещенных лучами спутника в любой данный временной интервал. Эта динамическая схема напрямую влияет на распределение ресурсов, таких как мощность и пропускная способность, и определяет потенциал для одновременного обслуживания пользователей в различных географических зонах.

Математически VIP в временном интервале t в системе Beam Hopping можно представить в виде вектора:

$$X_t = \{x_t^1, x_t^2, \dots, x_t^N \mid x_t^n = 0, 1\} \quad (36)$$

где:

- N : Общее количество ячеек;
- x_t^n : Двоичная переменная, указывающая статус освещения ячейки n в временном интервале t ;
- $x_t^n = 1$: Ячейка n освещена в временном интервале t ;
- $x_t^n = 0$: Ячейка n не освещена в временном интервале t .

VIP определяется графиком перемещения лучей, который задает активацию и деактивацию лучей с течением времени. Стратегическое управление VIP позволяет спутнику адаптироваться к изменяющимся требованиям трафика и приоритизировать обслуживание определенных регионов в зависимости от требований пользователей и целей качества обслуживания (QoS).

4.2.3 Постановка задачи минимизации задержек в спутниковых сетях

Установив модель коммуникационного канала и концепцию VIP, осуществляется переход к формальному определению задачи оптимизации распределения ресурсов в многолучевых спутниковых системах. Основные цели заключаются в минимизации задержки связи, потерь пакетов и потребления энергии при соблюдении присущих системе ограничений.

а) Цели

Наша задача оптимизации сосредоточена на минимизации трех ключевых показателей производительности:

- Задержка передачи в системе (d_{sys}): Этот показатель представляет собой среднее время, которое данные пакеты проводят в системе, включая задержки в очереди и время передачи.
- Коэффициент потерь данных (lr_{sys}): Этот показатель количественно выражает долю данных пакетов, потерянных из-за переполнения буфера или ошибок передачи.
- Нагрузка по потреблению энергии (Pl_{sys}): Этот показатель отражает общее потребление энергии спутниковой системой, учитывая как мощность передачи, так и другие бортовые системы.

б) Проблема оптимизации

Математически, проблему оптимизации можно сформулировать следующим образом:

Минимизировать:

$$G = \beta_1 d_{sys} + \beta_2 lr_{sys} + \beta_3 Pl_{sys} \quad (37)$$

где:

- d_{sys} : задержка передачи системы;
- lr_{sys} : коэффициент потерь пакетов данных;
- Pl_{sys} : нагрузка по потреблению энергии;
- $\beta_1, \beta_2, \beta_3$: коэффициенты весов, представляющие относительную важность каждой цели.

Коэффициенты весов позволяют нам приоритизировать определённые цели в зависимости от желаемой производительности системы и требований к качеству обслуживания (QoS). Например, если минимизация задержки имеет первостепенное значение, то β_1 будет назначен более высокое значение по сравнению с β_2 и β_3 .

в) Переменные решения

В задаче оптимизации используются переменные, такие как x_t^n, P_t^k и B^n , которые представляют статус освещения ячеек, распределение мощности по лучам и распределение полосы пропускания. Цель состоит в минимизации задержки, потерь пакетов и потребления энергии, при этом соблюдая системные ограничения и оптимизируя использование ресурсов.

г) С учётом ограничений

– C1 - Ограничение активации луча: $\sum_{i=1}^N x_t^i = K$ и $x_t^n \in \{0,1\}$. Это ограничение обеспечивает, чтобы количество активированных лучей в любой момент времени t было равно K , где K — максимальное количество одновременных лучей, допустимых системой.

– C2 - Ограничение по мощности: $\sum_{k=1}^K P_t^k \leq P_{tot}$. Это ограничение ограничивает общую мощность, распределяемую по всем лучам в момент времени t , в пределах доступного бюджета мощности спутника P_{tot} .

– C3 - Ограничение мощности индивидуального луча: $P_{min} \leq P_t^k \leq P_{max}$ для всех k . Это ограничение задает минимальные и максимальные уровни мощности для каждого отдельного луча, обеспечивая допустимое и эффективное распределение мощности.

– C4 - Ограничение полосы пропускания: $B^n \leq B_{tot}$ для всех n . Это ограничение ограничивает полосу пропускания, выделяемую для каждой ячейки n , в пределах общей доступной полосы пропускания B_{tot} .

4.2.4 Очередная модель обслуживания трафика в спутниковой сети

Мы рассматриваем модель очередей для каждой ячейки, чтобы точно захватить динамику потока данных и потенциальные задержки в многолучевой спутниковой системе. Эта модель отслеживает количество пакетов данных, ожидающих передачи в каждой ячейке, и их соответствующее время ожидания, предоставляя ценную информацию о производительности системы и влиянии решений по распределению ресурсов.

а) Структура очереди и поток данных

Каждая ячейка оснащена очередью, моделируемой как буфер с принципом «первый пришёл — первый вышел» (FIFO), где входящие пакеты данных хранятся перед передачей. Модель очереди помогает нам анализировать следующие аспекты:

- Длина очереди (Φ_t^n): Это количество пакетов данных, находящихся в очереди в ячейке n в временном интервале t .
- Задержка пакета (L_t): Это время, которое пакет данных проводит в очереди перед передачей.

б) Динамика очереди

На длину очереди в данный временной интервал влияют несколько факторов:

- Предыдущая длина очереди (Φ_{t-1}^n): количество пакетов, уже находящихся в очереди из предыдущего временного интервала;
- Новые поступления (ω_t^n): количество новых пакетов данных, поступающих в ячейку в текущем временном интервале;
- Скорость обслуживания: скорость передачи пакетов из очереди, которая определяется выделенной пропускной способностью и условиями канала.

Длина очереди изменяется согласно следующему уравнению:

$$\Phi_t^n = \Phi_{t-1}^n - \frac{x_{t-1}^n p_{t-1}^k R^n}{B^n} + \omega_t^n \quad (38)$$

где $\frac{x_{t-1}^n p_{t-1}^k R^n}{B^n}$ представляет количество пакетов, обслуженных в предыдущем временном интервале, с учетом состояния освещенности ячейки, выделенной мощности, достижимой скорости передачи данных и пропускной способности.

в) Вычисление задержки

Мы предполагаем дискретную временную систему с равными временными интервалами, что позволяет нам выразить время задержки для каждого пакета следующим образом:

$$\mathbf{L}_t = \begin{bmatrix} l_t^1 \\ l_t^2 \\ \vdots \\ l_t^N \end{bmatrix} = \begin{bmatrix} t_s & 2t_s & \cdots & t_{th} \\ t_s & 2t_s & \cdots & t_{th} \\ \vdots & \vdots & \ddots & \vdots \\ t_s & 2t_s & \cdots & t_{th} \end{bmatrix} \quad (39)$$

где l_t^n : время задержки пакета в ячейке n в временном интервале t относительно текущего временного интервала; t_s : продолжительность одного временного интервала; t_{th} : максимальное время порога очереди, представляющее собой максимально допустимую задержку, после которой пакет считается утерянным. Включив эту модель очереди в анализ, можно эффективно уловить динамическое поведение потока данных в многолучевой спутниковой системе и оценить влияние решений по распределению ресурсов на производительность по задержке и потерям пакетов.

4.2.5 Формулирование целевой функции и ограничений

Для эффективной оценки стратегий распределения ресурсов необходим количественный показатель. Это достигается с помощью целевой функции, которая интегрирует ключевые цели оптимизации: минимизацию задержки, потерь пакетов и потребления энергии. Такая функция позволяет провести математическую оценку компромиссов между этими целями и упрощает сравнение различных подходов к распределению ресурсов.

Целевая функция, обозначенная как G , формулируется как взвешенная сумма трех показателей производительности, как указано в формуле 37.

а) Расчет показателей производительности

Индивидуальные показатели производительности в рамках целевой функции рассчитываются следующим образом:

– Задержка передачи системы (d_{sys}): $d_{sys} = \sum_{i=1}^N \sum_{j=1}^{t_{th}} \phi_{t,j}^i l_j^i$. Эта формула вычисляет среднюю задержку, которую испытывают все пакеты в системе, учитывая задержку каждого пакета в каждой ячейке и количество пакетов, испытывающих эту задержку.

- Коэффициент потерь пакетов данных $(lr_{sys}):lr_{sys} = \frac{\sum_{i=1}^N \phi_{t,t_{th}}^i}{\sum_{i=1}^N \sum_{j=1}^{t_{th}} \phi_{t,j}^i}$. Эта

формула представляет собой отношение количества потерянных пакетов из-за превышения максимального порогового времени к общему количеству пакетов, поступивших в систему.

- Нагрузка по потреблению энергии $(Pl_{sys}):Pl_{sys} = \frac{\sum_{i=1}^N x_t^i P_t^i}{P_{tot}}$. Эта

формула рассчитывает долю от общего доступного запаса энергии, потребляемую освещёнными ячейками в данный временной интервал.

Формула 37 представляет собой сложную функцию оптимизации, подчеркивающую задачу учета множества конфликтующих целей. Эти цели, стремящиеся к экономии энергии при улучшении задержек и минимизации потерь пакетов в передаче данных спутниковых сетей, добавляют дополнительный уровень сложности к задаче оптимизации.

В многокритериальной оптимизации не доминирующие генетические алгоритмы, такие как NSGA-II (Non-dominated Sorting Genetic Algorithm II) или SPEA2 (Strength Pareto Evolutionary Algorithm 2), являются ценными для обработки сложных и конфликтующих целей. Они позволяют принимающим решения эффективно исследовать компромиссы и находить оптимальные решения по Парето, предоставляя диапазон сбалансированных вариантов для принятия решений.

4.3 Разработка алгоритмических решений на основе методов ИИ

Для решения поставленной задачи многокритериальной оптимизации были выбраны эволюционные алгоритмы. Эволюционные вычисления, к которым относятся генетические алгоритмы (такие как NSGA-II и SPEA2), являются одним из направлений искусственного интеллекта. Их основной принцип заключается в имитации процессов естественного отбора для поиска оптимальных решений в сложных задачах.

Наша работа использует генетический алгоритм для решения задачи, который включает три основных операции: отбор, кроссовер и мутацию. Мы рассматриваем NSGA-II [74] и SPEA2 [75]. Рисунок 14 иллюстрирует процесс реализации многокритериальной оптимизации с использованием алгоритмов генетического программирования с неконфликтным сортированием.

4.3.1 Оптимизация распределения ресурсов с помощью NSGA-II

Эти шаги описывают процесс: генерация начальной популяции, соблюдение ограничений, вычисление целевых функций, сортировка решений по неконфликтным фронтам, оценка уровней перегрузки, применение генетических операторов, смешивание популяций и проверка критериев остановки до достижения набора решений Плато.

Предоставленный псевдокод в таблице 5 описывает алгоритм NSGA-II, широко используемый эволюционный алгоритм для многокритериальных задач оптимизации.

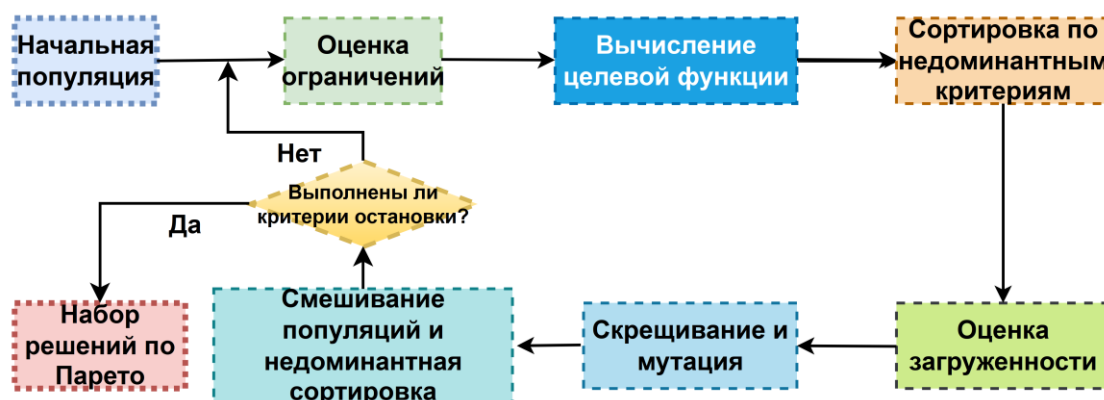


Рисунок 14. Многокритериальная генетическая оптимизация на основе алгоритма неконфликтного сортировки

В этом исследовании целевая функция вычисляется через последовательный процесс: сначала, на каждом временном интервале t_s определяется поступление пакетов (ϕ_t^n) на основе данных о скорости передачи в ячейках. Затем остаток пакетов в каждой очереди количественно оценивается с использованием Формулы (38). Далее оценивается целевая функция, включая

формулировки для d_{sys} , lr_{sys} и Pl_{sys} . Соблюдение требований к ограничениям (C1 – C4) имеет ключевое значение на протяжении всего этого процесса.

Таблица 5. Генетический алгоритм с недоминантной сортировкой II

Алгоритм 1: NSGA-II (Генетический алгоритм с недоминантной сортировкой II)

- 1: Инициализировать популяцию P случайными решениями
 - 2: Оценить целевые функции для каждого решения в P
 - 3: Инициализировать пустые множества F_1, F_2, F_k для хранения фронтов
 - 4: Инициализировать пустое множество F
 - 5: Установить счётчик поколений $t = 0$
 - 6: **while** условие завершения не выполнено **do**
 - 7: Создать потомков Q , выполняя генетические операции (скрещивание и мутация) на P
 - 8: Объединить популяции: $R = P \cup Q$
 - 9: Выполнить недоминантную сортировку на R , чтобы создать фронты F_1, F_2, F_k
 - 10: Инициализировать P' как пустое множество
 - 11: Установить $i = 1$
 - 12: **while** $|P'| + |F_i| \leq N$ **do**
 - 13: Присвоить ранг i индивидам из F_i
 - 14: Добавить F_i в P'
 - 15: Увеличить i
 - 16: **end while**
 - 17: Отсортировать оставшихся индивидов в F_i на основе дистанции загруженности
 - 18: Добавить индивидов из F_i с наибольшей дистанцией загруженности в P' , пока $|P'| = N$
 - 19: Обновить P , выбрав первых N индивидов из P'
 - 20: Увеличить счётчик поколений: $t = t + 1$
 - 21: **end while**
 - 22: **return** Аппроксимацию фронта Парето P
-

NSGA-II инициализирует популяцию решений случайным образом и оценивает их целевые функции. Через многоступенчатый процесс в каждом поколении он эволюционирует к оптимальному набору решений, известному как

фронт Парето [76]. Центральным элементом NSGA-II является механизм недоминированной сортировки, который организует решения в различные фронты на основе отношений доминирования. Каждый фронт представляет собой доминирование, при этом начальные фронты содержат недоминированные решения, а последующие фронты доминируются исключительно решениями из предыдущих фронтов.

4.3.2 Применение эволюционного алгоритма SPEA2 для минимизации задержек

Эффективное распределение ресурсов для многолучевых спутников в спутниковой связи является сложной задачей. Поскольку несколько лучей обслуживают определённые зоны, задача включает в себя оптимизацию распределения для максимизации результатов, возможно, учитывая противоречивые цели. Основные шаги алгоритма SPEA2 показаны в таблице 6.

Алгоритм SPEA2 является мощным инструментом для решения задач многокритериальной оптимизации. Он работает, имитируя естественный отбор для эволюции популяции потенциальных решений.

- Шаг (1) Инициализация: Инициализируйте популяцию P и архив A .
- Шаг (2) Основной цикл:
 - Оценка приспособленности: Рассчитайте исходную приспособленность и силу для каждого решения в P .
 - Оценка плотности: Рассчитайте оценку плотности для каждого решения в P на основе его расстояния до других решений
 - Архивирование: Обновите архив A ненадомозначенными решениями из P .
 - Отбор: Выберите индивидов из объединённой популяции $P \cup A$ для следующего поколения с использованием механизма отбора
 - Замена: Замените текущее поколение P выбранными индивидами

– Шаг (3) Критерии завершения: Повторяйте основной цикл до тех пор, пока не будут выполнены критерии завершения, такие как достижение максимального числа поколений или сходимость.

В отличие от традиционных алгоритмов, которые стремятся найти одно лучшее решение, SPEA2 сосредоточен на выявлении набора Парето-оптимальных решений. Эти решения представляют собой наилучшие возможные компромиссы между противоречащими целями.

Таблица 6. Алгоритм эволюции Strength Pareto 2 (SPEA-2)

Алгоритм 2: Алгоритм эволюции Strength Pareto 2 (SPEA-2)	
1:	Инициализировать популяцию P
2:	Инициализировать архив A
3:	repeat
4:	Рассчитать сырые значения приспособленности и силу для каждого решения в P
5:	Рассчитать оценку плотности для каждого решения в P
6:	Обновить архив A не доминирующими решениями из P
7:	Выбрать индивидов из $P \cup A$ для следующего поколения
8:	Заменить P выбранными индивидами
9:	until условие завершения не выполнено

4.4 Экспериментальные исследования и анализ результатов

4.4.1 Описание сценариев симуляции и параметров моделирования

Таблица 7 представляет параметры сценария симуляции, включая 21 обслуживающую ячейку с пятью передатчиками лучей и мощностью спутника 34 дБВт.

Частота нисходящего канала составляет 21,5 ГГц, а общая пропускная способность — 500 МГц. В наших экспериментах, проведённых с тщательным вниманием к деталям, максимальная общая скорость поступления данных k варьируется от 1000 до 2000 Мбит/с. Нормализованный трафик данных,

распределённый согласно закону Дирихле, обеспечивает суммирование трафика каждой ячейки до k . Размер каждого пакета данных, соответствующий стандартам протокола TCP/IP, составляет 1522 байта.

Таблица 7. Параметры сценария симуляции

Параметр	Значение
Высота орбиты h (км)	1560
Частота нисходящей линии связи (Ka) f_c (ГГц)	21.5
Количество ячеек N	21
Количество лучей K	5
Общая полоса частот B_{tot} (МГц)	500
Общая мощность спутника P_{tot} (дБВт)	34
Максимальная мощность луча P_{max} (дБВт)	32
Минимальная мощность луча P_{min} (дБВт)	21
Спектральная плотность мощности шума N_0 (дБм/Гц)	-187.6
Усиление антенны передатчика G_t (дБи)	33.3
Усиление антенны приемника G_r (дБи)	41.3
Путь потерь L_p (дБ)	202.1
Интервал времени прибытия пакетов данных t_s (мс)	10
Пороговое значение очереди t_{th} (мс)	400
Скорость поступления трафика данных λ (Мбит/с)	1000~2000
Размер пакетов данных (байты)	1522

Мы исследуем две стратегии: генетические алгоритмы NSGA-II и SPEA2. Исследование с комплексным подходом направлено на сравнение их эффективности в динамическом распределении ресурсов между несколькими лучами, оптимизации пропускной способности системы и сокращении потерь ресурсов.

4.4.2 Оценка производительности предложенных методов и анализ эффективности

В данном разделе проводится анализ взаимосвязи между стратегиями распределения ресурсов и основными показателями производительности в гибридных спутниковых сетях. На основе проведённых экспериментов были рассмотрены такие ключевые метрики, как потребление энергии, время задержки и потери пакетов. Результаты экспериментов позволили выявить, что оптимизация стратегий распределения ресурсов может значительно снизить сетевые задержки, а также минимизировать энергопотребление и потери пакетов. Данный анализ закладывает основу для дальнейшего совершенствования методов минимизации задержки в спутниковых каналах связи с использованием технологий искусственного интеллекта.

Рисунки (15a; 15b) показывают взаимосвязь между потреблением энергии и двумя ключевыми показателями производительности: средней задержкой передачи и коэффициентом потерь пакетов. Данные получены из симуляций, в которых использовались две вариации генетического алгоритма — NSGA-II и SPEA2 — для оптимизации распределения ресурсов в многолучевой спутниковой системе.

Рисунок (15a) показывает явную отрицательную корреляцию. При снижении потребления энергии средняя задержка передачи увеличивается как для NSGA-II, так и для SPEA2. Это указывает на компромисс: для достижения более низкой задержки требуется большее потребление энергии. NSGA-II стабильно поддерживает несколько более низкие задержки по сравнению с SPEA2, варьирующиеся от примерно 51,6 мс до 53,3 мс и от 54,3 мс до 65,8 мс соответственно в разных итерациях. Это свидетельствует о том, что NSGA-II более эффективно распределяет ресурсы для минимизации задержек связи, в то время как SPEA2, как правило, демонстрирует более высокие задержки, что может указывать на использование менее эффективных методов. Что касается

потребления энергии, то SPEA2 постоянно показывает более низкие значения (от 39,7% до 61,9%) по сравнению с NSGA-II (от 67,3% до 67,8%).

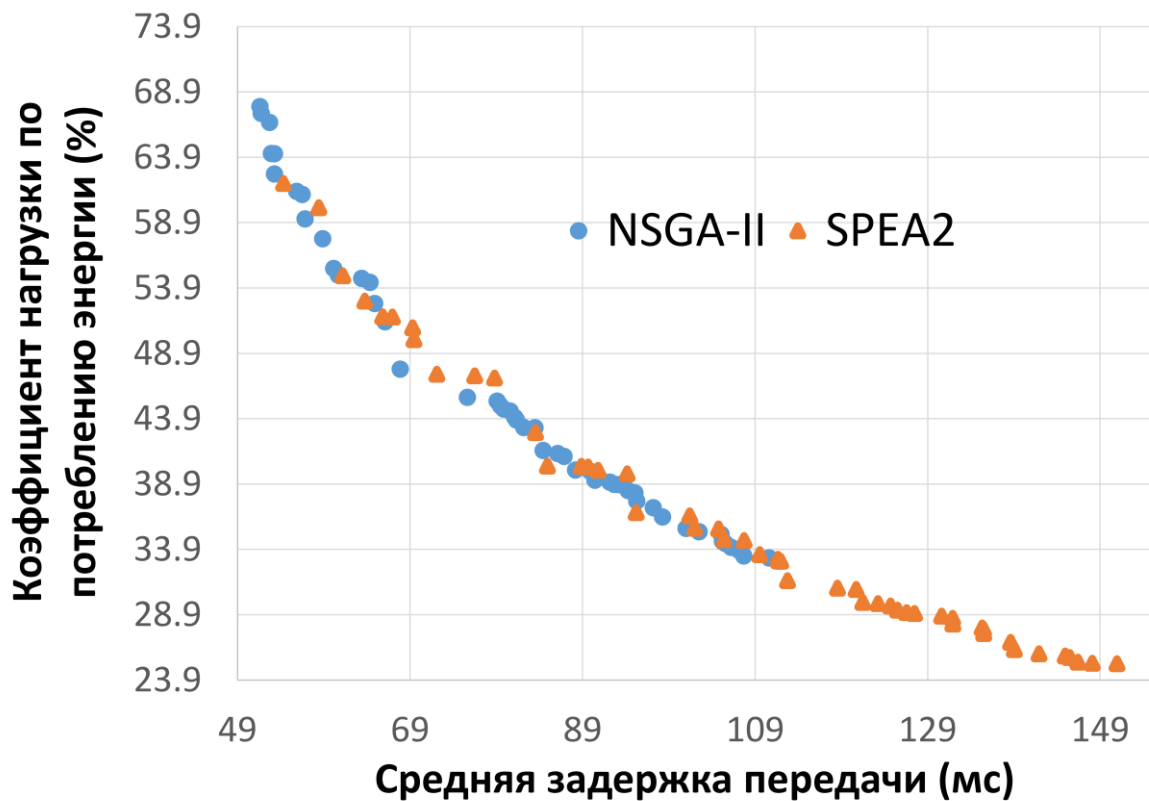


Рисунок 15а. Взаимосвязь между потреблением энергии спутниковой системы и средней задержкой передачи.

Рисунок (15b) также демонстрирует отрицательную корреляцию. Более низкое потребление энергии, как правило, приводит к более высоким коэффициентам потерь пакетов для обоих алгоритмов. Это предполагает, что для поддержания низкого уровня потерь пакетов требуется большее потребление энергии.

Однако NSGA-II стабильно демонстрирует более низкие значения потерь пакетов по сравнению с SPEA2 на протяжении всех итераций. Например, потери пакетов для NSGA-II составляют от примерно 0,31% до 0,36% в начале, тогда как для SPEA2 — от 0,50% до 5,66%. SPEA2 имеет значительное преимущество в потреблении энергии, варьирующемся от примерно 23,69% до 44,43% на начальных этапах, тогда как у NSGA-II этот показатель находится в пределах от 34,88% до 53,00%. Высокая энергоэффективность SPEA2 критична для

оптимизации распределения ресурсов в многолучевых спутниковых системах, снижения эксплуатационных расходов и увеличения срока службы оборудования.

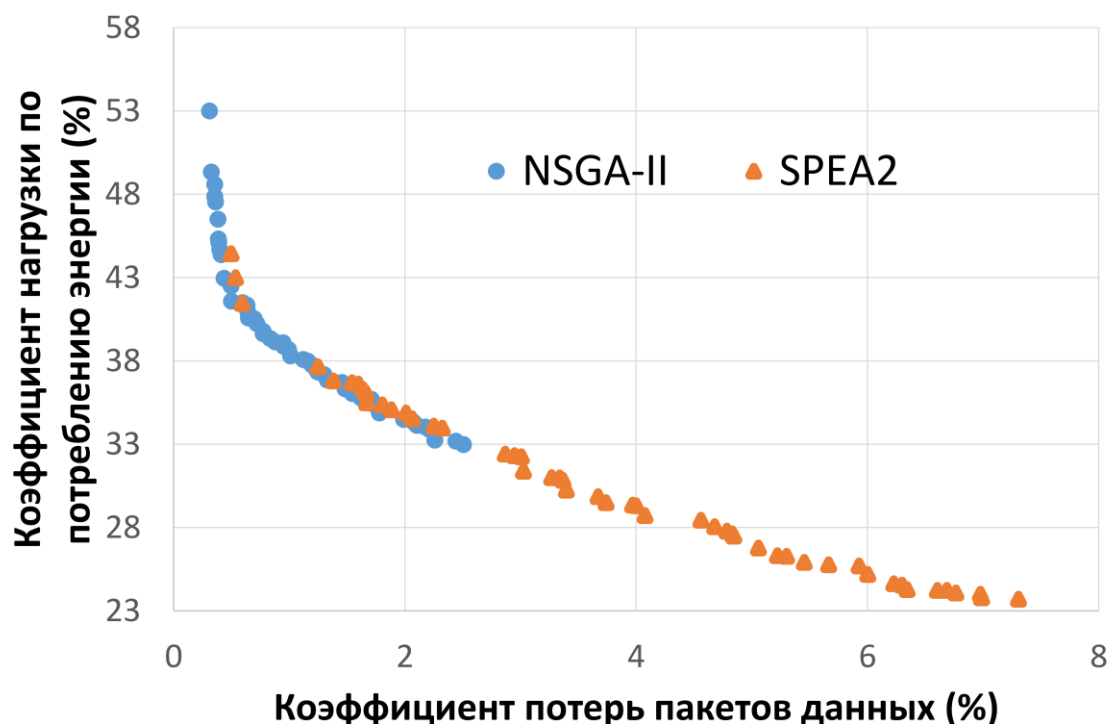


Рисунок 15b. Взаимосвязь между потреблением энергии спутниковой системы и коэффициентом потерь данных

Рисунки (16a; 16b) иллюстрируют среднюю задержку передачи (в миллисекундах) и коэффициент потерь данных (в процентах) при различных уровнях общего пользовательского трафика (измеряемых в Мбит/с).

Рисунок (16a) показывает положительную корреляцию между общим спросом на пользовательский трафик и средней задержкой передачи как для алгоритмов NSGA-II, так и для SPEA2. Обе линии растут по мере увеличения спроса на трафик. По мере роста спроса оба алгоритма испытывают более высокие задержки, что ожидаемо из-за увеличения перегруженности сети и конкуренции за ресурсы. Однако NSGA-II стабильно превосходит SPEA2, поддерживая более низкие средние задержки при всех уровнях спроса на трафик. Это свидетельствует о том, что NSGA-II более эффективно управляет распределением ресурсов и минимизирует задержки даже при высоких нагрузках на трафик.

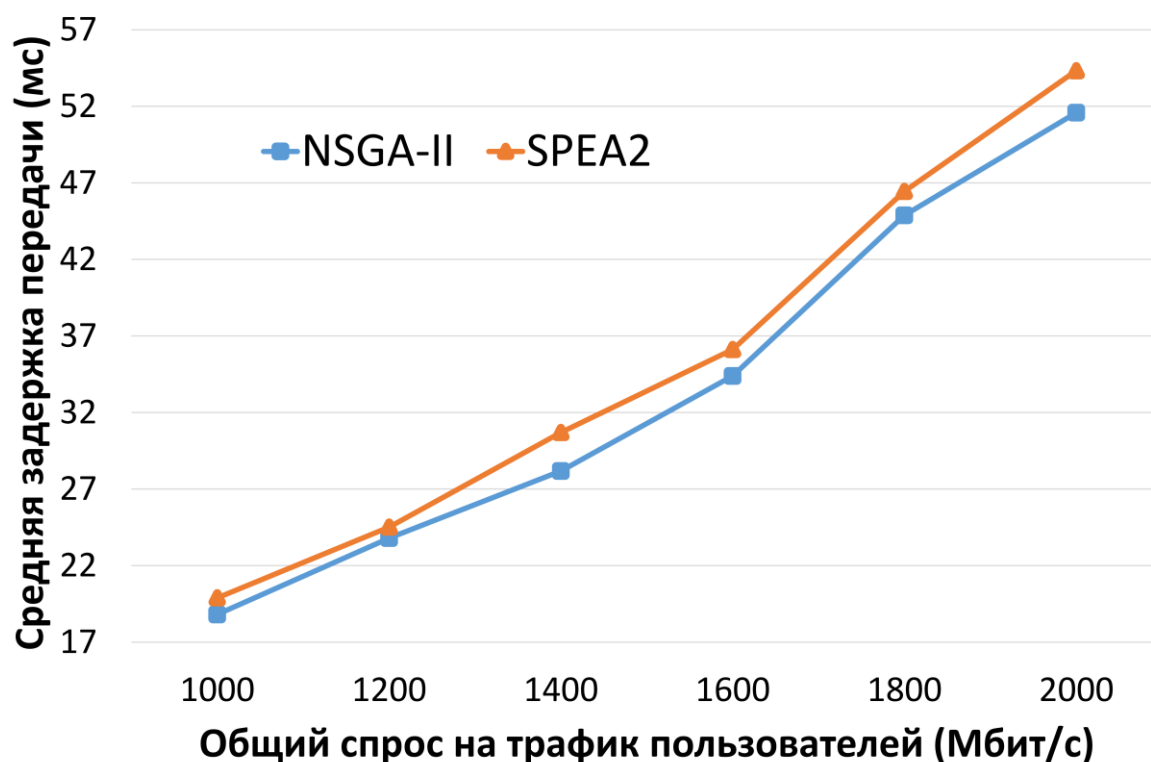


Рисунок 16а. Средняя задержка передачи при различных общих требованиях к пользовательскому трафику

Рисунок (16b) показывает положительную корреляцию между общим спросом на пользовательский трафик и коэффициентом потерь данных для алгоритмов NSGA-II и SPEA2. Подобно задержке, оба алгоритма демонстрируют положительную корреляцию между спросом на трафик и коэффициентом потерь пакетов. По мере увеличения спроса коэффициент потерь пакетов также возрастает, что указывает на более высокую степень потерь данных из-за перегруженности сети. Оба алгоритма показывают увеличение задержки с ростом общего спроса на трафик, что ожидаемо из-за нагрузки на ресурсы. NSGA-II стабильно поддерживает более низкие средние значения задержки по сравнению с SPEA2 при всех уровнях спроса. При 1000 Мбит/с задержка для NSGA-II составляет 18,81 мс, тогда как для SPEA2 — 19,89 мс. Эта тенденция сохраняется с увеличением спроса, что свидетельствует о превосходной способности NSGA-II минимизировать задержку. При 2000 Мбит/с задержка для NSGA-II составляет 51,58 мс по сравнению с 54,33 мс для SPEA2, что подчеркивает эффективность NSGA-II даже при высоком спросе. Это стабильное превосходство подчеркивает

эффективность NSGA-II в оптимизации распределения ресурсов для многолучевых спутников и улучшении качества обслуживания за счет минимизации задержек.

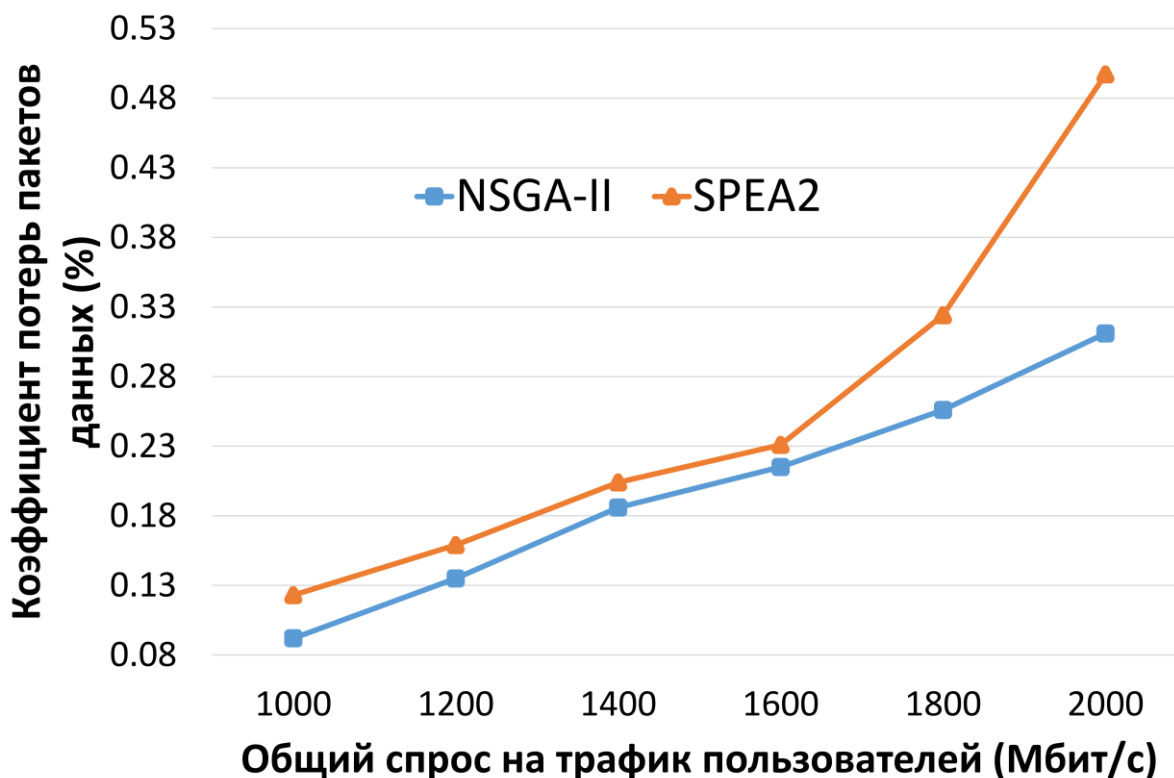


Рисунок 16б. Коэффициент потерь данных при различных общих требованиях к пользовательскому трафику.

В спутниковой связи потери пакетов отражают потерю данных из-за помех и перегрузки. Как NSGA-II, так и SPEA2 показывают увеличение потерь по мере роста трафика, что создаёт нагрузку на каналы связи.

Однако NSGA-II стабильно демонстрирует более низкий уровень потерь по сравнению с SPEA2 при всех уровнях трафика. При скорости передачи 1000 Мбит/с потери для NSGA-II составляют 0,092%, а для SPEA2 — 0,123%. Эта тенденция сохраняется при увеличении требований, что подчеркивает превосходную способность NSGA-II минимизировать потери пакетов и улучшать надежность канала связи в многолучевых спутниковых системах.

При сравнении показателей производительности SPEA2 и NSGA-II было отмечено, что время выполнения SPEA2 превышает время выполнения NSGA-II в 1,26 раза. Это различие подчеркивает врожденные вычислительные отличия в их

подходах к оптимизации распределения ресурсов в многолучевых спутниковых системах. Выявленное различие имеет значение для практического применения, подчеркивая необходимость глубокого изучения факторов, способствующих этим различиям в времени выполнения.

4.5 Практическая реализация методов минимизации задержки: тестирование на симуляторе

В данном разделе представлена методика тестирования разработанных методов минимизации сетевой задержки в фрагменте гибридной сети связи. Основное внимание уделяется оценке эффективности интеллектуальной приоритизации трафика и динамического управления ресурсами сети на основе технологий искусственного интеллекта.

Для проведения тестирования используется моделирование на основе реальных данных трафика и симуляции, реализованной в SimPy. Это позволяет проанализировать влияние предложенных методов на ключевые метрики сетевой задержки, включая задержки в очереди, сквозную задержку и пропускную способность сети в различных сценариях нагрузки.

4.5.1 Методологии приоритизации трафика с использованием ИИ

На рисунке 17 представлена схема выполнения методологии приоритизации трафика на основе ИИ для планирования пакетов в спутниковой системе связи GEO. Поток выполнения методологии приоритизации трафика на основе ИИ: процесс начинается с захвата сетевого трафика, затем выполняется классификация пакетов на основе ИИ, моделирование и управление очередью, а завершается оценкой производительности. Каждый этап в этом процессе гарантирует оптимальную приоритизацию критически важных данных для эффективной передачи.

а) Шаг 1 - Захват трафика

Первым этапом является захват реальных сетевых пакетов с использованием инструментов анализа трафика, таких как Wireshark или tcpdump. Эти пакеты содержат важные атрибуты, включая тип протокола, порты источника и назначения, размер пакета и время прибытия. Этот шаг обеспечивает точное представление характеристик реального спутникового сетевого трафика, предоставляя необходимые данные для дальнейшей обработки и классификации.

б) Шаг 2 - Классификация пакетов с использованием ИИ:

После захвата сетевого трафика пакеты классифицируются с помощью метода на основе ИИ. Глубокая нейронная сеть (DNN) используется для разделения пакетов на две группы: с **ВЫСОКИМ** приоритетом и с **НИЗКИМ** приоритетом. Классификация основывается на выбранных характеристиках захваченных пакетов, при этом пакеты с **ВЫСОКИМ** приоритетом требуют более быстрой передачи из-за их критической важности, в то время как пакеты с **НИЗКИМ** приоритетом менее чувствительны ко времени и могут обрабатываться с меньшей срочностью.

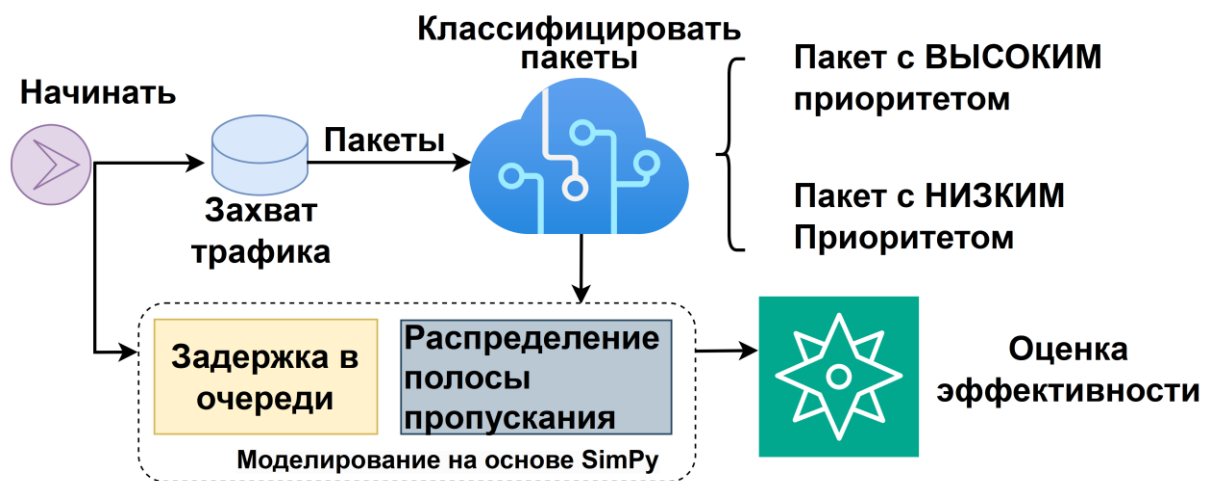


Рисунок. 17. Схема выполнения методологии приоритизации трафика на основе ИИ для планирования пакетов в спутниковой системе связи GEO.

в) Шаг 3 - Моделирование и управление очередью:

После классификации пакеты передаются в моделирование на основе SimPy, где используется приоритетная очередь M/M/1. В рамках моделирования выполняются два ключевых процесса:

Расчет задержки в очереди: система оценивает время ожидания каждого пакета перед передачей. Пакеты с ВЫСОКИМ приоритетом испытывают меньшую задержку, тогда как пакеты с НИЗКИМ приоритетом могут ожидать дольше из-за их более низкого приоритета.

Распределение полосы пропускания: динамическое выделение полосы пропускания выполняется на основе приоритета пакетов. Процесс распределения учитывает отношение сигнал-шум (SNR) и уровень шума в канале, применяя формулу пропускной способности Шеннона для оптимизации эффективного распределения полосы пропускания, обеспечивая соответствие системы требованиям производительности сети.

г) Шаг 4 - Оценка производительности:

Завершающий этап включает анализ эффективности механизма приоритизации трафика. Оцениваются ключевые метрики, такие как задержка в очереди, пропускная способность системы и общая эффективность планирования. Этот анализ позволяет определить, насколько успешно предложенная стратегия приоритизации управляет передачей пакетов и оптимизирует производительность сети спутниковой связи GEO.

4.5.2 Классификация сетевого трафика с применением глубокой нейронной сети (DNN)

Алгоритм классификации трафика, представленный в Алгоритме 3, начинается с извлечения данных сетевого трафика из файла PCAP. Это достигается с помощью инструментов анализа пакетов, таких как Wireshark или Scapy, которые позволяют выделять критические характеристики пакетов, включая тип протокола, размер пакета, время межпакетного прибытия, а также IP-адреса источника и назначения. Эти характеристики необходимы для последующей задачи классификации и предварительно обрабатываются, чтобы обеспечить их соответствие формату входных данных для модели глубокой нейронной сети (DNN). После предобработки характеристики каждого пакета

формируются в вектор признаков, который затем подается на вход предварительно обученной модели DNN для классификации.

В фазе классификации векторные признаки каждого пакета передаются через предварительно обученную модель DNN, которая вычисляет оценку классификации. На основе заранее заданного порога пакеты классифицируются как HIGH-priority или LOW-priority. Пакеты HIGH-priority, представляющие критические данные, получают более быстрый транзит, в то время как пакеты LOW-priority обрабатываются с меньшей срочностью.

Алгоритм 3: Классификация трафика с использованием DNN

- 1: **Вход:** Файл PCAP, предобученная модель DNN, признаки пакетов (тип протокола, размер пакета, межпакетное время и т. д.)
- 2: **Выход:** Классифицированный трафик (высокий или низкий приоритет)
- 3: **Шаг 1: Загрузка и извлечение данных пакетов**
- 4: Считать пакеты из файла PCAP с помощью анализатора трафика (например, Wireshark, Scapy)
- 5: Извлечь ключевые признаки: тип протокола, размер пакета, межпакетное время, IP-адреса источника и назначения
- 6: **Шаг 2: Предобработка признаков**
- 7: Нормализовать числовые признаки (например, размер пакета, межпакетное время)
- 8: Закодировать категориальные признаки (например, тип протокола с использованием one-hot encoding)
- 9: Сформировать векторы признаков f_i для каждого пакета
- 10: **Шаг 3: Загрузка и применение модели DNN**
- 11: Загрузить предобученную модель DNN (например, TensorFlow/PyTorch)
- 12: **for** каждый пакет в PCAP **do**
- 13: Вычислить оценку классификации: $\hat{y}_i = \text{DNN}(\mathbf{f}_i)$
- 14: **if** $\hat{y}_i > \text{порог}$ **then**

```

15:    Присвоить высокий приоритет
16:  else
17:    Присвоить низкий приоритет
18:  end if
19: end for
20: Шаг 4: Сохранение и возврат результатов
21: Сохранить классифицированные пакеты с метаданными (временная
метка, IP, классификация)
22: Вернуть результаты классификации для дальнейшей обработки
23: Шаг 5: Оценка качества модели
24: if доступны истинные метки then
25:   Вычислить точность, precision, recall и F1-score:

$$\text{Precision} = \frac{TP}{TP+FP}, \text{ Recall} = \frac{TP}{TP+FN}, F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

26: end if

```

Результаты классификации, включая метаданные, такие как временная метка, исходный/целевой IP-адрес и присвоенный приоритет, сохраняются для дальнейшей обработки. Наконец, если доступны метки истинных значений, эффективность модели оценивается с помощью расчета ключевых метрик, таких как точность, полнота, отзывчивость и F1-оценка. Эти метрики помогают оценить способность модели правильно классифицировать пакеты в соответствующие категории приоритетов и предоставляют информацию о ее общей эффективности.

4.5.3 Оптимальное выделение полосы пропускания в сети GEO-спутников

Процесс выделения полосы пропускания в сети GEO-спутников включает в себя распределение ресурсов (как полосу пропускания, так и мощность передачи) пакетам в зависимости от их уровня приоритета. Алгоритм в

Алгоритме 4 описывает ключевые этапы для достижения этой цели, обеспечивая соблюдение как приоритетов пакетов, так и ограничений ресурсов спутника.

Предобработка данных о трафике: Процесс начинается с извлечения данных о трафике, включая уровни приоритета каждого пакета. Эти уровни приоритета затем нормализуются для обеспечения согласованного и справедливого распределения ресурсов. Этот процесс нормализации помогает алгоритму выделить больше ресурсов для пакетов с высоким приоритетом и меньше для пакетов с низким приоритетом. После нормализации данные о трафике подготавливаются для последующих этапов выделения ресурсов.

Определение ограничений ресурсов: Следующий шаг - определение ограничений ресурсов системы. Общая доступная полоса пропускания должна быть распределена между всеми пакетами таким образом, чтобы сумма выделенной полосы пропускания не превышала общую доступную полосу пропускания сети. Точно так же общая мощность передачи, используемая спутниковой системой, также должна быть ограничена, обеспечивая, чтобы сумма всех распределений мощности пакетов не превышала мощности передачи спутника.

Алгоритм 4: Распределение ресурсов в сети GEO-спутника с учетом приоритета пакетов

1: **Вход:** Трафик пакетов с уровнями приоритета $\{P_1, P_2, \dots, P_N\}$, доступная пропускная способность B , параметры спутника (мощность передачи P_{sat} , энергетический баланс L , емкость канала C , высота спутника h), временное окно Δt

2: **Выход:** Выделенные ресурсы для каждого пакета: пропускная способность B_i , мощность передачи P_i

3: **Шаг 1: Предобработка данных**

4: Извлечение характеристик трафика и нормализация уровней приоритета $\{P_1, P_2, \dots, P_N\}$

5: **Шаг 2: Определение ограничений**

6: Соблюдение общего ограничения по пропускной способности: $\sum_{i=1}^N B_i = B$

7: Ограничение по общей мощности: $\sum_{i=1}^N P_i \leq P_{\text{sat}}$

8: Шаг 3: Приоритизация трафика

9: Выделение больших ресурсов для пакетов с высоким приоритетом.
Нормализация приоритета:

$$P'_i = \frac{P_i}{\sum_{i=1}^N P_i}$$

10: Вычисление выделенных ресурсов на основе приоритета:

$$B_i = P'_i \cdot B, P_i = P'_i \cdot P_{\text{sat}}$$

11: Шаг 4: Учет ограничений спутника

12: Вычисление мощности для каждого пакета

$$P_i = \frac{L \cdot C}{\text{gain}_i \cdot h^2}$$

13: Обеспечение соблюдения ограничений по емкости канала

14: Шаг 5: Корректировка на основе качества канала

15: Корректировка мощности и пропускной способности с учетом SNR:

$$P_i = P_i \cdot \left(\frac{\text{SNR}_i}{\text{SNR}_{\min}} \right), B_i = B_i \cdot \left(\frac{\text{SNR}_i}{\text{SNR}_{\min}} \right)$$

16: Шаг 6: Выходные данные

17: Вывести окончательные выделенные ресурсы: B_i и P_i

Приоритизация трафика: Для обработки разных уровней приоритета трафика алгоритм выделяет больше полосы пропускания и мощности передачи для пакетов с более высоким приоритетом. Процесс нормализации приоритета обеспечивает пропорциональное распределение общего объема выделенных ресурсов на основе относительной важности каждого пакета. Это позволяет системе отдавать приоритет более критическому трафику, например, для реального времени или высокоприоритетных коммуникаций, одновременно обеспечивая справедливую долю доступных ресурсов для трафика с низким приоритетом.

Учет ограничений спутника: С учетом физических характеристик спутников GEO необходимо учитывать дополнительные ограничения, такие как бюджет канала, пропускная способность канала и высота спутника. Алгоритм вычисляет требуемую мощность передачи для каждого пакета, исходя из этих параметров спутника. Эти вычисления обеспечивают правильное распределение мощности для каждого пакета, соответствуя операционным ограничениям спутниковой системы связи.

Корректировка с учетом качества канала: Качество канала может изменяться из-за факторов, таких как отношение сигнал/шум (SNR), что влияет как на мощность, так и на требования к полосе пропускания для каждого пакета. Алгоритм корректирует выделенную полосу пропускания и мощность в зависимости от SNR, обеспечивая, чтобы пакеты с лучшим качеством канала получали больше ресурсов. Эта динамическая корректировка помогает системе оптимизировать распределение ресурсов и поддерживать высокий уровень качества связи.

Наконец, после учета всех ограничений и корректировок, алгоритм выводит распределенные ресурсы для каждого пакета. Это включает как пропускную способность, так и мощность передачи, назначенные каждому пакету в зависимости от его приоритета и ограничений сети. Эти результаты могут быть использованы для дальнейшего анализа или оценки производительности, помогая оценить эффективность стратегии распределения ресурсов для обеспечения оптимальной производительности сети.

4.5.4 Результаты симуляции задержек в очередях на базе SimPy

Симуляция на базе SimPy используется для моделирования и оценки задержки в очереди в системе, использующей модель очереди M/M/1, которая представляет собой одно-серверную очередь с экспоненциальными скоростями поступления и обслуживания. В этой симуляции пакеты сети, классифицированные как пакеты высокого или низкого приоритета,

обрабатываются в очереди в зависимости от их уровня приоритета. Система моделирует поступление пакетов, соответствующие времена обслуживания и процесс очереди с использованием дискретной событийной симуляции. Пакеты высокого приоритета имеют более короткое время ожидания из-за более высоких скоростей обслуживания, в то время как пакеты низкого приоритета сталкиваются с большими задержками, что отражает дифференцированное обращение в системе. SimPy, библиотека для дискретной событийной симуляции на языке Python, позволяет моделировать эти процессы через простое планирование и обработку событий, что позволяет динамически оценивать поведение очереди при различных условиях трафика.

Помимо моделирования задержек в очереди, симуляция также включает распределение пропускной способности, что является критическим аспектом производительности сети в системе спутниковой связи GEO. Процесс распределения пропускной способности динамически корректируется в зависимости от приоритета пакетов и текущих условий сети, включая отношение сигнал/шум (SNR) и шум канала. С использованием формулы пропускной способности Шеннона вычисляется эффективная пропускная способность для каждого класса приоритетов, оптимизируя пропускную способность системы. Интегрируя как задержки в очереди, так и распределение пропускной способности в симуляцию на базе SimPy, модель может предоставить ценные данные о общей производительности системы спутниковой связи, особенно с точки зрения эффективности механизма управления трафиком на основе приоритета.

В данной работе мы проводим моделирование с использованием NVIDIA NX215 Carrier Board – AVerMedia, AI-решения, предназначенного для высокопроизводительных периферийных вычислений. Эта аппаратная платформа обеспечивает эффективную обработку предложенных нами моделей, используя передовые возможности GPU для ускорения задач глубокого обучения и оптимизации. На рисунке 18 представлена плата NVIDIA NX215 Carrier Board, используемая в наших экспериментах.

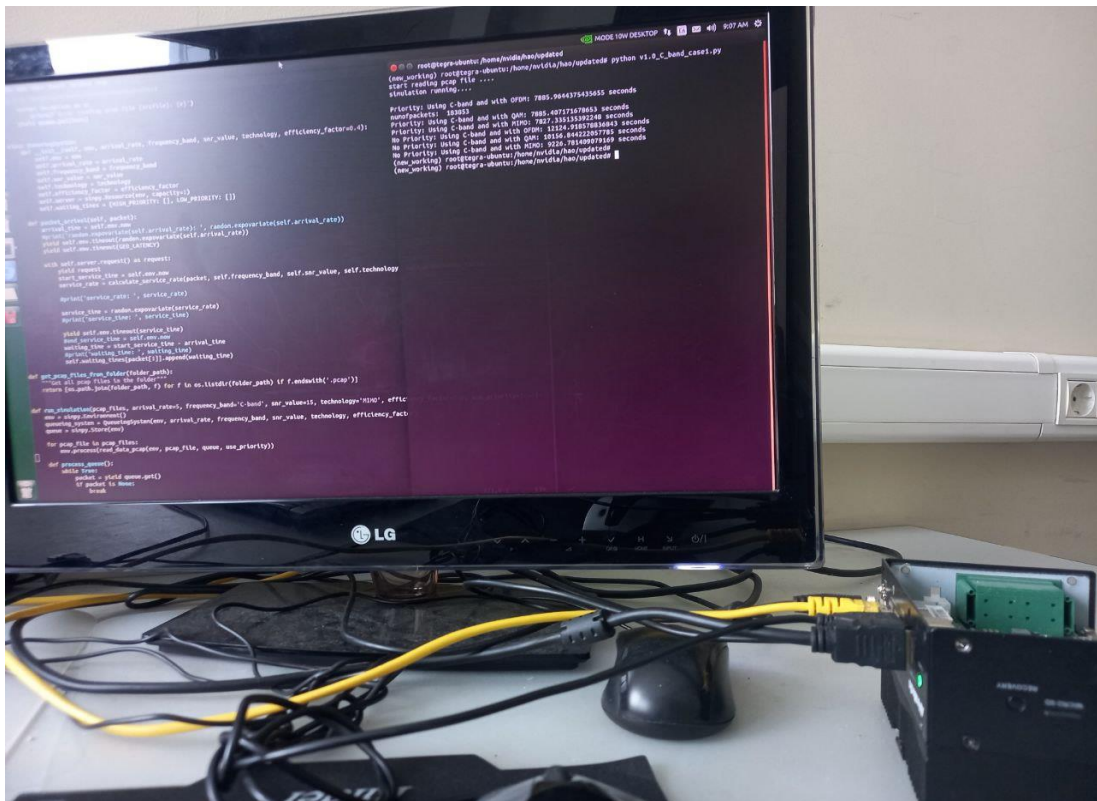


Рисунок. 18. NVIDIA NX215 Carrier Board – AVerMedia, используемая для симуляционных экспериментов

В данной работе мы используем С-диапазон в качестве базовой полосы частот для проведения экспериментов с SimPy. Мы интегрируем передовые технологии, такие как OFDM, QAM и MIMO, чтобы оценить среднее время ожидания каждого пакета в очереди (в миллисекундах) при использовании методов классификации пакетов на основе приоритета и без него. Анализируя поведение очереди и задержку при различных конфигурациях, мы оцениваем влияние приоритизации трафика, управляемой ИИ, на оптимизацию сетевой производительности в геостационарных спутниковых системах.

На рисунке 19 показано среднее время ожидания каждого пакета в очереди (мс) при использовании технологии MIMO с приоритетным и неприоритетным планированием. Результаты демонстрируют, что приоритетное планирование эффективно снижает задержку в очереди. В частности, метод с приоритизацией приводит к среднему времени ожидания 42.66 мс, тогда как метод без приоритизации достигает 50.19 мс.

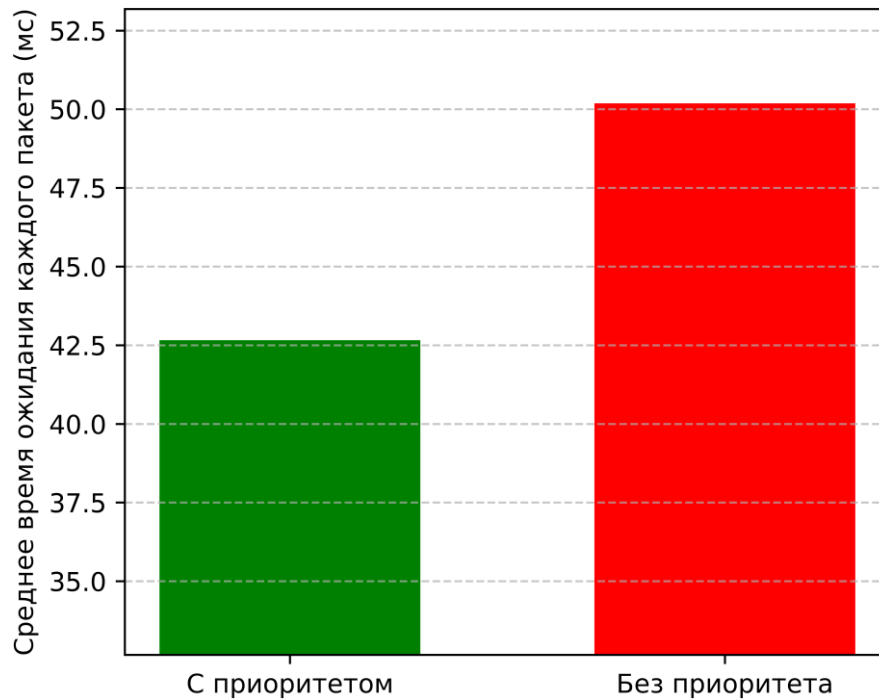


Рисунок. 19. Среднее время ожидания пакета в очереди (миллисекунды) при использовании технологии MIMO

Это означает сокращение времени ожидания в очереди примерно на 14.98% при использовании приоритизации пакетов. Улучшение подтверждает эффективность классификации на основе ИИ в сочетании с динамическим распределением полосы пропускания для оптимизации сетевой производительности, особенно в геостационарных спутниковых системах, где задержка является критически важным фактором.

Рис. 20 демонстрирует сравнение времени ожидания в очереди при использовании различных технологий, таких как OFDM, QAM и MIMO, с приоритетом и без приоритета. Из графика видно, что для всех технологий время ожидания в очереди с приоритетом значительно меньше, чем без приоритета. Конкретно, для OFDM, QAM и MIMO время ожидания с приоритетом составило 42.79, 42.87 и 42.66 соответственно, что заметно ниже, чем время ожидания без приоритета. Это свидетельствует о том, что использование приоритета улучшает время ожидания в очереди, независимо от выбранной технологии.

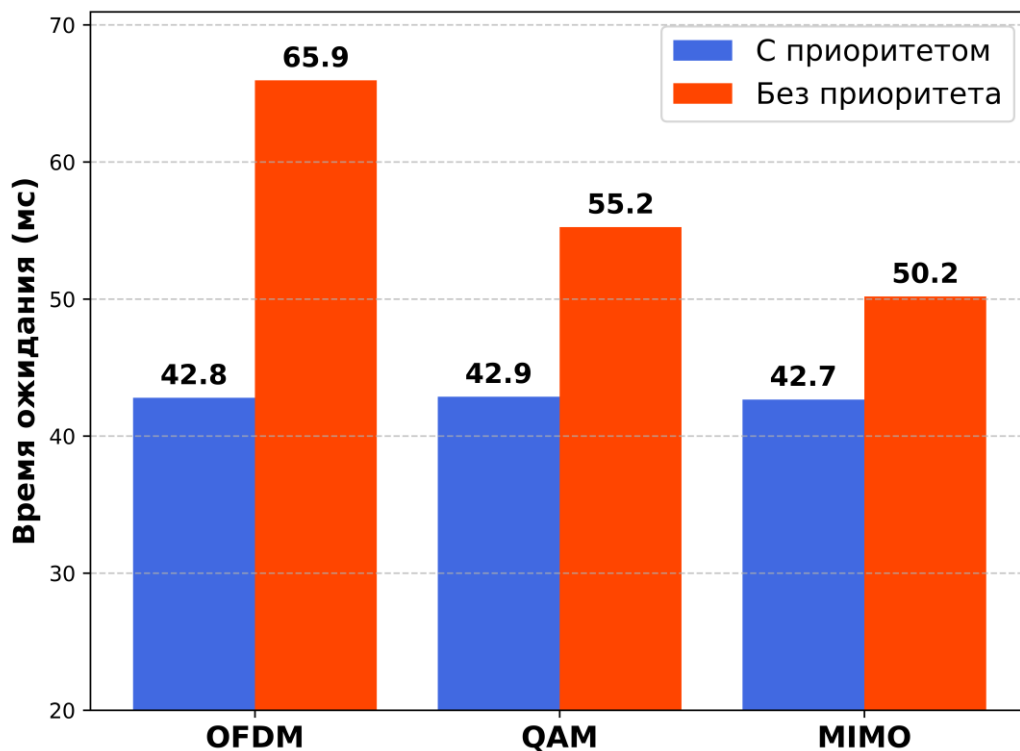


Рисунок. 20. Сравнение времени ожидания в очереди при использовании разных технологий

Таблица 8 показывает процентное сокращение времени ожидания в очереди при использовании метода с приоритетом для различных технологий: OFDM, QAM и MIMO. В таблице сравниваются значения времени ожидания с приоритетом и без приоритета, а также вычисляется процентное уменьшение задержки для каждой технологии. Это позволяет оценить, насколько эффективен метод с приоритетом в сокращении времени ожидания для каждой из выбранных технологий.

Из таблицы видно, что использование метода с приоритетом для всех технологий приводит к значительному сокращению времени ожидания в очереди. Для технологии OFDM время ожидания с приоритетом составило 42.79 миллисекундах, что на 35.14% меньше, чем при отсутствии приоритета (65.95 миллисекундах). Для QAM сокращение составило 22.50%, с 55.24 миллисекундах до 42.87 миллисекундах, а для MIMO — 15.04%, с 50.19 миллисекундах до 42.66 миллисекундах.

Таблица 8. Процентное сокращение времени ожидания в очереди при использовании метода с приоритетом для различных технологий

Технология	Время с приоритетом (мс)	Время без приоритета (мс)	Время без приоритета (мс)
OFDM	42.79	65.95	35.14
QAM	42.87	55.24	22.50
MIMO	42.66	50.19	15.04

Таким образом, метод с приоритетом демонстрирует наибольшее сокращение времени ожидания для технологии OFDM, но все технологии выигрывают от использования приоритетного подхода в плане уменьшения задержек.

4.6 Выводы главы 4

1. В данной главе были рассмотрены методы минимизации сетевой задержки в гибридных спутниковых сетях с использованием технологий искусственного интеллекта, а также проведено исследование распределения ресурсов многолучевых спутников с применением генетических алгоритмов NSGA-II и SPEA2. Экспериментальные результаты показали, что NSGA-II демонстрирует лучшие показатели в минимизации сетевых задержек и потерь пакетов, что важно для повышения качества передачи данных. В то же время, SPEA2 обеспечивает более высокую энергоэффективность, что играет ключевую роль в работе спутниковых систем, особенно в условиях ограниченных энергетических ресурсов.

2. Данное сравнительное исследование позволяет глубже понять компромиссы между потреблением энергии и задержками передачи, предлагая сбалансированный подход к оптимизации гибридных спутниковых сетей. Эти выводы создают основу для дальнейшего развития методов многокритериальной

оптимизации, включая изучение таких аспектов, как экономическая эффективность и влияние на окружающую среду. В перспективе интеграция дополнительных ограничений и целей, таких как экологические последствия, может расширить область применения предложенных методов, что усилит их актуальность для будущих разработок в сфере спутниковых технологий.

ГЛАВА 5. СЖАТИЕ ДАННЫХ БЕЗ ПОТЕРЬ В ГИБРИДНЫХ СПУТНИКОВЫХ СЕТЯХ НА ОСНОВЕ ST-GNN

5.1 Введение: Оптимизация распределения ресурсов в многолучевых спутниковых системах

Спутниковая связь является примером систем с ограниченными ресурсами, которые сталкиваются с ограничениями в спектре, мощности спутников и объеме хранилищ. Эта проблема усугубляется ростом инфраструктур спутниковых сетей на низкой околоземной орбите (LEO). Максимизация эффективности использования спектра, объема хранилищ и распределения мощности спутников при сохранении качества обслуживания (QoS) является критически важной.

Многолучевая антенна (МБА) [62] предлагает решение проблемы ограниченности ресурсов спутниковой связи, используя несколько узконаправленных лучей в отличие от традиционных методов с одним глобальным лучом. Однако повышенная гибкость систем спутников с многолучевой антенной (MBS) [63] также вносит сложности в управление динамическим распределением ресурсов (DRA).

Решение этих проблем требует инновационных подходов, выходящих за рамки традиционных методов. Методы беспотерной компрессии [78] сетевого трафика привлекли внимание как потенциальный способ улучшения производительности спутникового сегмента гибридной орбитально-наземной сети связи. Компрессионные технологии эффективно снимают ограничения по пропускной способности и минимизируют задержки передачи данных, уменьшая объем передаваемой информации при сохранении ее целостности. В этом контексте пространственно-временные графовые нейронные сети (ST-GNN) [79] представляют собой новаторскую и многообещающую стратегию для оптимизации компрессии сетевого трафика. ST-GNN используют пространственно-временные характеристики коммуникационных данных для

захвата сложных паттернов и зависимостей, что делает их особенно подходящими для задач компрессии трафика.

Данное исследование посвящено применению специально разработанных архитектур ST-GNN для оптимизации компрессии трафика и эффективного снижения задержек передачи данных в спутниковой сегменте гибридной орбитально-наземной сети связи. Выполняется глубокий сравнительный анализ влияния различных техник компрессии на задержку в этих сетях. Посредством строгой эмпирической оценки демонстрируются превосходные возможности ST-GNN в достижении конкурентоспособных уровней компрессии по сравнению с существующими методами, что значительно улучшает эффективность передачи данных [80] и снижает задержку.

Настоящая глава посвящена исследованию методов безпотерной компрессии трафика на основе ST-GNN и их роли в минимизации задержек в гибридных спутниково-наземных сетях. Также проводится сравнительный анализ различных техник компрессии и их воздействия на сетевую производительность. Во-первых, представлен метод безпотерной компрессии на основе ST-GNN. Во-вторых, выполняется всесторонний анализ, который проясняет воздействие различных техник компрессии на задержку, предоставляя ценные инсайты для улучшения производительности связи в этих сетях. В совокупности эти результаты подчеркивают практическую значимость предложенного подхода, отмечая его потенциал значительно повысить общую эффективность и надежность гибридных коммуникационных сетей [81, 82].

5.2 Обзор методов сжатия и подходов к снижению задержек в гибридных спутниковых сетях

Высокая задержка остаётся значительной проблемой в спутниковой сегменте гибридной орбитально-наземной сети связи из-за больших расстояний, которые сигналы проходят через спутниковые каналы связи. Различные исследования изучали различные методы снижения задержки, сосредотачиваясь

на оптимизации маршрутов, распределении ресурсов, кэшировании на краю сети и продвинутых методах сжатия. Несмотря на многообещающие результаты, каждый из этих подходов сталкивается с ограничениями и вызовами, особенно в отношении масштабируемости, сложности и практической реализации.

Таблица 9 ниже обобщает ключевые аспекты этих связанных исследований, подчеркивая их направленность, подходы и связанные с ними вызовы или ограничения. Этот обзор закладывает основу для нашего предлагаемого метода, который использует сжатие сетевого трафика без потерь с применением ST-GNN для решения проблем задержки в спутниковой сегменте гибридной орбитально-наземной сети связи.

Таблица 9. Обзор связанных исследований

Область исследования	Ключевые вклад	Ограничения
Оптимизация маршрутов и распределение ресурсов	- Эффективное определение маршрутов для сокращения задержек [83] - Оптимизированное распределение сетевых ресурсов [84]	- Проблемы масштабируемости при увеличении нагрузки на сеть
Кэширование на краю сети и сжатие без потерь	- Сокращение времени передачи данных за счёт кэширования контента и сжатия без потерь [85] - Хранение часто запрашиваемого контента ближе к конечным пользователям	- Ограниченное изучение в гибридных спутниково-наземных контекстах - Высокие затраты на инфраструктуру и её обслуживание
Перенос трафика и адаптивная маршрутизация	- Минимизация задержки от конца до конца за счёт перераспределения трафика и адаптивной маршрутизации [86] - Адаптивная оптимизация маршрутизации для минимизации задержек [87]	- Проблемы масштабируемости при увеличении нагрузки - Сложность модели и проблемы с её реализацией
Методы аппроксимации и Q-обучение	- Оптимальное развертывание спутниковых шлюзов на основе требований к задержке [88] - Алгоритм маршрутизации на	- Высокие затраты на реализацию - Сложная модель, затрудняющая

	основе Q-обучения для снижения задержек [89]	практическую реализацию
Федеративное обучение и вычисления на мобильных краевых узлах спутников	- Совмещение снижения задержек и эффективного использования пропускной способности с помощью глубокого обучения [90] - Решение проблем задержек от конца до конца и энергопотребления в вычислениях на мобильных краевых узлах спутников (SMEC) [91]	- Высокое время вычислений как потенциальный недостаток - Проблемы сложности модели
Графовые нейронные сети в спутниковой связи	- Графовые нейронные сети для снижения задержек от конца до конца [92]	- Сложность модели и требования к времени обработки

Как видно из таблицы, хотя каждое исследование внесло ценный вклад в эту область, всё ещё существует необходимость в комплексном подходе, который интегрирует передовые спутниковые технологии с методами сжатия без потерь для эффективного снижения задержек в гибридных спутниково-наземных сетях. Эта работа стремится закрыть этот пробел, предлагая новый метод, который использует сильные стороны обеих областей, опираясь на существующий массив знаний и устраняя ограничения предыдущих исследований. Таким образом, стремление заключается в разработке более комплексного и практического решения для снижения задержек в спутниковой сегменте гибридной орбитально-наземной сети связи, что открывает путь к улучшению производительности сети и пользовательского опыта.

5.3 Теоретические основы пространственно-временного сжатия данных

5.3.1 Архитектура гибридных спутниково-наземных сетей

Спутниковой сегменте гибридной орбитально-наземной сети связи представляют собой парадигмальный сдвиг в современной коммуникационной инфраструктуре [93]. Они интегрируют сильные стороны как спутниковых, так и

наземных сетей, чтобы предложить улучшенную связь и надёжность. Используя комбинацию спутниковых и наземных технологий связи, спутниковой сегменте гибридной орбитально-наземной сети связи преодолевают ограничения, связанные с использованием только спутниковых или только наземных систем.

Спутниковой сегменте гибридной орбитально-наземной сети связи обеспечивают бесшовную передачу данных по обширным географическим регионам, поддерживая различные приложения, такие как телекоммуникации, вещание, дистанционное зондирование и доступ в интернет. Интеграция этих технологий позволяет улучшить покрытие, особенно в удалённых или недостаточно обслуживаемых районах, где традиционные наземные сети могут быть недостаточны [94]. Интегрируя наземные компоненты в архитектуру сети, спутниковой сегменте гибридной орбитально-наземной сети связи предлагают жизнеспособное решение для смягчения проблем задержек, направляя данные через наземные каналы связи, когда это возможно, таким образом, уменьшая зависимость от спутниковых каналов и сокращая время передачи данных в оба конца.

На рисунке 21 представлена архитектура спутникового сегмента гибридной орбитально-наземной сети связи, которая обычно включает взаимосвязанные спутниковые и наземные узлы, соединённые посредством высокоскоростных каналов передачи данных. Такая архитектура способствует эффективной маршрутизации и передаче данных, обеспечивая надёжную и стабильную связь.

5.3.2 Модель канала связи в спутниковых системах

Производительность спутниковых коммуникационных каналов оценивается с помощью уравнения бюджета канала, которое рассчитывает мощность принимаемого сигнала на наземном терминале, учитывая усиления и потери вдоль пути передачи. Уравнение бюджета канала имеет следующий вид:

$$P_r = \frac{P_t G_t G_r}{L_{fs} L_m} \quad (40)$$

где:

- P_r : принимаемая мощность
- P_t : мощность передатчика
- G_t : усиление антенны передатчика
- G_r : усиление антенны приёмника
- L_{fs} : потери на свободном пространстве
- L_m : прочие потери (включая атмосферные затухания и потери на

оборудование)

Усиления антенн передатчика и приёмника рассчитываются по формуле

$$G = \eta \left(\frac{\pi D}{\lambda} \right)^2 \quad (41)$$

где:

- η : Эффективность антенны
- D : Диаметр антенны
- λ : Длина волны сигнала

Потери на свободном пространстве, представляющие собой затухание сигнала из-за расстояния, рассчитываются по формуле: $L_{fs} = \left(\frac{4\pi d}{\lambda} \right)^2$, где d — расстояние распространения сигнала.

Достижимая скорость передачи данных в коммуникационном канале напрямую связана с Отношением Сигнала к Шуму (SNR). В спутниковой связи мощность шума N_p рассчитывается по формуле: $N_p = N_0 B^{(n)}$, где N_0 : — плотность спектра мощности шума, $B^{(n)}$ — выделенная полоса пропускания для ячейки n .

Плотность спектра мощности шума далее определяется как $N_0 = k_b T_{sys}$, где k_b — постоянная Больцмана [95], T_{sys} — температура системного шума (включая вклад шума антенны, атмосферы и приёмника).

Учитывая интерференцию по одному каналу $I_{\text{cci}}^{(n)}$ от других ячеек, Отношение Сигнала к Интерференции и Шуму ($\text{SINR}^{(n)}$) для ячейки n рассчитывается по формуле:

$$\text{SINR}^{(n)} = \frac{P_r^{(n)}}{N_p + I_{\text{cci}}^{(n)}} \quad (42)$$

Наконец, применяя теорему о ёмкости Шеннона [96], можно определить теоретически максимальную скорость передачи данных $R^{(n)}$, достижимую в ячейке n :

$$R^{(n)} = B^{(n)} \log_2 (1 + \text{SINR}^{(n)}) \quad (43)$$

5.3.3 Основы пространственно-временных графовых нейронных сетей

Пространственно-временные графовые нейронные сети (ST-GNN) [97] — это передовые архитектуры нейронных сетей, предназначенные для эффективного моделирования и эволюции данных, структурированных в виде графов. Эти сети интегрируют пространственные и временные измерения, что делает их особенно подходящими для приложений с критическими пространственными зависимостями и временной динамикой. Используя Графовые Свёрточные Сети (GCN) [98] для захвата пространственных взаимосвязей и Рекуррентные Нейронные Сети (RNN) или временные свёртки для обработки временных последовательностей, ST-GNN могут обрабатывать сложные наборы данных, такие как те, что встречаются в прогнозировании трафика, распознавании человеческой активности и мониторинге окружающей среды.

ST-GNN предназначены для моделирования и обработки данных с пространственными и временными зависимостями, особенно в данных, структурированных в виде графов. Математическая формулировка ST-GNN может варьироваться в зависимости от конкретных архитектур и приложений, но обобщённое представление выглядит следующим образом:

Пусть $G = (V, E)$ — это граф, где V представляет собой множество узлов, а E — множество рёбер. Каждый узел $v_i \in V$ имеет связанный вектор признаков $x_i \in \mathbb{R}^d$, где d — размерность пространства признаков. Структура графа может быть представлена матрицей смежности A , где $A_{ij} = 1$, если существует ребро между узлами v_i и v_j , и $A_{ij} = 0$ в противном случае.

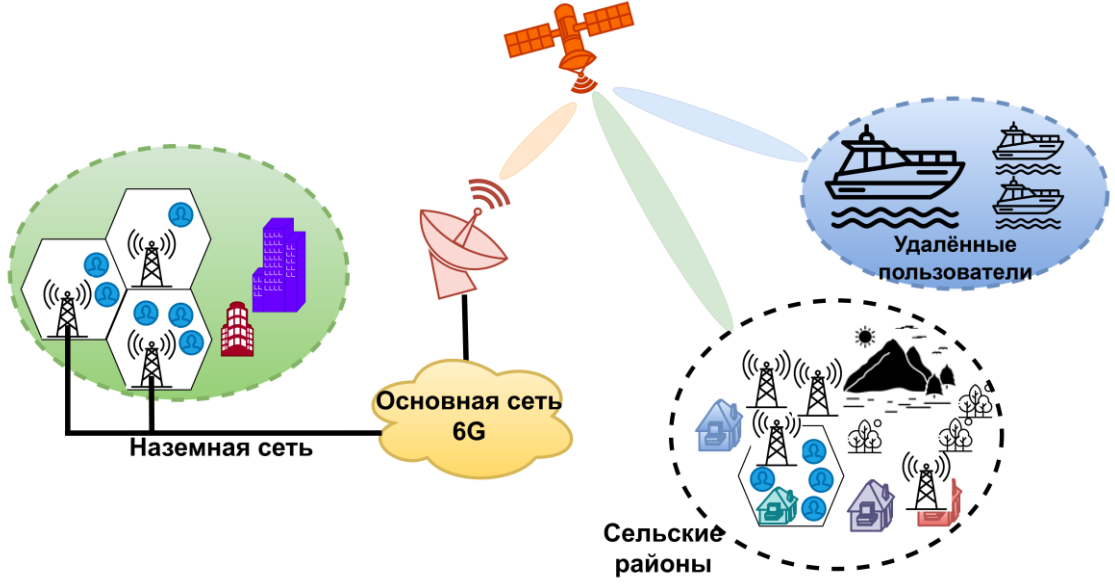


Рисунок 21. Архитектура спутникового сегмента гибридной орбитально-наземной сети связи

Пространственно-временная графовая нейронная сеть работает с последовательностями снимков графов, где каждый снимок представляет состояние графа в определённый момент времени. Пусть T — общее количество временных шагов в последовательности.

На каждом временном шаге t ST-GNN обрабатывает снимок графа $G_t = (V, E_t)$, где E_t представляет собой множество рёбер, присутствующих в графе в момент времени t . Признаки узлов обновляются на основе как пространственной, так и временной информации. Функция обновления может быть сформулирована следующим образом:

$$h_i^{(t)} = f^{(t)} \left(h_i^{(t-1)}, \sum_{j \in N(i)} g^{(t)} \left(h_i^{(t-1)}, h_j^{(t-1)} \right) \right) \quad (44)$$

где:

- $h_i^{(t)}$: представляет собой скрытое состояние узла v_i в момент времени t .
- $N(i)$: обозначает соседство узла v_i в снимке графа G_t .
- $f^{(t)}$: функция обновления узла, которая учитывает временную информацию в момент времени t .
- $g^{(t)}$: функция передачи сообщений, которая захватывает пространственные зависимости между соседними узлами в момент времени t .

После обработки всех узлов в снимке графа ST-GNN может выдавать прогнозы, классификации или другие представления на основе конечных скрытых состояний узлов. Конкретные формы $f^{(t)}$ и $g^{(t)}$ могут варьироваться в зависимости от архитектуры и решаемой задачи. Обычные варианты включают многослойные перцептроны (MLP) [99], рекуррентные нейронные сети (RNN) [100] и семейство алгоритмов RNN, таких как Долгосрочная Кратковременная Память (LSTM) [101] и Гейты Рекуррентного Узла (GRU) [102].

5.3.4 Подходы к сжатию данных без потерь в сетевом трафике

Сжатие [103] — это способ представления больших объемов данных, таких как числовой текст, изображения, видео, аудио или любые другие типы, с использованием меньшего размера данных. Сжатие делится на сжатие с потерями и без потерь. Сжатие с потерями означает, что декомпрессированные данные отличаются от оригинала, в то время как сжатие без потерь идентично как оригинальным, так и декомпрессированным данным. Выбор между двумя типами методов или алгоритмов сжатия зависит от типа данных, которые нужно сжать. Например, чтобы сжать изображение с помощью метода с потерями, нужно сохранить только достаточно информации, чтобы понять, что изображено на картине, например, автомобиль или человека. В отличие от этого, сжатие без потерь неподходящее для чувствительных данных, таких как финансовые или избирательные данные, которые используются для облегчения передачи через

интернет или хранения данных на USB-накопителях. Поэтому сжатие без потерь используется, когда каждое отдельное бит данных критически важно; в противном случае используется сжатие с потерями.

Методы сжатия сетевого трафика без потерь [104] предназначены для уменьшения объема данных, передаваемых по сети, без потери информации. Эти методы используют различные алгоритмы для выявления и устранения избыточностей в пакетах данных, обеспечивая возможность точной реконструкции оригинальных данных из сжатых. Оптимизация передачи данных в сетевой коммуникации [105] критична для обеспечения эффективного использования пропускной способности и снижения задержек. Техники сжатия без потерь направлены на уменьшение размера пакетов данных без ущерба для целостности передаваемой информации. Выявляя и удаляя избыточности в потоке данных, эти методы уменьшают объем передаваемой информации, сохраняя оригинальное содержание.

Арифметическое кодирование (AC) [106] выделяется среди алгоритмов, разработанных для сжатия без потерь, благодаря своей эффективности и универсальности. В отличие от традиционных методов, таких как кодирование Хаффмана, которое работает по принципу символ за символом, арифметическое кодирование кодирует всю последовательность символов как единое целое, предлагая лучшие коэффициенты сжатия.

Пусть S обозначает входную последовательность символов, а $P(s_i)$ представляет вероятность появления символа s_i в контексте предыдущих символов. Изначально единичный интервал $[0,1)$ делится на подс intervals, соответствующие каждому символу s_i в соответствии с их вероятностями. Длина каждого подс интервала пропорциональна вероятности связанного символа:

$$\text{Interval}(s_i) = [\text{Low}(s_i), \text{High}(s_i)) \quad (45)$$

$$\text{High}(s_i) = \sum_{j=1}^i P(s_j) \quad (46)$$

$$\text{Low}(s_i) = \sum_{j=1}^{i-1} P(s_j) \quad (47)$$

Во время кодирования, по мере обработки каждого символа s_i , текущий интервал дополнительно делится на подс интервал, связанный с s_i . Это деление выполняется так, чтобы размер интервала соответствовал вероятности последовательности символов, представленной обработанными символами. Результирующий подс интервал становится новым интервалом кодирования для последующих символов.

Процесс декодирования обращает эту процедуру, начиная с дробного числа, полученного из закодированных данных. Оригинальная последовательность символов может быть точно восстановлена, итеративно деля единичный интервал и определяя подс интервал, соответствующие декодированным символам.

5.3.5 Использование рекуррентных нейронных сетей в задаче сжатия

RNN (рекуррентные нейронные сети) — это класс искусственных нейронных сетей, специально разработанных для обработки последовательных данных. В отличие от традиционных прямых сетей, RNN включают цикл, который позволяет им обрабатывать информацию из предыдущих шагов последовательности, захватывая временные зависимости. Эта способность делает их идеальными для задач моделирования языка, распознавания речи и прогнозирования временных рядов.

а) Архитектура Vanilla RNN

Vanilla RNN — это фундаментальная архитектура, предназначенная для обработки последовательных данных с помощью рекуррентных связей, которые передают информацию во времени. Эта архитектура особенно эффективна для задач, где порядок данных имеет значение, поскольку она использует информацию из прошлого для информирования текущих шагов обработки. Если дана входная последовательность $X = \{x_1, x_2, \dots, x_T\}$, где x_t представляет входные данные на временном шаге t , скрытые состояния $H = \{h_1, h_2, \dots, h_T\}$, где h_t

представляет скрытое состояние на временном шаге t , и выходная последовательность $Y = \{y_1, y_2, \dots, y_T\}$, где y_t представляет выходные данные на временном шаге t , динамика Vanilla RNN может быть описана следующими уравнениями.

Уравнение обновления скрытого состояния:

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (48)$$

где W_{hx} — это матрица весов для соединений от входного состояния к скрытому состоянию, W_{hh} — матрица весов для соединений от скрытого состояния к скрытому состоянию, а b_h — смещение для скрытого состояния. Функция активации σ обычно представляет собой нелинейную функцию, такую как сигмоида или гиперболический тангенс.

Уравнение вычисления выхода:

$$y_t = \text{softmax}(W_{yh}h_t + b_y) \quad (49)$$

где W_{yh} — это матрица весов для соединений от скрытого состояния к выходу, а b_y — смещение для выхода. Функция softmax применяется для преобразования выходных данных в вероятности, что особенно полезно для задач классификации.

Vanilla RNN предоставляют надежную основу для моделирования последовательных данных, хотя они могут сталкиваться с проблемами, такими как исчезающие и взрывающиеся градиенты, что может ограничивать их способность изучать долгосрочные зависимости. Несмотря на эти сложности, их фундаментальные принципы проектирования служат основой для более сложных архитектур, таких как LSTM и GRU, которые решают эти ограничения и увеличивают способность моделировать сложные последовательности.

б) Долгосрочная кратковременная память (LSTM)

Сети LSTM были разработаны для преодоления ограничений Vanilla RNN, особенно проблемы исчезающих градиентов, что затрудняет Vanilla RNN обучение долгосрочным зависимостям. LSTM решают эту проблему с помощью специализированных механизмов управления, которые регулируют поток информации внутри сети. Единица LSTM состоит из нескольких ворот, которые

контролируют обновления состояния: входные ворота, ворота забывания и выходные ворота. Эти ворота определяют, какую информацию сохранять, удалять или выводить на каждом временном шаге. Математические операции внутри единицы LSTM определяются следующим образом:

Входные ворота (i_t) :

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (50)$$

Ворота забывания (f_t) :

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad (51)$$

Ворота ячейки (g_t) :

$$g_t = \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \quad (52)$$

Обновление состояния ячейки (c_t) :

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (53)$$

Выходные ворота (o_t) :

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_t + b_o) \quad (54)$$

Обновление скрытого состояния (h_t) :

$$h_t = o_t \odot \tanh(c_t) \quad (55)$$

Входные ворота контролируют, сколько новой информации из текущего входа и предыдущего скрытого состояния должно быть добавлено к состоянию ячейки. Ворота забывания определяют, сколько из предыдущего состояния ячейки должно быть перенесено в текущее состояние ячейки. Ворота ячейки создают новое кандидатное состояние ячейки, которое может быть добавлено к текущему состоянию ячейки. Состояние ячейки обновляется путем комбинирования предыдущего состояния ячейки, модулированного воротами забывания, с новым кандидатным состоянием ячейки, модулированным входными воротами. Здесь $\odot$ обозначает поэлементное умножение. Выходные ворота контролируют, сколько состояния ячейки должно быть передано в следующее скрытое состояние. Скрытое состояние обновляется путем модулирования нового состояния ячейки с помощью выходных ворот.

Матрицы весов $(W_{ix}, W_{ih}, W_{ic}, W_{fx}, W_{fh}, W_{fc}, W_{gx}, W_{gh}, W_{ox}, W_{oh}, W_{oc})$ являются обучаемыми параметрами, которые определяют, как входные данные, предыдущие скрытые состояния и предыдущие состояния ячеек влияют на текущее состояние ворот. Смещения b_i, b_f, b_g, b_o являются дополнительными обучаемыми параметрами, которые позволяют воротам более точно подгонять данные, смещая функции активации.

Функции активации σ and $\tanh$: Сигмоидальная функция (σ) используется в воротах для получения выходных значений между 0 и 1, указывая, сколько информации должно быть пропущено. Гиперболический тангенс $\tanh$ используется для создания кандидатных состояний ячеек со значениями между -1 и 1. Ворота (i_t, f_t, o_t, g_t) контролируют поток информации внутрь, наружу и внутри единицы LSTM, обеспечивая сохранение важной информации на длительный период и отбрасывание неактуальной информации.

в) Управляемые рекуррентные блоки (GRU)

GRU предлагает упрощённую, но эффективную альтернативу LSTM, сводя несколько механизмов управления LSTM к двум воротам: воротам обновления и воротам сброса. Это упрощение помогает GRU сохранять производительность, одновременно снижая вычислительную сложность. GRU разработаны для адаптивного захвата зависимостей в последовательных данных, что делает их эффективными для различных задач временных рядов и основанных на последовательностях.

Единица GRU использует два ворот для управления потоком информации: ворота обновления и ворота сброса. Эти ворота помогают GRU решать, какую информацию сохранять, а какую удалять на каждом временном шаге. Математические операции внутри единицы GRU определяются следующим образом:

Ворота обновления (z_t) :

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z) \quad (56)$$

Ворота сброса (r_t) :

$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r) \quad (57)$$

Активация кандидата ($\tilde{h}_t$) :

$$\tilde{h}_t = \tanh(W_x x_t + W_h(r_t \odot h_{t-1}) + b_h) \quad (58)$$

Обновление скрытого состояния (h_t) :

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (59)$$

Ворота обновления определяют, насколько сильно предыдущее скрытое состояние должно быть обновлено новой информацией. Они контролируют баланс между предыдущей памятью и текущим входом. Ворота сброса решают, насколько из предыдущего скрытого состояния следует забыть при включении новой информации. Они позволяют модели избирательно сбрасывать свою память.

Активация кандидата создает новое кандидатное скрытое состояние на основе текущего входа и предыдущего скрытого состояния, модифицированного воротами сброса. Здесь $\odot$ обозначает поэлементное умножение. Скрытое состояние обновляется путем смешивания предыдущего скрытого состояния с кандидатным скрытым состоянием, как контролируется воротами обновления. Этот механизм позволяет GRU сохранять долгосрочные зависимости, одновременно включая новую информацию.

Где матрицы весов ($W_{zx}, W_{zh}, W_{rx}, W_{rh}, W_x, W_h$): Эти матрицы являются обучаемыми параметрами, которые определяют, как входные данные и предыдущие скрытые состояния влияют на ворота и кандидатное скрытое состояние. Смещения (b_z, b_r, b_h): Эти дополнительные обучаемые параметры позволяют воротам точнее подгонять данные, смещая функции активации.

Функции активации (σ and $\tanh$): Сигмоидная функция (σ) используется в воротах для получения выходов в диапазоне от 0 до 1, что указывает на то, сколько информации следует пропускать. Гиперболический тангенс ($\tanh$) используется для создания кандидатных скрытых состояний со значениями от -1 до 1. Ворота (z_t, r_t): Эти ворота контролируют поток информации внутри блока

GRU, обеспечивая сохранение важной информации на длительные периоды времени и удаление несущественной информации.

5.4 Методология предложенного метода сжатия

Рисунок 22 представляет собой подробное визуальное представление рабочего процесса, связанного с реализацией метода сжатия сетевого трафика без потерь в спутниковой сегменте гибридной орбитально-наземной сети связи.



Рисунок 22. Основной поток сжатия и передачи данных

Иллюстрация охватывает весь процесс, включающий сжатие данных сетевого трафика, их передачу через инфраструктуру спутникового сегмента гибридной орбитально-наземной сети связи и последующую декомпрессию.

5.4.1 Алгоритм сжатия данных с использованием ИИ

Рис. 23 представляет собой подробный обзор модуля компрессора на основе искусственного интеллекта, объясняя его ключевые функции и иллюстрируя входные и выходные данные, связанные с сжатием сетевого трафика.

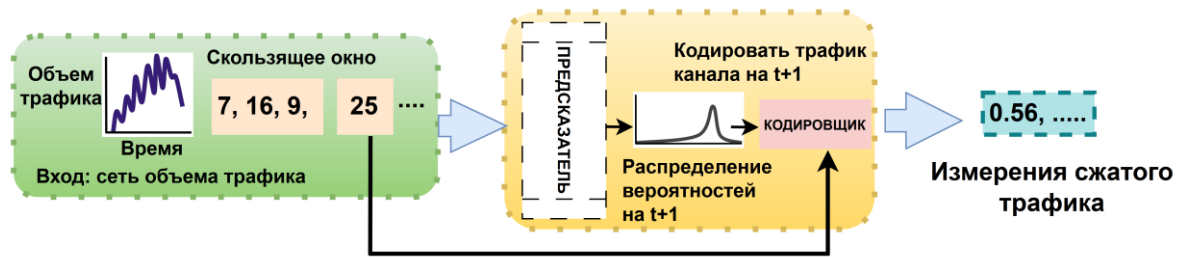


Рисунок 23. Обзор модуля компрессора на основе искусственного интеллекта с его входными и выходными данными.

Модуль принимает данные о трафике, организованные в дискретные временные интервалы, например, 5-минутные сегменты. Используя технику скользящего окна, модуль систематически проходит через эти временные интервалы, обрабатывая каждый сегмент для прогнозирования распределения вероятностей будущих интервалов. Этот прогнозный анализ осуществляется с помощью модели искусственного интеллекта, которая использует продвинутые алгоритмы для выявления закономерностей и прогнозирования будущих тенденций данных.

После фазы прогнозирования модуль компрессора использует кодировщик, который применяет алгоритм арифметического кодирования для кодирования прогнозируемых измерений трафика из каждого временного интервала. Результатом является сжатое представление данных трафика, что уменьшает объем данных. Этот процесс создает сжатые измерения трафика для каждого временного интервала, достигая основной цели оптимизации использования полосы пропускания и повышения эффективности сети с помощью методов компрессии на основе ИИ.

5.4.2 ST-GNN как предсказательная модель для сжатия

Рисунок 24 предоставляет подробное изложение механизма *ПРЕДСКАЗАТЕЛЯ* на основе ST-GNN, объясняя его рабочий процесс по предсказанию распределения вероятностей в скользящем окне размером два.

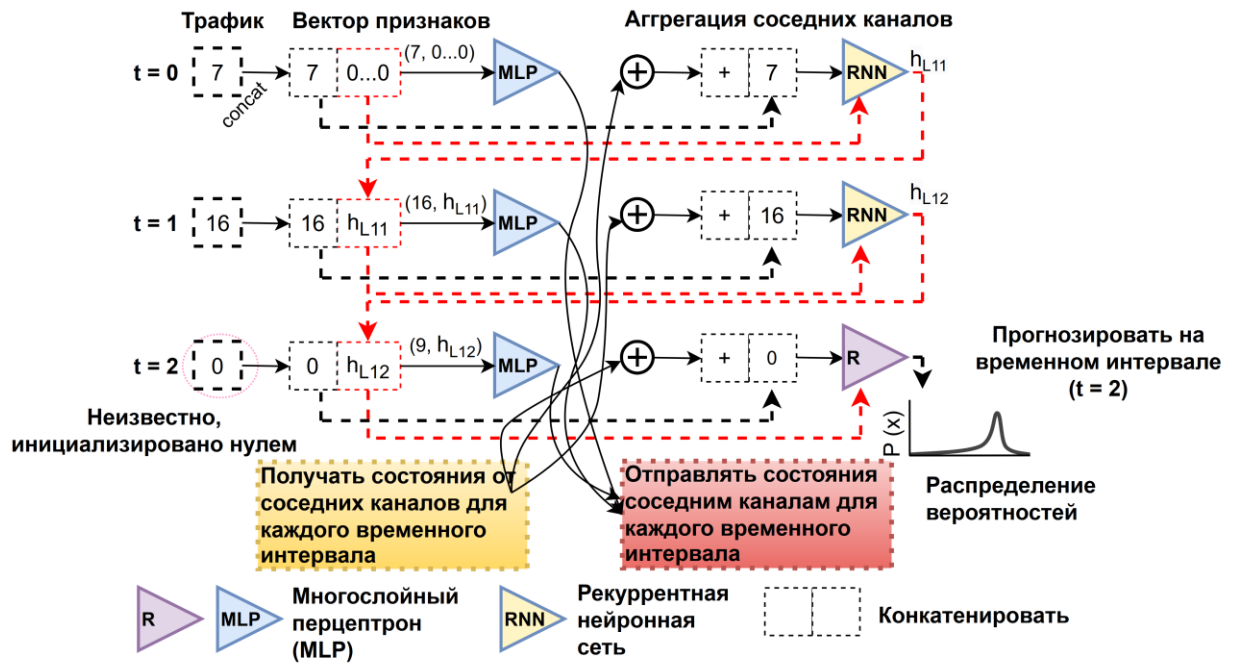


Рисунок 24. Обзор модуля компрессора на основе ИИ с его входами и выходами

Процесс начинается с инициализации векторных признаков для связей на временном интервале $t = 0$. Эти векторы включают значения трафика и заполняются нулями для упрощения последующих вычислений. Инициализированные векторы признаков затем обрабатываются многослойным перцептроном (MLP), и выходные данные от MLP распространяются на соседние связи. Эти распространенные выходные данные агрегируются для формирования агрегированного представления состояния. Одновременно текущая связь включает скрытые состояния от своих соседей в свое внутреннее представление наряду с фактическим значением трафика. Это расширенное представление проходит дальнейшую обработку с помощью RNN, что приводит к получению конечного скрытого состояния для текущего временного интервала. Это конечное скрытое состояние служит основой для инициализации вектора признаков в последующем временном интервале, продолжая тем самым цикл предсказания в рамках скользящего окна.

Более того, конечное скрытое состояние, полученное на последнем временном интервале в пределах скользящего окна, обрабатывается другим MLP (обозначенным как R), который выдает параметры распределения вероятностей,

указывающие на прогнозируемые тенденции трафика. Эти параметры распределения вероятностей затем используются алгоритмом арифметического кодирования для сжатия значений трафика для текущей связи на временном интервале $t = 2$.

Обратите внимание, что в данном исследовании также изучается эффективность различных вариантов RNN, таких как LSTM и GRU, в сжатии данных сетевого трафика. Включение этих вариантов направлено на выявление наиболее эффективных алгоритмических фреймворков для повышения эффективности сжатия и улучшения работы сети.

5.4.3 Процесс кодирования и декодирования данных

Важно подчеркнуть, что предложенный метод обеспечивает сжатие данных без потерь (lossless), что означает абсолютно точное восстановление исходных данных после декодирования. Этот принцип гарантируется разделением ролей между нейронной сетью и классическим алгоритмом энтропийного кодирования (например, арифметическим кодированием).

Роль нейронной сети ST-GNN в данной архитектуре — это роль интеллектуального предсказателя. Она не генерирует сжатые данные напрямую. Вместо этого, на каждом шаге модель анализирует предыдущие данные и их пространственную структуру, чтобы сформировать распределение вероятностей для следующего возможного значения в потоке данных. Например, модель может предсказать: "следующее значение будет '7' с вероятностью 90%, '16' с вероятностью 5%, и т.д."

Затем это распределение вероятностей передается на вход стандартному алгоритму энтропийного кодирования. Этот алгоритм, являясь по своей природе безпотерным, выполняет фактическое сжатие: он присваивает более короткие битовые последовательности более вероятным значениям и более длинные — менее вероятным.

Процесс декодирования происходит в обратном порядке. Декодер, имея идентичную и синхронизированную модель ST-GNN, на каждом шаге генерирует точно такое же распределение вероятностей, что и кодер. Получив короткую битовую последовательность, он может однозначно и без потерь восстановить исходное значение ('7'), так как знает, какому символу эта последовательность соответствует в данном контексте вероятностей. Таким образом, нейронная сеть отвечает за повышение эффективности сжатия, а классический энтропийный кодер – за его безошибочность и обратимость.

Рисунок 25 предоставляет всесторонний обзор процесса сжатия, подчеркивая ключевую роль модуля предсказания, который реализован с использованием вариантов алгоритма RNN. В этой реализации модуль предсказания построен с использованием ST-GNN, что позволяет использовать значения трафика соседних ссылок в качестве дополнительных входных признаков для повышения точности прогнозирования.

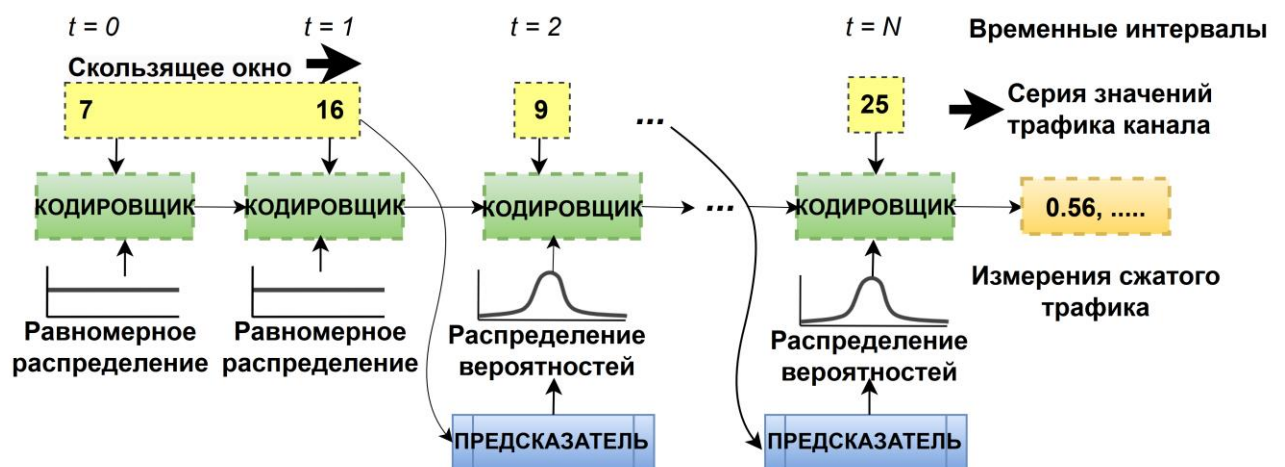


Рисунок 25. Обзор модуля сжатия на основе ИИ с его входами и выходами

Процесс сжатия начинается с модуля предсказания, который обрабатывает данные трафика в пределах скользящего окна. Первоначальные значения трафика в этом окне кодируются с использованием равномерных распределений вероятностей. Такой подход облегчает эффективное представление данных и сжатие в рамках управления сетевым трафиком. Используя ST-GNN, модуль предсказания улучшает точность распределений вероятностей для последующих

временных интервалов. Затем модуль кодирования использует эти распределения вероятностей для эффективного сжатия значений трафика.

Процесс в деталях:

- Инициализация с равномерным распределением: на $t = 0$ значения трафика кодируются с использованием равномерного распределения
- Предсказание и кодирование: когда окно перемещается на $t = 2$, предсказатель генерирует распределение вероятностей для значений трафика на основе предыдущих данных и значений соседних ссылок. Кодировщик затем использует это предсказанное распределение для сжатия данных.
- Продолжение сжатия: этот процесс продолжается для последующих временных интервалов ($t = N$), при этом предсказатель постоянно улучшает точность распределений вероятностей, а кодировщик сжимает данные о трафике соответственно.

5.4.4 Разработка декомпрессора на основе ИИ

В фазе декомпрессии модель использует параллельный подход к процессу компрессии, применяя ту же модель на основе нейронной сети (NN) [107] для восстановления оригинальной последовательности поэтапно из сжатого файла. Изначально временные интервалы проходят декомпрессию с использованием равномерных распределений вероятностей [108], аналогично начальному кодированию во время компрессии. Затем модуль предсказания использует восстановленные значения для вычисления распределений вероятностей для связей, отражая поток информации, наблюдаемый во время компрессии. Каждый временной интервал предоставляет идентичную входную информацию для предсказателя, как это было во время компрессии, что обеспечивает последовательность в предсказательных механизмах на обоих этапах.

Аналогично, модуль декодирования, получая ту же входную информацию, что и на этапе компрессии, выполняет инвертированные операции,

соответствующие декодированию. Этот процесс обеспечивает точное восстановление оригинальной последовательности трафика, сохраняя верность её состоянию до компрессии. Этот синхронизированный подход подчеркивает систематическую организацию операций компрессии и декомпрессии в рамках управления сетевым трафиком. Передовые модели на основе нейронных сетей и предсказательные алгоритмы способствуют бесшовной передаче и восстановлению данных, улучшая эффективность и надежность современных коммуникационных инфраструктур.

5.5 Экспериментальные исследования и анализ результатов

В данном исследовании стремится исследовать и сравнить эффективность различных метрик, связанных с коэффициентами сжатия. Это сравнение особенно актуально в контексте PREDICTOR, как показано на рисунке 22, где оцениваются возможности сжатия моделей LSTM и GRU. Исследование основывается на наборе данных, упомянутом в статье [7], который служит основой для эмпирического анализа.

Чтобы обеспечить надежность и обобщаемость, наша программа обучения следует стандартному протоколу. Мы выделяем 70% набора данных для обучения и оставляем оставшиеся 30% для строгой оценки. Эксперименты делятся на две основные области:

- Эксперимент 1: В этом эксперименте проводится сравнительный анализ коэффициентов сжатия между вариантами LSTM и GRU. Оцениваются такие метрики, как среднее время кодирования и декодирования, а также различия в размере моделей, чтобы оценить производительность.

- Эксперимент 2: Этот эксперимент исследует влияние сжатия на время задержки в спутниковой сегменте гибридной орбитально-наземной сети связи. Сравнивая время задержки до и после применения методов сжатия, оценивается эффективность данных методов в реальных сценариях передачи данных.

5.5.1 Оценка эффективности предложенного метода (Эксперимент 1)

В первом эксперименте была проведена оценка нескольких критических метрик для моделей сжатия сетевого трафика. В частности, были сравниваемы параметры, такие как размер модели, среднее время кодирования, среднее время декодирования и коэффициент сжатия. Этот анализ был проведен в рамках модели сжатия сетевого трафика, основанной на ST-GNN, с особым акцентом на использование PREDICTOR, реализованного через архитектуры GRU и LSTM.

Таблица 10 представляет собой всесторонний обзор сравнительных показателей производительности для двух различных конфигураций в модели сжатия сетевого трафика, использующих архитектуры GRU и LSTM, с PREDICTOR в качестве основной механики. Замечено, что размер модели записан в килобайтах, причем конфигурация GRU имеет немного меньший размер – 230 килобайт по сравнению с конфигурацией LSTM, размер которой составляет 290 килобайт. Это различие в размере модели можно отнести к присущим архитектурным различиям между GRU и LSTM, где первая обычно требует меньшего количества параметров, что способствует более компактному представлению модели.

Таблица 10. Сравнительные показатели производительности для двух различных конфигураций PREDICTOR.

Предсказатель	Размер модели (кб)	Среднее время кодирования (с)	Среднее время декодирования (с)	Коэффициент сжатия
LSTM	290	0.933	0.496	2.426
GRU	230	0.946	0.482	3.397

Более того, среднее время кодирования и декодирования дает представление о вычислительной эффективности соответствующих конфигураций. Реализация на основе GRU демонстрирует немного более короткое среднее время кодирования – примерно 0,946 секунд, по сравнению с временем кодирования конфигурации LSTM, которое составляет примерно 0,933

секунд. Аналогично, в контексте среднего времени декодирования конфигурация GRU показывает несколько лучшие результаты, требуя примерно 0,482 секунд, в то время как декодирование LSTM занимает примерно 0,496 секунд.

Однако стоит отметить, что GRU демонстрирует более высокий коэффициент сжатия, примерно 3.397, что указывает на его способность достигать более эффективного сжатия данных. В то время как конфигурация LSTM показывает коэффициент сжатия примерно 2.426, что подразумевает несколько менее эффективный механизм сжатия. Эти результаты подчеркивают тонкие компромиссы, связанные с выбором между архитектурами GRU и LSTM в контексте сжатия сетевого трафика, где такие аспекты, как вычислительная эффективность и эффективность сжатия, требуют тщательного рассмотрения.

Эти результаты можно объяснить несколькими факторами:

- Архитектурная эффективность: Архитектуры GRU обычно более эффективны с точки зрения параметров по сравнению с LSTM, что может способствовать их меньшему размеру модели и немного лучшим показателям времени кодирования и декодирования. GRU разработаны быть более простыми, что часто приводит к более быстрым вычислениям и меньшим требованиям к памяти.

- Сложность модели: Более простая структура GRU, использующая меньшее количество ворот по сравнению с LSTM, может привести к более эффективным процессам обучения и инференса. Эта простота может позволить GRU более эффективно сжимать данные, что приводит к более высоким коэффициентам сжатия, наблюдаемым в результатах.

- Временные зависимости: GRU могут более эффективно обрабатывать определенные временные зависимости в данных сетевого трафика по сравнению с LSTM. Это может объяснить их превосходные результаты в отношении коэффициента сжатия, поскольку они могут захватывать основные паттерны в данных с меньшим количеством параметров.

– Динамика обучения: Динамика обучения и поведение сходимости GRU могут быть более благоприятными для конкретного набора данных и задач, рассмотренных в этом исследовании. Это может привести к лучшему обобщению и производительности во время фаз сжатия и декомпрессии.

5.5.2 Сравнение с существующими подходами (Эксперимент 2)

Второй эксперимент направлен на оценку эффективности уменьшения времени задержки в спутниковой сегменте гибридной орбитально-наземной сети связи путем применения предложенной процедуры сжатия. Эта оценка сравнивает время задержки до и после применения методов сжатия, предоставляя информацию о влиянии этих техник на производительность сети.

Таблица 11 детализирует основные параметры для оценки времени задержки в спутниковой сегменте гибридной орбитально-наземной сети связи, включая высоту орбиты, частоту передачи данных, общую полосу пропускания, общую мощность спутника, спектральную плотность мощности шума, усиление передающей антенны, усиление приемной антенны и потери на пути передачи.

Таблица 11. Параметры симуляционного сценария

Параметр	Значение
Высота орбиты h (км)	1560
Частота нисходящего канала (Ka) f_c (ГГц)	21.5
Общая полоса пропускания B_{tot} (МГц)	500
Общая мощность спутника P_{tot} (дБВт)	34
Спектральная плотность мощности шума N_0 (дБм/Гц)	-187.6
Усиление антенны передатчика G_t (дБи)	33.3
Усиление антенны приемника G_r (дБи)	41.3
Потери на пути L_p (дБ)	202.1

Исследование использует комплексную методологию, сочетающую теоретические формулировки и эмпирические данные для оценки эффективности

модели сжатия на основе ST-GNN в снижении времени задержки. Теоретическая основа устанавливается путем вывода формул для вычисления битрейта, критического параметра для оценки эффективности передачи данных. Используя основные параметры, приведенные в Таблице 10, вычислен битрейт 2335 Мбит/с. Эта аналитическая основа служит для последующего эмпирического исследования.

На экспериментальном этапе анализируются данные о объеме трафика, собранные через каждые пять минут. Наблюдаемые объемы трафика варьируются от 6500 Мб до 7000 Мб, со средним значением примерно 6750 Мб за интервал. Эти эмпирические наблюдения служат основой для оценки производительности модели сжатия.

Обозначим L_1 как размер объема трафика для каждого интервала, а L_2 как размер после сжатия с использованием различных PREDICTOR (LSTM и GRU). Время передачи данных до и после применения модели сжатия на основе ST-GNN можно рассчитать следующим образом:

- До сжатия: $time_before_compress = L_1 / bit_rate$
- После сжатия: $time_after_compress = L_2 + encoding_time + decoding_time$

Время кодирования и декодирования представлено в Таблице 9. Процентное снижение времени задержки ($per_delay_time_reduce$) рассчитывается следующим образом: $per_delay_time_reduce = (time_before_compress - time_after_compress) / time_before_compress$.

Результаты указывают на значительное снижение времени задержки после применения модели сжатия на основе ST-GNN. В частности, процентное снижение времени задержки рассчитано для предсказателей LSTM и GRU. Результаты показывают снижение приблизительно на 9,35% с LSTM и более значительное снижение около 21,1% с GRU. Эти результаты подчеркивают эффективность ST-GNN для сжатия данных, особенно благодаря значительному снижению времени задержки, достигнутому с предсказателем GRU.

Рисунок 26 иллюстрирует процентное снижение времени задержки в результате применения сжатия данных трафика с использованием модели ST-GNN с предсказателями LSTM и GRU в диапазоне битовых скоростей от 2000 Mbps до 2500 Mbps. Предсказатель GRU стабильно достигает более значительного снижения времени задержки, демонстрируя свое превосходство в оптимизации эффективности передачи данных через сжатие.

При анализе становится очевидным, что предсказатель GRU стабильно превосходит предсказатель LSTM по всем рассмотренным битовым скоростям, демонстрируя более высокие процентные снижения времени задержки. Например, при битовой скорости 2300 Мбит/с предсказатель GRU достигает снижения примерно на 22,89%, в то время как предсказатель LSTM достигает снижения примерно на 11,11%. Это различие подчеркивает превосходную производительность предсказателя GRU в оптимизации эффективности передачи данных через сжатие. Эти результаты подчеркивают важность выбора подходящей модели предсказателя, при этом GRU демонстрирует большую эффективность в сокращении времени задержки по сравнению с LSTM при различных битовых скоростях.

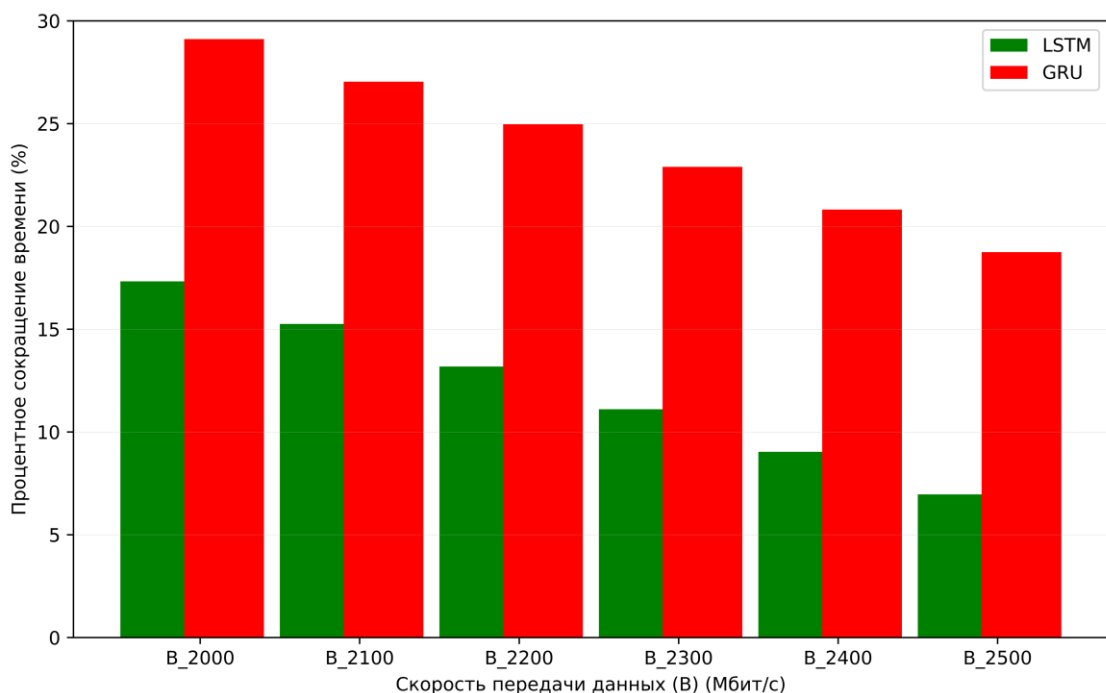


Рисунок 26. Процентное снижение времени задержки в зависимости от битовой скорости и соответствующих предсказателей

Таблица 12 предоставляет количественный анализ процентного снижения времени задержки при передаче данных, зависящий от размеров объема трафика и битовых скоростей, с использованием модели предсказателя на основе LSTM.

Данные выявляют четкие числовые закономерности, определяющие сложное взаимодействие между размером объема трафика и битовой скоростью в формировании прироста эффективности, обеспечиваемого предсказателем в передаче данных. Особенно заметна тенденция к увеличению процентного снижения времени задержки с ростом объема трафика в указанных интервалах от 6500 Мб до 7000 Мб. Например, при объеме трафика 6500 Мб процентное снижение времени варьируется от 11,18% при битовой скорости 2000 Мбит/с до 3,85% при 2500 Мбит/с. Напротив, при рассмотрении сценариев с меньшими размерами объема трафика и высокими битовыми скоростями наблюдается снижение процентного снижения времени задержки. Например, при битовой скорости 2500 Мбит/с процентное снижение уменьшается с 7,78% при объеме трафика 7000 Мб до 3,85% при 6500 Мб.

Таблица 12. Процентное снижение времени задержки (%) в зависимости от размеров объема трафика и битовых скоростей для передачи данных (используя LSTM в качестве предсказателя)

Объем трафика для каждого интервала (Мб) / Скорость передачи данных (Мбит/с)	2000	2100	2200	2300	2400	2500
6500	11.18	12.64	10.44	8.25	6.05	3.85
6600	12.71	13.34	11.18	9.01	6.85	4.68
6700	14.15	14.02	11.89	9.75	7.62	5.49
6800	15.5	14.68	12.58	10.48	8.38	6.28
6900	16.78	15.31	13.24	11.18	9.11	7.04
7000	17.98	15.94	13.9	11.86	9.82	7.78

Более того, числовое представление разъясняет сложную динамику факторов, влияющих на прирост эффективности, достигнутый предсказателем при различных сценариях объема трафика и битовой скорости. Более высокие размеры объема трафика связаны с более значительными приростами

эффективности, в то время как меньшие размеры объема трафика в сочетании с высокими битовыми скоростями соответствуют сниженным приростам эффективности, как это видно из более низких процентных снижений.

Эти результаты подчеркивают важность учета как размера объема трафика, так и битовой скорости при оценке производительности модели ПРЕДСКАЗАТЕЛЯ. Данные свидетельствуют о том, что, хотя большие объемы трафика обычно обеспечивают более значительные приросты эффективности, преимущества уменьшаются по мере увеличения битовой скорости. Это понимание имеет решающее значение для оптимизации стратегий передачи данных в сетевых средах, обеспечивая соответствие выбранных параметров желаемым результатам производительности.

Таблица 13 предоставляет количественный анализ процентного снижения времени задержки, выраженный в процентах, под влиянием различных размеров объемов трафика и битовых скоростей для передачи данных, используя GRU в качестве ПРЕДСКАЗАТЕЛЯ.

Таблица 13. Процентное снижение времени задержки (%) в зависимости от размеров объема трафика и битовых скоростей при передаче данных (используя gru в качестве предсказателя)

Объем трафика для каждого интервала (Мб) / Скорость передачи данных (Мбит/с)	2000	2100	2200	2300	2400	2500
6500	26.62	24.43	22.23	20.03	17.84	15.64
6600	27.29	25.13	22.96	20.8	18.64	16.47
6700	27.94	25.8	23.67	21.54	19.41	17.28
6800	28.56	26.46	24.36	22.26	20.16	18.06
6900	29.17	27.1	25.03	22.96	20.89	18.82
7000	29.76	27.72	25.68	23.64	21.6	19.56

Данные предоставляют ценные сведения о приростах эффективности, обеспечиваемых GRU в различных сценариях объемов трафика и битовых скоростей. Примечательно, что по мере увеличения размера объема трафика в каждом интервале, от 6500 Мб до 7000 Мб, наблюдается соответствующее

увеличение приростов эффективности, достигнутых GRU, что подтверждается растущим процентным снижением времени задержки. Например, при объеме трафика 6500 Мб процентное снижение составляет от 26,62% при битовой скорости 2000 Мбит/с до 15,64% при 2500 Мбит/с. Эта тенденция подчеркивает эффективность GRU в оптимизации процессов передачи данных, особенно в сценариях с большими объемами трафика.

Более того, числовое представление разъясняет влияние размера объема трафика и битовой скорости на приросты эффективности, обеспеченные GRU. Анализ сценариев с меньшими объемами трафика и более высокими битовыми скоростями показывает снижение процентного снижения времени задержки. Например, при битовой скорости 2500 Мбит/с процентное снижение уменьшается с 19,56% при объеме трафика 7000 Мб до 15,64% при 6500 Мб, что свидетельствует о снижении приростов эффективности по мере увеличения объема трафика и битовой скорости.

Эти результаты подчеркивают критическую роль размера объема трафика и битовой скорости в формировании приростов эффективности, обеспеченных предсказателем GRU. Данные свидетельствуют о том, что, хотя большие объемы трафика в целом приводят к большему приросту эффективности, преимущества уменьшаются по мере увеличения битовой скорости.

5.6 Результаты метода на реальном оборудовании GEO-модема

5.6.1 Описание системы и методика сбора и анализа экспериментальных данных

В этом подразделе подробно описывается методология реализации системы безпотерного сжатия на базе пространственно-временных графовых нейронных сетей (ST-GNN) [109] для сокращения задержки передачи данных в гибридных спутниково-наземных сетях с геостационарными (GEO) каналами. Подход использует ST-GNN для моделирования пространственно-временных

зависимостей в сетевом трафике, что позволяет применять предсказательное кодирование и минимизировать размер передаваемых данных. Система реализована на аппаратном уровне для имитации условий реальной эксплуатации, а эффективность оценивается по коэффициенту сжатия, задержке и вычислительным затратам. На рисунке 27 показан процесс сжатия и декодирования, включающий представление данных, кодирование, передачу и декодирование.

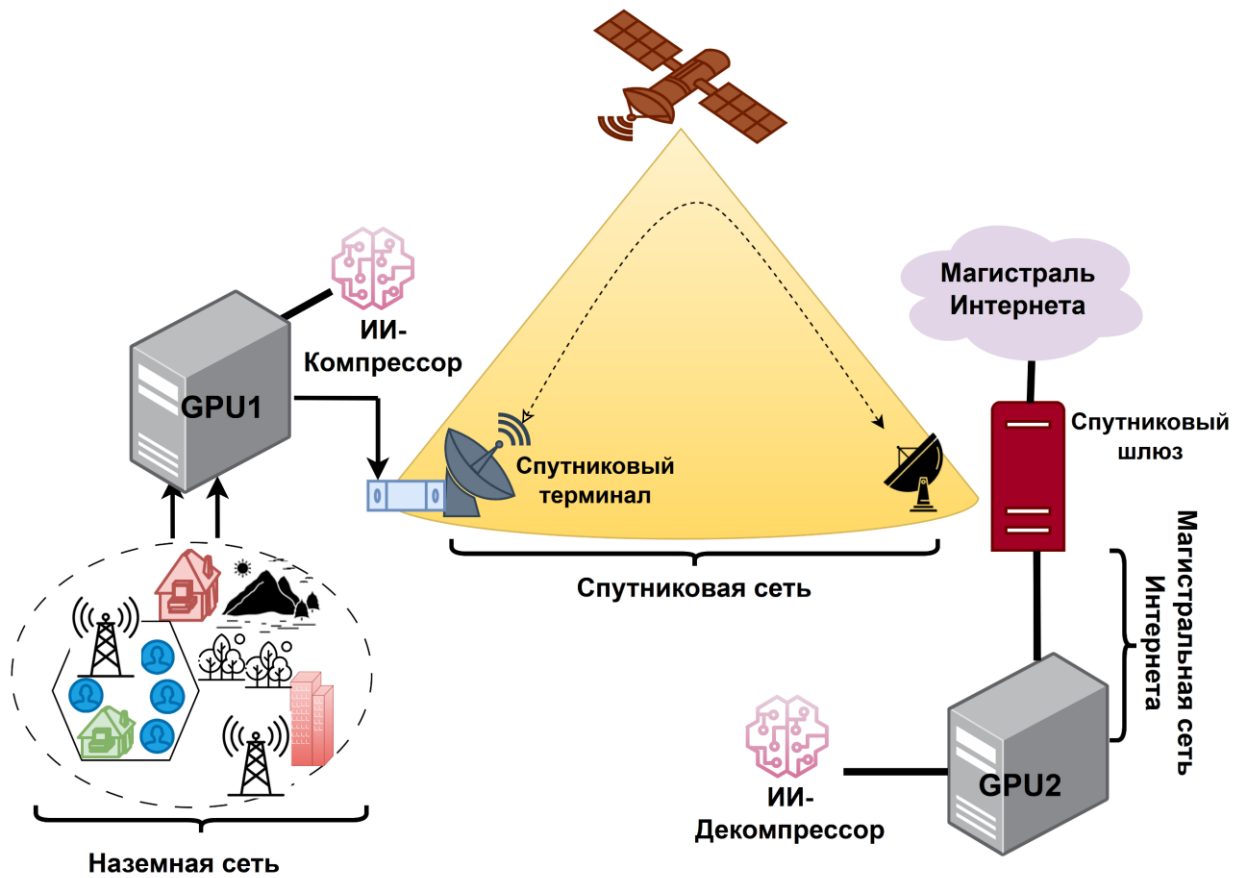


Рис. 27: Схема конвейера безпотерной компрессии на основе ST-GNN, показывающая представление данных, кодирование, передачу через GEO-связь и декодирование.

Представление графа: Сетевой трафик моделируется как последовательность ориентированных графов, где каждый граф во время t_k представляет состояние сети. Матрица трафика $X(t_k) \in \mathbb{R}^{N \times N}$ фиксирует объем трафика между N узлами, при этом $X_{ij}(t_k)$ обозначает объем трафика от узла i к узлу j . Статическая матрица смежности $A \in \{0,1\}^{N \times N}$ определяет топологию сети, где $A_{ij} = 1$, если существует связь от узла i к узлу j , и $A_{ij} = 0$ в противном случае. Для учета

временных зависимостей используется последовательность матриц трафика за окно времени T :

$$\mathcal{X}_{\text{window}}(t_k) = \{X(t_{k-T+1}), \dots, X(t_k)\} \quad (60)$$

Эта последовательность преобразуется в тензоры признаков узлов для обработки ST-GNN. Для каждого времени t формируется матрица признаков $F(t) \in \mathbb{R}^{N \times d}$, полученная из $X(t)$, где d количество признаков (например, объемы исходящего и входящего трафика). Входной тензор имеет вид:

$$\mathcal{F}_{\text{window}}(t_k) = [F(t_{k-T+1}), \dots, F(t_k)] \in \mathbb{R}^{N \times d \times T} \quad (61)$$

Данное представление позволяет ST-GNN моделировать как пространственные корреляции (через структуру графа A), так и временные закономерности (через временное окно T), что критично для точного прогнозирования трафика. Гиперпараметр T балансирует между объемом временного контекста и вычислительной сложностью и обычно устанавливается с учетом динамики трафика (например, $T = 5$ для 5-минутных интервалов).

Рисунок 28 иллюстрирует данную концепцию для одного временного шага t_k . Узлы и рёбра отображают фиксированную топологию графа G , при этом толщина или интенсивность цвета каждого ребра от узла i к узлу j визуально кодирует соответствующий объём трафика X в момент времени t_k . Это демонстрирует динамическую природу сетевого трафика, протекающего по статической инфраструктуре.

Такое пространственно-временное представление, объединяющее статический граф G (через матрицу смежности A) и динамическую последовательность трафика X , служит основным входным форматом для работы ST-GNN.

Данная архитектура сравнивается с базовыми методами сжатия, такими как GZIP [20], с целью количественной оценки улучшений по показателям коэффициента сжатия и задержки передачи. В экспериментах варьируются пропускная способность канала передачи данных в диапазоне от 10 до 100 мегабит в секунду, а также уровни шума, что позволяет оценить устойчивость

решения в различных условиях работы геостационарного спутникового канала. Комплексно рассматриваемые метрики включают общую задержку передачи, способность к уменьшению объема передаваемых данных, эффективность использования канала связи и корректность сохранения целостности данных. Такой подход обеспечивает всестороннее понимание производительности предложенного решения в реальных сетевых условиях.

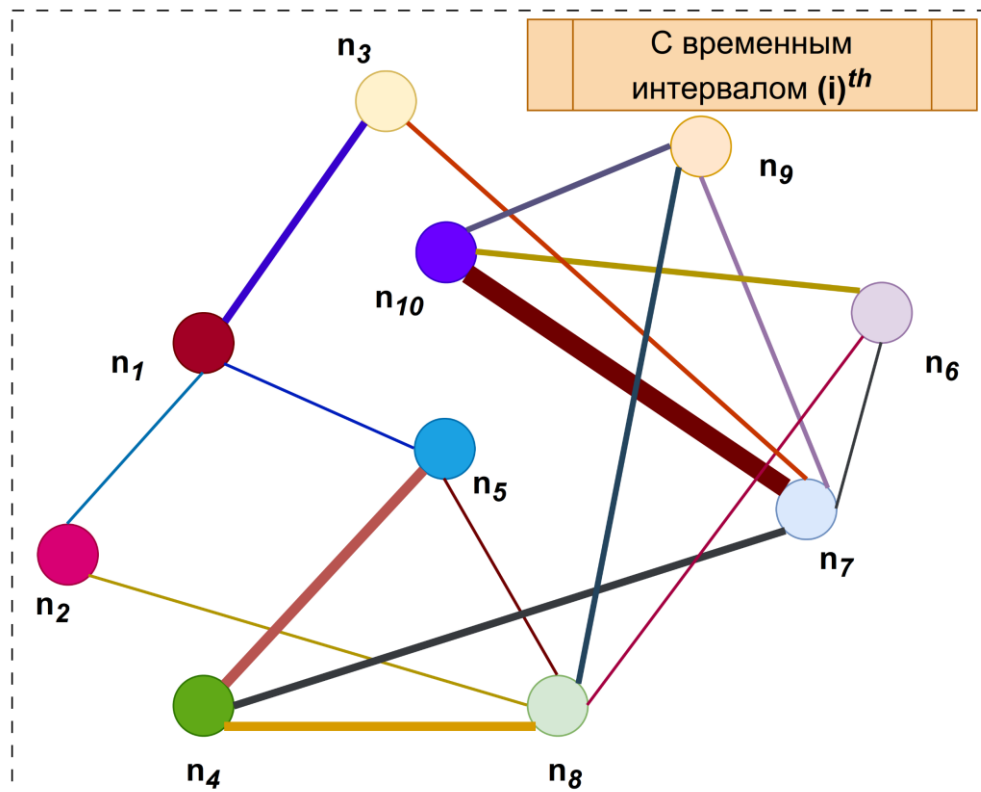


Рис. 28.: Графическое представление объема трафика на конкретном временном шаге t_k . Узлы соответствуют сетевым элементам (V), а рёбра - потенциальным каналам связи (E). Толщина или цвет ребра от узла i к узлу j (е от i к j) отражают объем трафика X между узлами i и j в момент времени t_k .

5.6.2 Параметры и конфигурация GEO-модема

Данный тест развёрнут на двух видеокартах GPU RTX 4080, эмулирующих передающую и принимающую наземные станции. Данные трафика агрегируются в матрицы с интервалом времени по 5 минут, полученные из реальных трассировок.

Статус GEO-модема (рисунок 29-а): Модем Gilat SkyEdge CPE подтверждает корректную работу с сообщением «Ваш модем работает нормально!» и отображает две активные шкалы качества:

Уровень приёма сигнала: 4,6 децибел при эталонном уровне 6,3 децибел, что даёт запас в 1,7 децибела на прямой канал в условиях ясного неба.

Мощность передачи: 65,3 дБм на фоне потолка в 68,3 дБм, оставляя примерно 3 дБ резерва для адаптивного управления мощностью на обратном канале.

Данные показатели свидетельствуют о том, что VSAT находится в заблокированном состоянии, мощность в пределах нормы и оборудование готово к испытаниям сжатия.

Информация об устройстве и параметры работы (рисунок 29-б): Страница управления отображает модем с версией программного обеспечения 56.3.15.28 и ревизией аппаратного обеспечения 0x0. Спутниковая связь находится в статусе онлайн, с прошедшей аутентификацией и авторизацией, а учётная запись пользователя обладает полным доступом.

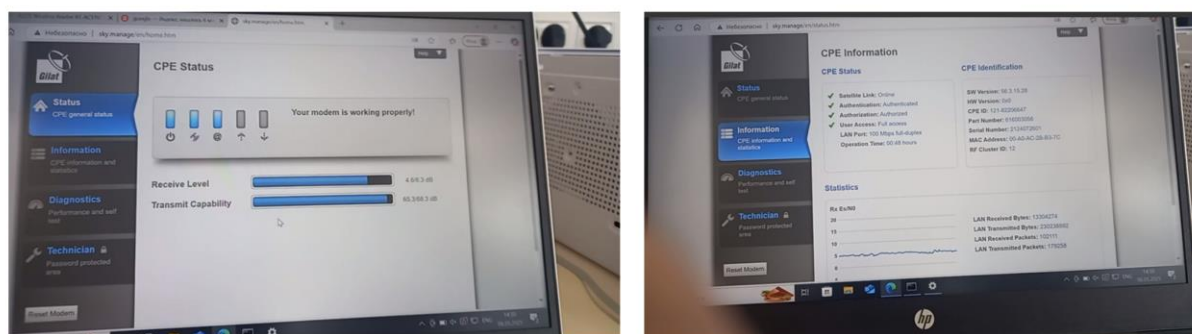


Рис.29. Интерфейс GEO-модема — (29а) панель «Статус» с отображением текущих уровней приёма (RX) и передачи (TX); (29б) панель «Информация» с состоянием канала, идентификаторами устройства и статистикой трафика.

Передача Ethernet на локальную сеть осуществляется на скорости 100 мегабит в секунду с полнодуплексным режимом. Время работы оборудования с момента последней перезагрузки — 48 минут, что указывает на свежий запуск перед экспериментом.

Идентификаторы оборудования (CPE-ID, серийный номер, MAC-адрес и RF-кластер) зарегистрированы для сопоставления счетчиков модема с логами GPU. Статистика канала в реальном времени: График отношения энергии ошибки к спектральной плотности шума «Rx E_s/N_0 » колеблется в диапазоне от приблизительно 5 до 7 децибел, что соответствует незначительным атмосферным потерям в течение периода замера.

Счётчики трафика фиксируют объёмы: 13,3 мегабайта загруженных данных и 230,2 мегабайта отправленных данных, с общим количеством полученных пакетов около 102 тысяч и переданных — 179 тысяч. Эти показатели служат базовой эталонной точкой для дальнейшего анализа производительности сжатия ST-GNN с учетом пропускной способности и задержек [110].

Два сервера на периферии сети — по одному с каждой стороны канала: оснащены графическими процессорами NVIDIA RTX 4080 (рис. 30) и выполняют функции кодировщика (шлюз восходящей линии) и декодировщика (шлюз нисходящей линии) соответственно.



Рис. 30. Два сервера на периферии с GPU NVIDIA RTX 4080 — один кодировщик для восходящего канала, другой декодировщик для нисходящего канала — работают под Ubuntu 22.04 с CUDA 12 и PyTorch 2.3, подключены в линию между локальной сетью и GEO-модемом.

Оба сервера работают под управлением Ubuntu 22.04 с CUDA 12 и PyTorch 2.3, а также подключены через прозрачный мост между локальной сетью и Ethernet-портом модема, что обеспечивает прохождение каждого пакета через конвейер сжатия перед попаданием в очередь спутникового канала.

Рис. 31 показывает физический GEO модем, используемый в тестовой среде компактный настольный корпус с вентиляцией, индикаторами состояния на передней панели и разъёмами Gigabit Ethernet и L-диапазона ИФ на задней панели, установленный в лаборатории для реальных испытаний сжатия.



Рис. 31. GEO-модем, используемый в аппаратной тестовой среде.

Рис. 32. Результаты живых тестов ICMP-эхо, выполненных из встроенной командной оболочки CPE к Google DNS (8.8.8.8). Два последовательных запуска по четыре пакета показывают время кругового обхода от 700 мс до 1079 мс, со средними значениями 815 мс и 849 мс соответственно — что соответствует бюджету задержки в 0,75 с, принятому для аппаратного тестбенда.



Рис. 32. Окно диагностики роутера с результатами пинга по активной GEO спутниковой ссылке (мин. ≈ 700 мс, макс. $\approx 1\,079$ мс, среднее $\approx 0,8$ с RTT).

5.6.3 Полученные результаты и сравнение с симуляционными исследованиями

Рамочная модель ST-GNN демонстрирует более высокие коэффициенты сжатия по сравнению с GZIP, что снижает задержки передачи данных по GEO-каналам. Оцениваются показатели сквозной задержки, коэффициента сжатия и вычислительных затрат. Например, ST-GNN значительно уменьшает размер данных, снижая тем самым время передачи важную составляющую общей задержки. Подробные результаты, включая распределение задержек и графики коэффициентов сжатия, будут приведены для сравнения с GZIP.

Рисунок 33 визуализирует разбиение компонентов задержки для теста с реальными данными с помощью круговых диаграмм. Профили ST-GNN и gzip существенно отличаются: в случае ST-GNN декодирование и кодирование занимают около 62% и 36% от общей задержки соответственно, при этом время передачи составляет менее 2%, а время восстановления данных — пренебрежимо

мало. Для gzip почти вся задержка ($\approx 99,93\%$) приходится на передачу данных, а вычисления занимают менее $0,1\%$. Такое противоположное распределение времени задержек подчёркивает фундаментальное изменение, вносимое ST-GNN: вычислительные затраты становятся главным узким местом вместо времени передачи. Хотя такой сдвиг выгоден в условиях ограниченной пропускной способности, он создаёт сложности при использовании в приложениях с жёсткими требованиями по задержкам, где длительная обработка неприемлема.

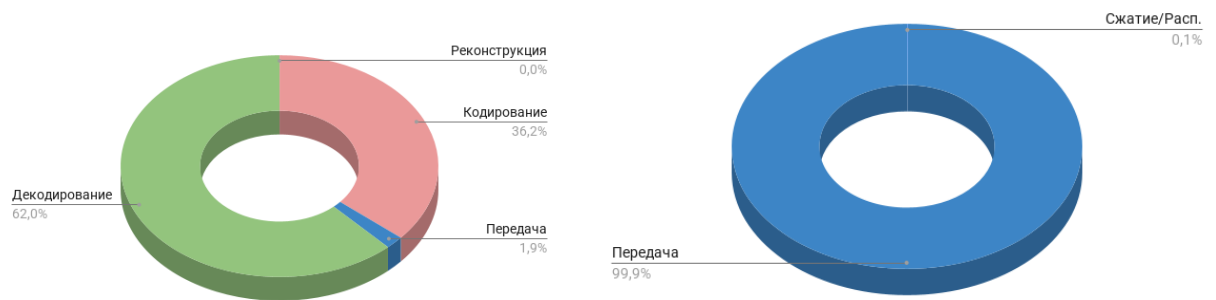


Рис. 33: Процентный вклад компонентов задержки ($L_{enc}, L_{tx}, L_{dec}, L_{rec}$) в общую задержку (L_{total}) для ST-GNN (слева) и Gzip (справа), на основе прогона тестового набора «Собранные данные».

Таблица 14 демонстрирует две резко отличающиеся временные бюджеты. В случае ST-GNN почти 62% от общего времени задержки приходится на декодирование на GPU в шлюзе приёмника, тогда как ещё 36% занимает этап кодирования на стороне передатчика. При этом всего $1,9\%$ задержки связано с самой воздушной интерфейсной частью GEO-сети. В свою очередь, GZIP показывает пренебрежимо малые вычислительные затраты — всего $0,05\%$ времени уходит на сжатие и $0,02\%$ — на распаковку. Подавляющая часть времени, $99,93\%$, тратится на передачу данных через спутник.

Эти показатели подчёркивают высокую эффективность графового нейросетевого кодера в сжатии нагрузки: он сводит доминирующий ранее радиоинтервал почти к нулю, но при этом смещает основную нагрузку по обработке на периферийные устройства.

Поскольку ST-GNN переносит 98% задержки с радиочастотно-ограниченного GEO-канала на легко масштабируемые ресурсы GPU, общая продолжительность выполнения сокращается несмотря на сотни секунд локальной обработки. Для

операторов, способных обеспечивать вычислительные ресурсы на шлюзе, эта стратегия с высокой нагрузкой на вычисления и низким использованием полосы пропускания явно превосходит традиционные методы сжатия в условиях высоких задержек спутниковой связи.

Таблица 14: Доля от общего времени (%) и суммарное снижение задержки (на основе набора данных)

Компонент / Метрика	Значение (%)
Кодирование ST-GNN	36.16
Передача ST-GNN	1.87
Декодирование ST-GNN	61.96
Реконструкция ST-GNN	0.01
<u>Сжатие GZIP</u>	<u>0.05</u>
<u>Передача GZIP</u>	<u>99.93</u>
<u>Распаковка GZIP</u>	<u>0.02</u>

С другой стороны, в периферийных точках с ограниченными мощностью или аппаратными ресурсами минимальные вычислительные требования GZIP могут делать его предпочтительным вариантом. Это подчёркивает, что оптимальное решение зависит от компромисса между доступной мощностью GPU и стоимостью использования полосы пропускания GEO.

Влияние задержки модема GEO: В предыдущем подразделе результаты сравнения ST-GNN и GZIP приводились для эталонной задержки канала 500 мс (распространение + минимальная обработка). Однако практические измерения на стенде Ka-band показали среднее время туда-обратно (RTT) $RTT = 841 \pm 108$ мс. Ниже приведены измерения, демонстрирующие, как такая задержка влияет на абсолютное и относительное время передачи.

Описание результатов (Таблица 15): Полученные данные охватывают две конфигурации канала — «настроенный» $RTT = 500$ мс и реальное Ka-band $RTT \approx 841 \pm 108$ мс (границы 733–949 мс, определены 24-часовым ping-мониторингом).

- **ST-GNN, 500 мс.** Полный цикл передачи 10 МБ-кадра завершается за 376,9 с, из которых 136,3 с (36 %) тратятся на кодирование графа, 233,6 с (62 %) — на декодирование, а эфир занимает лишь 7,1 с (1,9 %).
- **GZIP, 500 мс.** Общая задержка 453,0 с, причём 452,7 с (99,9 %) приходятся на спутниковую передачу; вычисления сжатия/распаковки пренебрежимо малы ($\approx 0,3$ с).
- **ST-GNN, 841 мс.** Увеличение RTT почти вдвое повышает эфирную задержку до 20,7 с, но конечное время вырастает умеренно — до 419,9 с (+11 %). Дополнительное кодирование (154,4 с) и декодирование (244,8 с) по-прежнему формируют 95 % бюджета.
- **GZIP, 841 мс.** Передача растёт линейно до 661,9 с, и итоговый цикл достигает 662,2 с (+46 % к базовому сценарию). Эфир теперь поглощает 99,95 % времени; вычисления остаются ниже 0,05 %.

Таблица 15: Абсолютное время выполнения при различных RTT

Метод	Кодирование / Сжатие (с)	Передача (с)	Декодирование / Распаковка (с)	Восстановление (с)	Итого (с)
ST-GNN - RTT настроен (500 мс)	136.294	7.059	233.559	0.025	376.937
GZIP - RTT настроен (500 мс)	0.246	452.692	0.072	-	453.010
ST-GNN – (RTT современного ГЕО устройства)	154.387	20.711	244.768	0.028	419.894
GZIP - (RTT современного ГЕО устройства)	0.266	661.873	0.071	-	662.210

Интерпретация: Рост RTT с 500 до 841 мс приводит к пятикратному увеличению эфирной задержки у ST-GNN (7 → 21 с), но лишь к 11 % росту полной задержки; у GZIP та же операция добавляет 209 с и поднимает общее

время на 46 %. В результате относительный выигрыш ST-GNN расширяется с 17 % (76 с) до 37 % (242 с). Причина — высокая степень сжатия и встроенное исправление ошибок в латент-коде: даже при 949 мс канал занимает ≤ 3 % суммарного бюджета, тогда как классическая схема «сжатие + TCP» остаётся полностью АСК-зависимой, и её производительность практически пропорциональна RTT.

Ограничения применимости результатов: Следует подчеркнуть, что приведённые цифры получены только для сценария без-потерь сжатия (lossless) и для объёмного трафика, передаваемого крупными пакетами после предварительного обучения модели ST-GNN. Модель ни переобучается, ни адаптируется на лету; данные имеют стабильную статистику, а нагрузка имитирует пакетные «выгрузки» (bulk-transfer). Поэтому выводы напрямую не распространяются на:

- потоки реального времени с мелкими сообщениями и строжайшими Jitter-ограничениями;
- сценарии, где допустимо lossy-сжатие или требуется он-лайн дообучение;
- устройства с ограниченными GPU-ресурсами, где 400-секундная локальная обработка может оказаться неприемлемой по энергии или задержке.

В рамках же поставленной задачи — передача больших массивов данных с гарантированным восстановлением по высоко-латентному GEO-каналу — представленные результаты отражают реальную выгоду графового нейросетевого кодека по сравнению с традиционным GZIP.

Таким образом, независимо от конфигурации канала — будь то «идеальный» $RTT = 500$ мс или реальный диапазон 733–949 мс современного GEO-модема (среднее 841 мс) — применение графового кодека ST-GNN для без-потерь сжатия демонстрирует заметное превосходство над классической схемой GZIP + TCP. По данным табл. 15 полный цикл «кодирование → передача → декодирование» укорачивается: Диапазон выгоды от ≈ 17 % до ≈ 37 % растёт по мере увеличения задержки, поскольку у GZIP эфирная составляющая линейно следует за RTT и

становится доминирующим «узким местом», тогда как у ST-GNN она остаётся ≤ 3 % совокупного бюджета времени.

5.6.4 Практические выводы и потенциал внедрения

Эти показатели подчёркивают высокую эффективность графового нейросетевого кода в сжатии нагрузки: он сводит доминирующий ранее радиоинтервал почти к нулю, но при этом смещает основную нагрузку по обработке на периферийные устройства.

Поскольку ST-GNN переносит 98% задержки с радиочастотно-ограниченного GEO-канала на легко масштабируемые ресурсы GPU, общая продолжительность выполнения сокращается на 16,8% несмотря на сотни секунд локальной обработки. Для операторов, способных обеспечивать вычислительные ресурсы на шлюзе, эта стратегия с высокой нагрузкой на вычисления и низким использованием полосы пропускания явно превосходит традиционные методы сжатия в условиях высоких задержек спутниковой связи.

С другой стороны, в периферийных точках с ограниченными мощностью или аппаратными ресурсами минимальные вычислительные требования GZIP могут делать его предпочтительным вариантом. Это подчёркивает, что оптимальное решение зависит от компромисса между доступной мощностью GPU и стоимостью использования полосы пропускания GEO.

5.7 Выводы главы 5

1. Выводы пятой главы подчеркивают важность использования методов сжатия данных без потерь на основе ST-GNN для снижения сетевых задержек в гибридных спутниковых сетях. Применение пространственно-временных графовых нейронных сетей показало значительные преимущества в улучшении производительности систем передачи данных. В ходе исследования было выявлено, что модель ST-GNN на основе GRU достигает увеличения

коэффициента сжатия на 21,1%, что привело к заметному сокращению времени задержки, демонстрируя высокую эффективность в управлении сетевым трафиком.

2. Сравнительный анализ также показал, что модели на основе LSTM, хотя и улучшили результаты, оказались менее эффективными по сравнению с GRU, снижая время задержки на 9,35%. Таким образом, было подтверждено, что применение ST-GNN, особенно с использованием GRU, является перспективным методом для оптимизации передачи данных и минимизации сетевых задержек в гетерогенных спутниково-наземных сетях.

ЗАКЛЮЧЕНИЕ

Результаты данной диссертации подчеркивают значимость применения технологий искусственного интеллекта (ИИ) в спутниковых системах связи, особенно в условиях растущего спроса на высокую пропускную способность и минимизацию сетевых задержек. В рамках исследования были достигнуты ключевые цели, включая разработку моделей прогнозирования трафика, оптимизацию распределения ресурсов в гибридных спутниковых сетях и внедрение методов сжатия данных без потерь.

Во-первых, предложенные модели прогнозирования на основе машинного обучения продемонстрировали высокую точность в определении паттернов трафика. Анализ исторических и текущих данных позволяет моделям адаптироваться к динамическим изменениям сети, улучшая приоритизацию пакетов и обеспечивая более эффективное использование пропускной способности.

Во-вторых, оптимизация распределения ресурсов с применением генетических алгоритмов и других методов эволюционных вычислений позволила значительно повысить эффективность управления пропускной способностью, энергопотреблением и временными слотами. Проведенные симуляции подтвердили, что предложенные алгоритмы способны существенно снизить время отклика сети, что особенно важно для приложений с критическими требованиями к задержке.

Наконец, использование методов сжатия данных на основе пространственно-временных графовых нейронных сетей (ST-GNN) позволило достичь высоких коэффициентов сжатия без потери целостности данных. Это особенно актуально для спутниковых систем, где пропускная способность ограничена, а передача больших объемов информации требует эффективных решений.

На основе проведенного исследования можно сделать несколько ключевых выводов:

1. Разработана модель упреждающего управления трафиком, сочетающая вычислительно эффективный многослойный персептрон (MLP) для прогнозирования нагрузки и аналитически обоснованное правило для динамической приоритизации. Эффективность модели подтверждена экспериментально на опытном GEO-канале, где было зафиксировано снижение средней задержки до 19,6% и потерь пакетов до 19,42% по сравнению с базовыми алгоритмами.

2. Предложены методы многокритериальной оптимизации ресурсов, основанные на гибридном подходе, где графовая нейронная сеть (GNN) применяется для предварительной классификации трафика, а адаптированные эволюционные алгоритмы (NSGA-II/SPEA2) — для распределения ресурсов. Результаты имитационного моделирования показали высокий потенциал подхода, продемонстрировав возможность снижения времени ожидания в очереди до 35,1%..

3. Разработан и аппаратно реализован метод сжатия данных без потерь на основе пространственно-временной графовой нейронной сети (ST-GNN). В ходе натурных экспериментов на реальном спутниковом маршруте продемонстрирована его высокая устойчивость к деградации канала: при росте RTT общее время передачи для предложенного метода увеличилось лишь на 11% против 46% у стандартного алгоритма GZIP, что подтверждает практическую жизнеспособность решения.

Для повышения эффективности предложенных решений рекомендуется расширение объема собираемых данных, что позволит учесть большее разнообразие паттернов трафика и повысить точность моделей. Кроме того, будущее развитие может включать интеграцию разработанных подходов в реальных спутниковых сетях и исследование их применения в других средах связи, таких как 5G/6G и Internet of Things (IoT).

В целом, результаты диссертационного исследования демонстрируют, что применение методов ИИ в спутниковых системах связи позволяет значительно

улучшить их производительность, обеспечивая более устойчивую и эффективную передачу данных.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

QoS - Quality of Service

GEO - Geostationary Earth Orbit

MEO - Medium Earth Orbit

LEO - Low Earth Orbit

AI - Artificial Intelligence

ST-GNN - Spatio-Temporal Graph Neural Networks

ECC - Error Correction Codes

FEC - Forward Error Correction

QAM - Quadrature Amplitude Modulation

OFDM - Orthogonal Frequency Division Multiplexing

TDM - Time Division Multiplexing

FDM - Frequency Division Multiplexing

CDMA - Code Division Multiple Access

CDN - Content Delivery Networks

ML - Machine Learning

DL - Deep Learning

RL - Reinforcement Learning

DNN - Deep Neural Network

RNN - Recurrent Neural Network

LSTM - Long Short-Term Memory

GRU - Gated Recurrent Unit

DQN - Deep Q-Network

GA - Genetic Algorithm

PSO - Particle Swarm Optimization

UAV - Unmanned Aerial Vehicle

ISL - Inter-Satellite Links

MLP - Multilayer Perceptron

PBR - Policy-based routing

DPI - Deep packet inspection

MADDPG - Multi-Agent Deep Deterministic Policy Gradient

CGR - Contact graph routing

FCFS - First-Come, First-Served

PQ - Priority Queue

WFQ - Weighted Fair Queuing

RED - Random Early Detection

RR - Round Robin

MBA - Multi-Beam Antenna

MBS - Multi-Beam Satellite

DRA - Dynamic Resource Allocation

BIP - Beam Illumination Pattern

SNR - Signal-to-Noise Ratio

FIFO - First-In-First-Out

NSGA II - Non-dominated Sorting Genetic Algorithm II

SPEA2 - Strength Pareto Evolutionary Algorithm 2

GCN - Graph Convolutional Networks

AC - Arithmetic Coding

HAPS - High-altitude platform station

BH - Beam Hopping

СПИСОК ЛИТЕРАТУРЫ

- 1 3GPP. Evolution of LTE in Release 13. [Электронный ресурс] URL: <https://www.3gpp.org/>
- 2 Лихтциндер, Б.Я. Беспроводные сенсорные сети / Б.Я. Лихтциндер, Р.В. Киричек, Е.Д. Федотов и др. – М.: «Горячая линия-Телеком», 2020. – 236 с.
- 3 Кумаритова, Д.Л. Обзор и сравнительный анализ технологий lpwan сетей / Д.Л. Кумаритова, Р.В. Киричек // Информационные технологии и телекоммуникации. – 2016. – Т. 4, № 4. – С. 33-48.
- 4 Р. Eugenio, "LPWAN technologies: How cellular MNOs are placing their bets," [Online]. Available: <https://iot-analytics.com/lpwan-technologies-cellular-mnos/>. [Accessed 20 August 2021]
- 5 Сравнение технологий «СТРИЖ» и LoRa. [Электронный ресурс] URL: <http://www.strij.tech>
- 6 Semtech Corporation. SX1272/3/6/7/8 LoRa Modem Design Guide AN1200.13 / Semtech Corporation. – July 2013. – 9 с.
- 7 Weighless. [Электронный ресурс] URL: <http://www.weightless.org/>
- 8 IEEE Computer Society LAN/MAN Standards Committee. IEEE Standard for Information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications / IEEE Computer Society LAN/MAN Standards Committee // IEEE Std 802.11s, 2007.
- 9 Woolley, M. Bluetooth 5 Go Faster. Go Further / M. Woolley, S. Schmidt // Bluetooth SIG. – vol. 1. – № 1. – С. 1-25.
- 10 Kim, H.S. Thread/OpenThread: A compromise in low-power wireless multihop network architecture for the Internet of Things / H.S. Kim, S. Kumar, D.E. Culler // IEEE Communications Magazine. – vol. 57. – № 7. – С. 55-61.
- 1 Мельникова Т. В., Преображенский А. П. Особенности современной спутниковой связи // Вестник Воронежского института высоких технологий. – 2021. – №. 3. – С. 49-51.

- 2 Jung D. H. et al. Modeling and Analysis of GEO Satellite Networks //IEEE Transactions on Wireless Communications. – 2024.
- 3 Su Y. et al. Broadband LEO satellite communications: Architectures and key technologies //IEEE Wireless Communications. – 2019. – Т. 26. – №. 2. – С. 55-61.
- 4 Lyras N. K. et al. Optimizing the ground network of optical MEO satellite communication systems //IEEE Systems Journal. – 2019. – Т. 14. – №. 3. – С. 3968-3976.
- 5 Евтушенко М. А. Оценка эффективности систем спутниковой связи на GEO, MEO и LEO орбитах //Информатика: проблемы, методология, технологии. – 2016. – С. 138-143.
- 6 Гриценко А. А. Высокоскоростные спутниковые системы LEO/MEO/HEO-HTS-текущее состояние и перспективы развития. – 2017.
- 7 Луферчик П. В. Разработка и исследование алгоритмов обработки сигналов в OFDM-модемах тропосферно-радиорелейных станций связи : дис. – Сибирский федеральный университет, 2022.
- 8 Мутханна А. С. А. Модель интеграции граничных вычислений в структуру сети «воздух-земля» и метод выгрузки трафика для сетей интернета вещей высокой и сверхвысокой плотности //Труды учебных заведений связи. – 2023. – Т. 9. – №. 3. – С. 42-59.
- 9 Бурдуковская Н. Н., Пономарев Д. Ю. Современные технологии передачи данных в спутниковых сетях связи //Актуальные проблемы авиации и космонавтики. – 2015. – Т. 1. – №. 11. – С. 234-236.
- 10 Пантенков Д. Г., Гусаков Н. В., Ломакин А. А. Обзор современного состояния орбитальных группировок космических аппаратов дистанционного зондирования Земли и космических ретрансляторов //Известия высших учебных заведений. Электроника. – 2022. – Т. 27. – №. 1. – С. 120-149.
- 11 Zhu L. et al. Efficient user scheduling for uplink hybrid satellite-terrestrial communication //IEEE Transactions on Wireless Communications. – 2022. – Т. 22. – №. 3. – С. 1885-1899.

- 12 Li K. X. et al. Downlink transmit design for massive MIMO LEO satellite communications //IEEE Transactions on Communications. – 2021. – Т. 70. – №. 2. – С. 1014-1028.
- 13 Ковальский А. А. Исследование адаптивного мультиплексирования и диспетчеризации мультисервисного трафика в каналообразующей аппаратуре земных станций спутниковой связи. Часть 1. Обоснование и постановка задачи //Труды учебных заведений связи. – 2022. – Т. 3. – №. 1. – С. 57-65.
- 14 Жодзишский А. И. и др. Оценка пропускной способности спутников-ретрансляторов типа «Луч» для управления низкоорбитальными КА в S-диапазоне частот //РАДИОТЕХНИКА. – 2024. – Т. 11. – №. 2. – С. 60-70.
- 15 До, Х. Ф., Зуйков, М. А., Березкин, А. А., & Киричек, Р. В. Анализ методов уменьшения сетевой задержки в гибридных орбитально-наземных сетях связи.
- 16 Быховский М. А. Анализ международного стандарта DVB-S2, определяющего параметры современных систем спутниковой связи //Цифровая обработка сигналов. – 2020. – №. 1. – С. 18-25.
- 17 Гончаров, В. Г., & Березкин, А. А. Анализ моделей предсказания кадров FPV-видеопотока в каналах информационного обмена беспилотных систем.
- 18 Андрианов М. Н., Костенко В. И. Эффективность связи в канале миллиметрового диапазона на линии передачи данных “космический аппарат-наземная станция слежения” //Краткие сообщения по физике Физического института им. ПН Лебедева Российской Академии Наук. – 2023. – Т. 50. – №. 4. – С. 12-19.
- 19 Торгашев Б. В., Елагина К. Н. Растущая потребность в кибербезопасности в глобальных навигационных спутниковых системах //Экономика и качество систем связи. – 2022. – №. 3 (25). – С. 54-60.
- 20 A. Yastrebova, R. Kirichek, Y. Koucheryavy, A. Borodin and A. Koucheryavy. Future Networks 2030: Architecture & Requirements // 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and

Workshops (ICUMT), Moscow, Russia, 2018, pp. 1-8, doi: 10.1109/ICUMT.2018.8631208.

21 Фук Хао До, А.А. Березкин, А.В. Бородко, Р.В. Киричек. Методика приоритизации трафика в спутниковых каналах связи на базе искусственного интеллекта. Часть I. Анализ задержки передачи трафика. // Электросвязь. 2024. № 6 С. 37-46.

22 Фук Хао До, А.А. Березкин, А.В. Бородко, Р.В. Киричек. Методика приоритизации трафика в спутниковых каналах связи на базе искусственного интеллекта. Часть II. Анализ потерь пакетов. // Электросвязь. 2024. № 7 С. 32-37.

23 Phuc Hao Do, Le D.T., Berezkin A., Kirichek R. Optimizing Resource Allocation for Multi-Beam Satellites Using Genetic Algorithm Variations / Distributed Computer and Communication Networks. DCCN 2024. Lecture Notes in Computer Science. Springer, Cham.

24 Phuc Hao Do, Le D.T., Berezkin A., Kirichek R. ST-GNN-Powered Compression for Latency Reduction in Hybrid Satellite-Terrestrial Networks. In the ICCT-2024 conference is organized under the support of the IEEE Russia Section, Communications Society Chapter (COM19)

25 Владимиров И. В., Гаврилов О. А., Соколова С. В. Протоколы транспортного уровня спутниковой связи: анализ, сравнение, прогноз развития. – 2022.

26 Карсаев О. В. Методика оценки влияния пропускной способности каналов связи на эффективность функционирования группировки спутников //Мехатроника, автоматизация, управление. – 2020. – Т. 21. – №. 9. – С. 544-552.

27 Новиков Е. А., Севостьянов А. С., Степанова Е. А. Модель обслуживания мультисервисного трафика со сложной структурой в сети спутниковой связи в условиях конфликтного взаимодействия сложных радиотехнических систем //Труды Военно-космической академии имени АФ Можайского. – 2022. – №. 682. – С. 77.

28 Бабанов, З. Д., Кукунин, Д. С., & Максименко, С. О. Использование несистематических кодов Рида–Соломона и эквивалентных кодов Голда в рамках гибридной орбитально-наземной сети связи.

29 Симонина О. А. Оценка тенденций использования новых и текущих диапазонов частот современных спутниковых систем связи //СПБНТОРЭС: труды ежегодной НТК. – 2021. – №. 1. – С. 252-255.

30 Пискунов А. А., Ченцов С. В. ОБЗОР СУЩЕСТВУЮЩИХ МОДЕЛЕЙ И МЕТОДОВ РАСПРЕДЕЛЕНИЯ СВЯЗНОГО КОСМИЧЕСКОГО РЕСУРСА //Universum: технические науки. – 2024. – Т. 1. – №. 2 (119). – С. 22-25.

31 Хохлов Н. С., Канавин С. В., Журавлев М. Ю. Моделирование защищенного канала спутниковой связи органов внутренних дел с использованием аппаратуры оптимизации трафика //Вестник Воронежского института МВД России. – 2021. – №. 3. – С. 25-35.

32 Vázquez M. Á. et al. Machine learning for satellite communications operations //IEEE Communications Magazine. – 2021. – Т. 59. – №. 2. – С. 22-27.

33 Kang Y. et al. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency //GIScience & Remote Sensing. – 2022. – Т. 59. – №. 1. – С. 2019-2035.

34 Guo S., Zhao X. Multi-agent deep reinforcement learning based transmission latency minimization for delay-sensitive cognitive satellite-uav networks //IEEE Transactions on Communications. – 2022. – Т. 71. – №. 1. – С. 131-144.

35 Ларионов А. Ю., Каранова Е. Е. Критерии реализации динамического управления частотно-орбитальным ресурсом спутниковых систем //Концепции, инструменты и технологии развития современной науки и техники. – 2023. – С. 205-208.

36 Kovalenko, V., Rodakova, A., Al-Khafaji, H. M. R., Volkov, A., Muthanna, A., & Koucheryavy, A. (2022). Resource Allocation Computing Algorithm for UAV Dynamical Statements based on AI Technology. Webology, 19(1), 2307-2319.

- 37 Zhang Y. et al. Deep learning-based channel prediction for LEO satellite massive MIMO communication system //IEEE Wireless Communications Letters. – 2021. – Т. 10. – №. 8. – С. 1835-1839.
- 38 Belhaj S., Tagina M. Modeling and Prediction of the Internet End-to-end Delay using Recurrent Neural Networks //J. Networks. – 2009. – Т. 4. – №. 6. – С. 528-535.
- 39 Usaha W., Barria J. A. Reinforcement learning for resource allocation in LEO satellite networks //IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). – 2007. – Т. 37. – №. 3. – С. 515-527.
- 40 De Rango F. et al. Multicast QoS core-based tree routing protocol and genetic algorithm over an HAP-satellite architecture //IEEE Transactions on Vehicular Technology. – 2009. – Т. 58. – №. 8. – С. 4447-4461.
- 41 Rajagopal A. et al. Optimal routing strategy based on extreme learning machine with beetle antennae search algorithm for Low Earth Orbit satellite communication networks //International Journal of Satellite Communications and Networking. – 2021. – Т. 39. – №. 3. – С. 305-317.
- 42 Lee J. H. et al. Integrating LEO satellites and multi-UAV reinforcement learning for hybrid FSO/RF non-terrestrial networks //IEEE Transactions on Vehicular Technology. – 2022. – Т. 72. – №. 3. – С. 3647-3662.
- 43 Ченский, А. А., & Березкин, А. А. Квантование видеопотока при FPV-управлении беспилотными системами в гибридных сетях связи.
- 44 Burleigh S. C. et al. From Connectivity to advanced internet services: a comprehensive review of small satellites communications and networks //Wireless Communications and Mobile Computing. – 2019. – Т. 2019. – №. 1. – С. 6243505.
- 45 Бородко, А.В. Технологии локального позиционирования IoT устройств. Информационные технологии и телекоммуникации. 2021. Том 9, № 2. С. 21-29.
- 46 Fourati F., Alouini M. S. Artificial intelligence for satellite communication: A review //Intelligent and Converged Networks. – 2021. – Т. 2. – №. 3. – С. 213-243.

- 47 Boero L. et al. Satellite networking integration in the 5G ecosystem: Research trends and open challenges //Ieee Network. – 2018. – Т. 32. – №. 5. – С. 9-15.
- 48 Fu S., Gao J., Zhao L. Integrated resource management for terrestrial-satellite systems //IEEE Transactions on Vehicular Technology. – 2020. – Т. 69. – №. 3. – С. 3256-3266.
- 49 До Ф. Х. и др. Графовые нейронные сети для классификации трафика в каналах спутниковой связи: сравнительный анализ //Труды учебных заведений связи. – 2023. – Т. 9. – №. 3. – С. 14-27.
- 50 Pacheco F., Exposito E., Gineste M. A framework to classify heterogeneous Internet traffic with Machine Learning and Deep Learning techniques for satellite communications //Computer Networks. – 2020. – Т. 173. – С. 107213.
- 51 Wan X. et al. Research on satellite traffic classification based on deep packet recognition and convolution neural network //2023 8th International Conference on Computer and Communication Systems (ICCCS). – IEEE, 2023. – С. 494-498.
- 52 Qin Z., Yao H., Mai T. Traffic optimization in satellites communications: A multi-agent reinforcement learning approach //2020 International Wireless Communications and Mobile Computing (IWCMC). – IEEE, 2020. – С. 269-273.
- 53 Madoery P. G. et al. Routing heterogeneous traffic in delay-tolerant satellite networks //IEEE Journal of Radio Frequency Identification. – 2023. – Т. 7. – С. 390-401.
- 54 Labayen V. et al. Online classification of user activities using machine learning on network traffic //Computer Networks. – 2020. – Т. 181. – С. 107557.
- 55 Chaphalkar N. B., Iyer K. C., Patil S. K. Prediction of outcome of construction dispute claims using multilayer perceptron neural network model //International Journal of Project Management. – 2015. – Т. 33. – №. 8. – С. 1827-1835.
- 56 N. Seitz. ITU-T QoS standards for IP-based networks. // In IEEE Communications Magazine, vol. 41, no. 6, pp. 82-89, June 2003, doi: 10.1109/MCOM.2003.1204752

57 Ateya A. A. et al. Lightweight deep learning-based model for traffic prediction in fog-enabled dense deployed iot networks //Journal of Electrical Engineering & Technology. – 2023. – T. 18. – №. 3. – C. 2275-2285.

58 Liao H. et al. Learning-based queue-aware task offloading and resource allocation for space–air–ground-integrated power IoT //IEEE Internet of Things Journal. – 2021. – T. 8. – №. 7. – C. 5250-5263.

59 Koucheryavy A. E. et al. A First-Priority Set of Telepresence Services and a Model Network for Research and Education //International Conference on Distributed Computer and Communication Networks. – Cham : Springer Nature Switzerland, 2022. – C. 208-219.

60 Kirichek R., Paramonov A., Koucheryavy A. Flying ubiquitous sensor networks as a queueing system //2015 17th International Conference on Advanced Communication Technology (ICACT). – IEEE, 2015. – C. 127-132.

61 Kirichek R., Paramonov A., Koucheryavy A. Swarm of public unmanned aerial vehicles as a queueing network //Distributed Computer and Communication Networks: 18th International Conference, DCCN 2015, Moscow, Russia, October 19-22, 2015, Revised Selected Papers 18. – Springer International Publishing, 2016. – C. 111-120.

62 Jin S. C. et al. High integration Ka-band multi-beam antenna for LEO communication satellite //2021 International Conference on Microwave and Millimeter Wave Technology (ICMMT). – IEEE, 2021. – C. 1-3.

63 Du X. et al. Dynamic Resource Allocation for Beam Hopping Satellites Communication System: An Exploration //2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). – IEEE, 2022. – C. 1296-1301.

64 Marcano N. J. H. et al. On the queueing delay of time-varying channels in low earth orbit satellite constellations //IEEE Access. – 2021. – T. 9. – C. 87378-87390.

65 Lin Z. et al. Dynamic beam pattern and bandwidth allocation based on multi-agent deep reinforcement learning for beam hopping satellite systems //IEEE Transactions on Vehicular Technology. – 2022. – T. 71. – №. 4. – C. 3917-3930.

66 Huang Y. et al. Sequential dynamic resource allocation in multi-beam satellite systems: A learning-based optimization method //Chinese Journal of Aeronautics. – 2023. – T. 36. – №. 6. – C. 288-301.

67 Shi D., Liu F., Zhang T. Resource allocation in beam hopping communication satellite system //2020 International Wireless Communications and Mobile Computing (IWCMC). – IEEE, 2020. – C. 280-284.

68 Diao R. et al. A Muti-beam Placement Optimization Scheme in LEO Beam Hopping Satellite Systems //2023 International Conference on Wireless Communications and Signal Processing (WCSP). – IEEE, 2023. – C. 658-663.

69 Mirjalili S. Evolutionary algorithms and neural networks //Studies in computational intelligence. – 2019. – T. 780. – C. 43-53.

70 Huang Y. et al. Sequential dynamic resource allocation in multi-beam satellite systems: A learning-based optimization method //Chinese Journal of Aeronautics. – 2023. – T. 36. – №. 6. – C. 288-301.

71 Ya'acob N. et al. Link Budget and Noise Calculator for Satellite Communication //Journal of Physics: Conference Series. – IOP Publishing, 2019. – T. 1152. – №. 1. – C. 012021.

72 Zhang S. et al. Multi-objective satellite selection strategy based on entropy //2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). – IEEE, 2021. – C. 1-5.

73 Ahmed Y. Recipes for Communication and Signal Processing. – Springer, 2023. – C. 1-188.

74 Sai J. P., Rao B. N. Non-dominated sorting genetic algorithm II and particle swarm optimization for design optimization of shell and tube heat exchanger //International Communications in Heat and Mass Transfer. – 2022. – T. 132. – C. 105896.

75 Li X. et al. Achievement scalarizing function sorting for strength Pareto evolutionary algorithm in many-objective optimization //Neural Computing and Applications. – 2021. – T. 33. – C. 6369-6388.

- 76 Panichella A. An improved Pareto front modeling algorithm for large-scale many-objective optimization //Proceedings of the Genetic and Evolutionary Computation Conference. – 2022. – C. 565-573.
- 77 Wei T. et al. Hybrid satellite-terrestrial communication networks for the maritime Internet of Things: Key technologies, opportunities, and challenges //IEEE Internet of things journal. – 2021. – T. 8. – №. 11. – C. 8910-8934.
- 78 Nasif A., Othman Z. A., Sani N. S. The deep learning solutions on lossless compression methods for alleviating data load on IoT nodes in smart cities //Sensors. – 2021. – T. 21. – №. 12. – C. 4223.
- 79 Ta X. et al. Adaptive spatio-temporal graph neural network for traffic forecasting //Knowledge-based systems. – 2022. – T. 242. – C. 108199.
- 80 Shukla S. et al. Improving latency in Internet-of-Things and cloud computing for real-time data transmission: a systematic literature review (SLR) //Cluster Computing. – 2023. – C. 1-24.
- 81 Berezkin, A. A., R. M. Vivchar, and R. V. Kirichek. "Multi-Criteria Evaluation of UAV Control Efficiency in Hybrid Communication Networks." Proceedings of Telecommunication Universities 10.1 (2024): 18-25.
- 82 Berezkin A. A., Vivchar R. M., Kirichek R. V. Approach to Video Stream Quality Requirements Justification for FPV Control of Unmanned Systems //Proceedings of Telecommunication Universities. – 2024. – T. 10. – №. 4. – C. 7-15.
- 83 Cui G. et al. Latency and energy optimization for MEC enhanced SAT-IoT networks //IEEE Access. – 2020. – T. 8. – C. 55915-55926.
- 84 Zhang S. et al. Joint computing and communication resource allocation for satellite communication networks with edge computing //China Communications. – 2021. – T. 18. – №. 7. – C. 236-252.
- 85 Almasan P. et al. Atom: Neural Traffic Compression with Spatio-Temporal Graph Neural Networks //Proceedings of the 2nd on Graph Neural Networking Workshop 2023. – 2023. – C. 1-6.

- 86 Abderrahim W. et al. Latency-aware offloading in integrated satellite terrestrial networks //IEEE Open Journal of the Communications Society. – 2020. – T. 1. – C. 490-500.
- 87 Xing Z. et al. A multipath routing algorithm for satellite networks based on service demand and traffic awareness //Frontiers of Information Technology & Electronic Engineering. – 2023. – T. 24. – №. 6. – C. 844-858.
- 88 Torkzaban N. et al. Joint satellite gateway placement and routing for integrated satellite-terrestrial networks //ICC 2020-2020 IEEE international conference on communications (ICC). – IEEE, 2020. – C. 1-6.
- 89 Yin Y. et al. Reinforcement Learning-Based Routing Algorithm in Satellite-Terrestrial Integrated Networks //Wireless Communications and Mobile Computing. – 2021. – T. 2021. – №. 1. – C. 3759631.
- 90 Liu Y. et al. Energy-efficient space–air–ground integrated edge computing for internet of remote things: A federated DRL approach //IEEE Internet of Things Journal. – 2022. – T. 10. – №. 6. – C. 4845-4856.
- 91 Zhang Z., Zhang W., Tseng F. H. Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques //IEEE network. – 2019. – T. 33. – №. 1. – C. 70-76.
- 92 Do P. H. et al. Graph Neural Networks for Traffic Classification in Satellite Communication Channels: A Comparative Analysis //Труды учебных заведений связи. – 2023. – Т. 9. – №. 3. – C. 14-27.
- 93 Jia Y. et al. Collaborative transmission in hybrid satellite-terrestrial networks: Design and implementation //2020 IEEE Wireless Communications and Networking Conference (WCNC). – IEEE, 2020. – C. 1-6.
- 94 Wang P. et al. Convergence of satellite and terrestrial networks: A comprehensive survey //IEEE access. – 2019. – Т. 8. – C. 5550-5588.
- 95 Zhang S. et al. Multi-objective satellite selection strategy based on entropy //2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). – IEEE, 2021. – C. 1-5.

- 96 Ahmed Y. Recipes for Communication and Signal Processing. – Springer, 2023. – C. 1-188.
- 97 Sahili Z. A., Awad M. Spatio-temporal graph neural networks: A survey //arXiv preprint arXiv:2301.10569. – 2023.
- 98 Wu F. et al. Simplifying graph convolutional networks //International conference on machine learning. – PMLR, 2019. – C. 6861-6871.
- 99 Kruse R. et al. Multi-layer perceptrons //Computational intelligence: a methodological introduction. – Cham : Springer International Publishing, 2022. – C. 53-124.
- 100 Salehinejad H. et al. Recent advances in recurrent neural networks //arXiv preprint arXiv:1801.01078. – 2017.
- 101 Hua Y. et al. Deep learning with long short-term memory for time series prediction //IEEE Communications Magazine. – 2019. – T. 57. – №. 6. – C. 114-119.
- 102 Xu C. et al. An intrusion detection system using a deep neural network with gated recurrent units //IEEE Access. – 2018. – T. 6. – C. 48697-48707.
- 103 Abdulzahra S. A., Al-Qurabat A. K. M., Idrees A. K. Compression-based data reduction technique for IoT sensor networks //Baghdad Science Journal. – 2021. – T. 18. – №. 1. – C. 0184-0184.
- 104 Routray S. K. et al. Lossless compression techniques for low bandwidth networks //2020 3rd international conference on intelligent sustainable systems (ICISS). – IEEE, 2020. – C. 823-828.
- 105 Sathish Kumar L. et al. Modern Energy Optimization Approach for Efficient Data Communication in IoT-Based Wireless Sensor Networks //Wireless Communications and Mobile Computing. – 2022. – T. 2022. – №. 1. – C. 7901587.
- 106 Ma Z. et al. Deep Lossless Compression Algorithm Based on Arithmetic Coding for Power Data //Sensors. – 2022. – T. 22. – №. 14. – C. 5331.
- 107 Wei J., Hou Z. Compression and Decompression Using Deep Neural Network //Machine Learning and Artificial Intelligence. – IOS Press, 2023. – C. 1-8.
- 108 Viti A., Terzi A., Bertolaccini L. A practical overview on probability distributions //Journal of thoracic disease. – 2015. – T. 7. – №. 3. – C. E7.

109 Фук Хао До. Методика тестирования методов минимизации сетевой задержки на фрагменте гибридной сети связи. Часть I. Интеллектуальная приоритезация и сокращение задержек в очередях// Электросвязь. 2025. № 7 С. 62-72.

110 Фук Хао До. Методика тестирования методов минимизации сетевой задержки на фрагменте гибридной сети связи. Часть II. Сжатие ST-GNN и снижение задержки передачи// Электросвязь. 2025. № 8 С. 17-26.

ПРИЛОЖЕНИЕ А.

**АО «НПП «ИСТА-СИСТЕМС»**

ул. Харченко, д.5, литер А,
С.-Петербург, Россия, 194100,
тел (812) 960 0610, факс (812) 960 0611
E-mail: info@ista.ru, http://www.ista.ru
ОКПО 35521209, ОГРН 1027807594495
ИНН/КПП 7814100536/780201001

«__» _____ 2025 г.

АКТ

**о внедрении научных результатов, полученных в диссертационной работе
До Фук Хао «Разработка моделей и методов минимизации сетевой задержки
на фрагменте гибридной сети связи с использованием технологий
искусственного интеллекта» в деятельность
АО «НПП «ИСТА-Системс»**

Комиссия в составе:

1. Кукунин Дмитрий Сергеевич, зам. технического директора, к.т.н.
2. Головин Валерий Вячеславович, программист.
3. Хоробрых София Алексеевна, инженер.
4. Кукин Роман Сергеевич, специалист внедрения и сопровождения ПО.

составила настоящий акт в том, что научные результаты, полученные До Фук Хао в диссертации «Разработка моделей и методов минимизации сетевой задержки на фрагменте гибридной сети связи с использованием технологий искусственного интеллекта», а именно:

- метод оптимизации распределения сетевых ресурсов на основе графовых нейронных сетей и эволюционных алгоритмов для минимизации задержки в спутниковой сегменте гибридной орбитально-наземной сети связи;
- метод сжатия данных без потерь на основе пространственно-временной графовой нейронной сети для снижения сетевой задержки в спутниковом канале связи;

были использованы в научно-исследовательских и опытно-конструкторских работах АО «НПП «ИСТА-Системс», в частности, при проектировании и разработке резервных спутниковых каналов связи для синхронизации данных комплексной системы безопасности распределенных объектов Балтийского Химического Комплекса.

Председатель комиссии,
Зам. тех. директора
АО «НПП «ИСТА-Системс»,



Д.С. Кукунин

ПРИЛОЖЕНИЕ Б.

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»
(СПбГУТ)

АКТ

14.09. 2025г.

№ 84/54-99

о реализации научных результатов,
полученных До Фук Хао

Комиссия в составе: председателя – и.о. заведующего кафедрой ПИИВТ Краевой Е.В. и членов комиссии: доцента кафедры ПИИВТ, к.т.н. Бирюкова М.А. и доцента кафедры ПИИВТ, к.т.н. Пачина А.В., составила настоящий акт о том, что алгоритмические элементы результатов научных исследований До Фук Хао, а именно:

– Модель прогнозирования и динамической приоритизации трафика на основе многослойного персептрона (MLP) для снижения задержки и потерь пакетов в спутниковом сегменте гибридной орбитально-наземной сети;

– Методы оптимизации распределения сетевых ресурсов на основе графовых нейронных сетей и эволюционных алгоритмов для минимизации задержки в спутниковой сегменте гибридной орбитально-наземной сети связи

внедрены в учебный процесс, в рамках дисциплины «Предиктивная аналитика в сетях и системах связи» и «Разработка приложений искусственного интеллекта в киберфизических системах».

Использование предложенных научных результатов позволило усовершенствовать существующие методы оптимизации сетевых ресурсов, а также расширить класс моделей для прогнозирования и динамической приоритизации трафика в киберфизических системах.

Председатель комиссии:	и.о. завкафедрой		Е.В. Краева
Члены комиссии:	доцент		Т.И. Белая
	доцент		А.В. Пачин