

Казахский Национальный Университет имени Аль-Фараби

УДК 004.852.621.391(043)
рукописи

На правах

НУРАХОВ ЕДИЛЬ СЕРГАЗИЕВИЧ

**Разработка методов прогнозирования показателей качества в 5G
Standalone сетях с использованием машинного обучения**

8D06102 Компьютерная инженерия

Диссертация на соискание ученой степени
доктора философии (PhD)

Отечественный научный консультант:
д.т.н., профессор Ахмед-Заки Д. Ж.

Зарубежный научный консультант:
PhD Электротехники, Профессор Minso Hahn
(KAIST; Korea Advanced
Institute of Science and Technology
г. Тэджон, Республика Корея)

Республика Казахстан
Алматы, 2026

СОДЕРЖАНИЕ	
НОРМАТИВНЫЕ ССЫЛКИ	4
ОПРЕДЕЛЕНИЯ	5
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	6
ВВЕДЕНИЕ	8
1 ОБЗОР АРХИТЕКТУРЫ 5G СЕТЕЙ И МЕТОДОВ АНАЛИЗА	12
1.1 Архитектура 5G Standalone	12
1.2 Open RAN и программно-определяемые подходы	14
1.3 Показатели качества сети в 5G	15
1.4 Методы анализа и прогнозирования	17
1.5 Проблема неконтролируемых данных и мотивация использования testbed	20
2 АРХИТЕКТУРА ПРОГРАММНО-ОПРЕДЕЛЯЕМОЙ ТЕСТОВОЙ СРЕДЫ	25
2.1 Общая архитектура стенда (Open5GS + srsRAN + ZMQ)	25
2.2 Модуль эмуляции пользователей (multi-UE, namespaces)	27
2.3 Эмуляция радиосреды (pathloss, noise, fading)	30
2.4 Процесс запуска и сценарии работы системы	31
2.5 Сбор и хранение KPI (метрики, pipeline, CSV)	35
2.6 Ограничения и особенности тестовой среды	39
2.6.1 Ограничения программно-определяемой реализации	40
2.6.2 Ограничения измерений и влияние транспортных механизмов	40
2.6.3 Угрозы валидности экспериментальной методики	41
3 АНАЛИЗ ПОКАЗАТЕЛЕЙ КАЧЕСТВА СЕТИ	44
3.1 Формирование датасета	44
3.2 Выбор ключевых показателей качества	45
3.3 Формализация исследуемых зависимостей	46
3.4 Анализ зависимостей показателей качества	48
3.5 Анализ узких мест системы (bottleneck)	51
3.6 Причины деградации	52
3.7 Динамика поведения системы	53
3.8 Модель состояний системы	58
3.8.1 Состояние Normal	58
3.8.2 Состояние Loaded	59
3.8.3 Состояние Degraded	60
3.8.4 Состояние Breakdown	61
3.8.5 Переходы между состояниями	62
3.9 Выводы по главе 3	63
4 ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ КАЧЕСТВА	66
4.1 Постановка задачи прогнозирования	66
4.2 Подготовка данных	68
4.2.1 Формирование временных рядов	68
4.2.2 Нормализация данных	70
4.2.3 Формирование обучающих выборок	71

4.3 Модели прогнозирования.....	75
4.4 Результаты прогнозирования по фазово-структурированным данным	79
4.5 Результаты прогнозирования по стохастическим данным.....	83
4.6 Анализ ошибок и ограничений прогнозирования	87
4.7 Анализ прогнозируемости деградации сети	90
4.8 Выводы по главе 4	94
ЗАКЛЮЧЕНИЕ	97
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	99

НОРМАТИВНЫЕ ССЫЛКИ

Диссертация подготовлена с учетом нормативных и методических документов, регламентирующих подготовку и оформление диссертационных работ.

ОПРЕДЕЛЕНИЯ

В настоящей диссертации используются термины и определения, принятые в области сетей связи, машинного обучения и анализа показателей качества обслуживания.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

- 5G - Fifth Generation - сеть пятого поколения
5G SA - 5G Standalone - автономная архитектура 5G
5GC - 5G Core - ядро сети пятого поколения
5GS - 5G System - система пятого поколения
AMF - Access and Mobility Management Function - функция управления доступом и мобильностью
ARIMA - Autoregressive Integrated Moving Average - авторегрессионная интегрированная модель скользящего среднего
AUSF - Authentication Server Function - функция сервера аутентификации
BLER - Block Error Rate - доля блоков с ошибками
CNN - Convolutional Neural Network - свёрточная нейронная сеть
COTS - Commercial Off-The-Shelf - готовое коммерческое оборудование
CSV - Comma-Separated Values - формат хранения табличных данных
DL - Downlink - нисходящий канал
GRU - Gated Recurrent Unit - управляемый рекуррентный блок
GTP-U - GPRS Tunnelling Protocol – User Plane - туннельный протокол пользовательской плоскости
ICMP - Internet Control Message Protocol - протокол управляющих сообщений
KPI - Key Performance Indicator - ключевой показатель эффективности
LSTM - Long Short-Term Memory - долгая краткосрочная память
MAC - Medium Access Control - управление доступом к среде передачи
ML - Machine Learning - машинное обучение
NFV - Network Function Virtualization - виртуализация сетевых функций
NRF - Network Repository Function - функция репозитория сети
O-RAN - Open Radio Access Network - открытая сеть радиодоступа
PCF - Policy Control Function - функция управления политиками
PDCP - Packet Data Convergence Protocol - протокол конвергенции пакетных данных
PHY - Physical Layer - физический уровень
PRB - Physical Resource Block - физический блок ресурсов
QoS - Quality of Service - качество обслуживания
RAN - Radio Access Network - сеть радиодоступа
RIC - RAN Intelligent Controller - интеллектуальный контроллер RAN
RLC - Radio Link Control - управление радиоканалом
RTT - Round-Trip Time - время двойного пробега пакета (задержка)
SCTP - Stream Control Transmission Protocol - протокол передачи с управлением потоком
SDAP - Service Data Adaptation Protocol - протокол адаптации служебных данных
SDN - Software-Defined Networking - программно-определяемые сети
SMF - Session Management Function - функция управления сессиями
TCP - Transmission Control Protocol - протокол управления передачей
UDM - Unified Data Management - единое управление данными

UE - User Equipment - пользовательское оборудование
UL - Uplink - восходящий канал
UPF - User Plane Function - функция пользовательской плоскости
XGBoost - eXtreme Gradient Boosting - экстремальный градиентный бустинг
ZMQ - ZeroMQ - высокопроизводительная библиотека асинхронного обмена сообщениями

ВВЕДЕНИЕ

Актуальность темы исследования

Современные телекоммуникационные системы переходят к архитектурам пятого поколения (5G), в которых ключевую роль играют программно-определяемые подходы, виртуализация сетевых функций и гибкое управление ресурсами. В частности, сети 5G Standalone (SA) представляют собой полностью автономные системы, в которых управление, обработка данных и взаимодействие компонентов реализуются программными средствами.

Одной из центральных задач в таких системах является обеспечение требуемого уровня показателей качества сети, таких как задержка (latency), пропускная способность (throughput) и потери пакетов (packet loss). Эти показатели формируются под влиянием множества факторов, включая число пользователей, характер трафика, параметры распределения радиоресурсов и условия функционирования сети.

В спецификации 3GPP TS 23.501 показатели качества рассматриваются как наблюдаемые параметры 5G System, включающие packet delay, congestion information и data rate per QoS flow [1]. Анализ и прогнозирование показателей качества в 5G Standalone сетях являются не только инженерной, но и нормативно определённой задачей.

В реальных сетях анализ поведения показателей качества затруднён рядом факторов. Во-первых, параметры системы в production-среде не контролируются и изменяются под воздействием внешних условий. Во-вторых, наблюдаемые данные содержат шум и скрытые зависимости, что усложняет выявление причинно-следственных связей. В-третьих, отсутствует возможность воспроизводить одинаковые сценарии работы системы, что ограничивает применение классических методов анализа и прогнозирования.

Дополнительную сложность представляет кросс-слойный характер деградации в 5G системах. Согласно спецификации TS 38.300, деградация может формироваться не только на уровне радиоинтерфейса, но и вследствие переполнения буферов, retransmission-механизмов и особенностей обработки данных на уровнях SDAP, PDCP, RLC и MAC [2].

В современных исследованиях Open RAN и программно-определяемых сетей воспроизводимость экспериментов становится одним из ключевых требований к исследовательской инфраструктуре. Использование открытых программных стеков Open5GS и srsRAN совместно с эмуляцией радиосреды через ZeroMQ позволяет исключить влияние неконтролируемых радиочастотных факторов и обеспечить повторяемость сценариев нагрузочного тестирования [10, 9].

Подобный подход делает возможным проведение контролируемых экспериментов, исследование причин деградации системы и анализ зависимостей между параметрами нагрузки, буферизации, транспортного уровня и показателями качества сети.

Дополнительную актуальность задаче придаёт рост масштабов внедрения 5G Standalone сетей. Согласно Ericsson Mobility Report, к концу 2025 года число

5G-подключений ожидается на уровне около 2.9 млрд, а количество операторов, внедряющих 5G SA, продолжает расти [5]. Это приводит к увеличению объёмов трафика, усложнению механизмов управления и повышению требований к методам анализа и прогнозирования поведения сети.

Дополнительную сложность представляет нелинейный характер поведения системы. При увеличении нагрузки могут возникать пороговые эффекты, при которых незначительное изменение входных параметров приводит к резкому ухудшению показателей качества. Выявление таких эффектов и определение условий деградации системы является важной научной и практической задачей.

В этой связи актуальным направлением является разработка методов анализа, прогнозирования и исследования управления показателями качества в 5G Standalone сетях на основе контролируемых экспериментальных данных и методов машинного обучения.

Степень разработанности проблемы

Вопросы анализа, прогнозирования и управления показателями качества в телекоммуникационных сетях активно исследуются в современной научной литературе.

Стандарты 3GPP подробно определяют архитектуру 5G System, механизмы QoS и параметры мониторинга качества обслуживания [1, 2]. Развитие концепции Open RAN и архитектуры O-RAN [4, 8] дополнительно вводит механизмы программно-определяемого управления сетью через Near-RT RIC и Non-RT RIC [3, 4].

Одновременно активно развиваются открытые программно-определяемые платформы для построения экспериментальных 5G Standalone testbed-сред. Использование Open5GS, srsRAN и ZeroMQ позволяет создавать воспроизводимые сценарии нагрузочного тестирования без применения специализированного радиочастотного оборудования [10, 9].

Значительное количество исследований посвящено проблемам bufferbloat, queueing delay и влиянию TCP-механизмов на показатели качества сети. Работы Gettys и Nichols [17], а также последующие исследования transport-level degradation в мобильных сетях показывают, что механизмы буферизации способны становиться ключевым источником роста задержек и нестабильности системы.

Отдельным направлением являются методы прогнозирования KPI и сетевого трафика. Современные исследования используют как классические статистические модели (ARIMA, SARIMAX), так и методы машинного обучения и глубокого обучения, включая XGBoost, LSTM, GRU и Transformer-подходы [37, 38].

Несмотря на высокий уровень разработанности отдельных направлений, существующие исследования в большинстве случаев рассматривают либо архитектуру сети, либо методы прогнозирования, либо отдельные аспекты QoS/QoE и управления.

При этом в современной литературе недостаточно представлены воспроизводимые экспериментальные методологии, которые в рамках единого открытого 5G Standalone/Open RAN testbed объединяют:

- анализ деградации системы
- выявление bottleneck-механизмов
- прогнозирование KPI
- исследование простых контуров управления

Тем самым, существует необходимость разработки единой воспроизводимой методологии анализа, прогнозирования и исследования управления показателями качества в программно-определяемых 5G Standalone сетях.

Цель и задачи исследования

Целью диссертационной работы является разработка и экспериментальное обоснование подхода к анализу, выявлению закономерностей, прогнозированию и исследованию методов управления показателями качества 5G Standalone системы на основе воспроизводимой тестовой среды.

Для достижения поставленной цели решаются следующие задачи:

1. Разработка воспроизводимой архитектуры 5G Standalone тестовой среды на базе Open5GS, srsRAN и ZeroMQ.
2. Организация контролируемого сбора данных показателей качества в условиях multi-UE сценариев.
3. Анализ зависимостей показателей качества от параметров системы и нагрузки.
4. Выявление условий деградации системы и определение узких мест (bottleneck).
5. Исследование динамики изменения показателей качества, включая эффекты буферизации и транспортных протоколов.
6. Формализация зависимости показателей качества от входных параметров системы.
7. Применение методов машинного обучения для аппроксимации выявленных зависимостей.
8. Интерпретация моделей с целью выявления значимых факторов.
9. Прогнозирование показателей качества на основе временных рядов.
10. Исследование влияния методов управления параметрами сети на показатели качества.
11. Сравнительная оценка различных подходов управления в контролируемых условиях.

Объект и предмет исследования

Объектом исследования являются 5G Standalone сети как динамические системы с изменяемой нагрузкой и управляемыми параметрами.

Предметом исследования являются методы анализа, выявления закономерностей, прогнозирования и исследования подходов управления показателями качества сети на основе контролируемой экспериментальной среды и методов машинного обучения.

Научная новизна

Научная новизна работы заключается в следующем:

1. Предложена воспроизводимая 5G Standalone тестовая среда, обеспечивающая контролируемое варьирование параметров сети.
2. Выявлена нелинейная зависимость показателей качества от уровня нагрузки с наличием пороговых эффектов деградации.
3. Установлено, что ограничивающим фактором системы являются механизмы буферизации и транспортного уровня.
4. Предложен подход к описанию состояния системы через переходы между режимами функционирования.
5. Разработан подход к анализу и прогнозированию на основе экспериментальных данных.
6. Показана возможность выявления значимых факторов с использованием интерпретируемых моделей машинного обучения.
7. Проведено исследование влияния методов управления на показатели качества.

Практическая значимость

Практическая значимость работы заключается в следующем:

- разработана тестовая среда, позволяющая проводить исследования без использования RF-оборудования
- обеспечена возможность анализа поведения системы при изменении нагрузки
- реализована основа для прогнозирования показателей качества
- создана платформа для разработки и тестирования методов управления
- полученные результаты могут быть использованы при разработке систем мониторинга и управления

Структура диссертации

Диссертационная работа состоит из введения, пяти глав, заключения и списка использованных источников. В первой главе представлен обзор архитектуры 5G Standalone сетей и методов анализа. Во второй главе описана экспериментальная тестовая среда. В третьей главе выполнен анализ поведения показателей качества и выявлены закономерности. В четвертой главе рассмотрены методы прогнозирования. В пятой главе исследовано влияние управления на показатели качества. В заключении приведены основные результаты и направления дальнейших исследований.

1 ОБЗОР АРХИТЕКТУРЫ 5G СЕТЕЙ И МЕТОДОВ АНАЛИЗА

Переход мобильных сетей к архитектуре 5G Standalone означает не просто рост пропускной способности и снижение задержки, а изменение самой природы исследуемой системы. В 5G SA сеть становится программно-определяемой, сервисно-ориентированной и телеметрически наблюдаемой, вследствие чего задача анализа состояния сети, фиксации деградации, прогнозирования KPI и исследования простых контуров управления приобретает не только прикладной, но и нормативно и архитектурно обоснованный характер. В спецификации 3GPP TS 23.501 QoS monitoring прямо включает измерение UL/DL/round-trip packet delay, congestion information и data rate per QoS flow, то есть проблема исследования формулируется в координатах стандартизованных наблюдаемых параметров, а не произвольных эмпирических индикаторов [1]. Дополнительную актуальность задаёт масштаб внедрения 5G: в отчёте Ericsson Mobility Report за ноябрь 2025 года указано, что 5G-подписки уже составляют около одной трети всех мобильных подписок, а мобильный трафик вырос на 20 процентов между Q3 2024 и Q3 2025 [5].

Вместе с тем научная сложность задачи определяется не отсутствием измеримых метрик, а конфликтом двух исследовательских требований. С одной стороны, реалистичные операторские сети позволяют наблюдать подлинное поведение системы, но почти не дают контроля над нагрузкой, скрытыми переменными и повторяемостью сценариев. С другой стороны, чистые симуляционные модели обеспечивают контроль, но часто упрощают транспортные механизмы, буферизацию и программную архитектуру сети. Поэтому ключевой методологический вопрос для настоящей диссертации состоит в том, как построить воспроизводимую, но при этом инженерно содержательную экспериментальную среду, в которой можно совместно исследовать архитектуру 5G SA, контуры Open RAN, KPI/QoS, транспортные узкие места и методы прогнозирования. Именно этот конфликт «reproducibility vs realism» далее рассматривается как центральная исследовательская ось главы.

1.1 Архитектура 5G Standalone

Архитектура 5G SA принципиально отличается от унаследованных мобильных систем тем, что ядро сети 5GC строится как service-based system. В TS 23.501 указано, что сетевые функции control plane внутри 5GC должны взаимодействовать через service-based interfaces, а reference points и point-to-point взаимодействия описываются как архитектурные представления процедур, а не как единственно допустимый способ организации сети [1]. Для задач настоящего исследования это означает, что 5G SA естественно интерпретируется как программно-конфигурируемая система, в которой состояние сети и качество обслуживания могут анализироваться через набор формализованных взаимодействий между AMF, SMF, UPF, PCF и другими функциями 5GC [1].

Для первой главы важно подчеркнуть, что архитектура 5G SA релевантна диссертационной теме не потому, что она «современна», а потому, что она делает возможной формализацию наблюдаемости сети. Под наблюдаемостью далее

понимается наличие синхронизированного набора телеметрических и QoS-параметров, позволяющих описывать поведение системы в виде вектора состояния за фиксированный временной интервал. Такое понимание не противоречит стандарту: TS 23.501 прямо отмечает, что параметры QoS monitoring должны описывать качество, «experienced in the 5GS by the application», то есть речь идёт не только о внутренних конфигурационных профилях, но и о наблюдаемом результате функционирования сети для приложения [1]. Следовательно, в контексте диссертации состояние системы может быть определено как совокупность нормированных параметров QoS и диагностической телеметрии, доступных для воспроизводимого анализа.

Однако сама архитектура 5G SA не означает, что деградация производительности может быть сведена к одному уровню анализа. В TS 38.300 описание NR/NG-RAN подчёркивает многослойный характер обработки трафика: QoS flows сопоставляются с радионесителями, а дальнейшая обработка проходит через SDAP, PDCP, RLC и MAC. При этом спецификация прямо фиксирует, что gNB не может гарантировать полное отсутствие L2 buffer overflow и что при переполнении буферов UE может отбрасывать пакеты из L2 buffer [2]. Эта деталь имеет принципиальное методологическое значение: она показывает, что деградация в 5G SA/Open RAN не должна пониматься исключительно как следствие радиоканала; она может возникать в точках буферизации, повторных передач, планирования и обслуживания очередей. Иначе говоря, уже на уровне стандарта задача исследования должна формулироваться как кросс-слойная.

Для целей настоящей работы под событием деградации далее понимается временной интервал, в котором хотя бы один критичный KPI устойчиво отклоняется от целевого профиля качества или от экспериментального baseline. Это определение вводится как рабочее: оно ещё не является результатом главы 3, но задаёт язык, на котором первая глава связывает архитектуру 5G SA с последующим анализом деградации, прогнозированием и исследованием управляющих воздействий. Архитектура 5G SA в диссертации рассматривается не как фоновое описание стандарта, а как основание для формализации состояния системы и для постановки причинно-следственного анализа сетевого поведения [1, 2].

Промежуточный вывод. Архитектура 5G SA задаёт нормативно и технически обоснованный язык исследования: сеть может быть описана как сервисно-ориентированная, наблюдаемая и кросс-слойно деградирующая система. Это создаёт фундамент для перехода от общего обзора архитектуры к анализу Open RAN как архитектурного каркаса для замыкания контуров анализа и управления [1, 2].

Сети 5G представляют собой развитие мобильных телекоммуникационных систем, направленное на обеспечение высокой пропускной способности, низкой задержки и гибкости управления. В отличие от предыдущих поколений, 5G вводит принципиально новую архитектуру, основанную на разделении функций, виртуализации и программной управляемости.

В рамках стандартизации 3GPP выделяются два режима функционирования 5G: Non-Standalone (NSA) и Standalone (SA). В режиме NSA сеть 5G функционирует совместно с ядром 4G, тогда как в режиме SA используется полностью автономная архитектура, включающая собственное ядро сети (5G Core).

Архитектура 5G SA включает следующие основные компоненты:

- UE (User Equipment) - пользовательские устройства, генерирующие и принимающие трафик
- gNB (Next Generation Node B) - базовая станция, обеспечивающая радиодоступ
- 5G Core (5GC) - ядро сети, реализующее функции управления и маршрутизации

В составе 5G Core выделяются следующие функциональные элементы [22, 42, 43]:

- AMF (Access and Mobility Management Function) - управление доступом и мобильностью
- SMF (Session Management Function) - управление сессиями передачи данных
- UPF (User Plane Function) - обработка пользовательского трафика
- PCF (Policy Control Function) - управление политиками

Особенностью архитектуры 5G является разделение на control plane и user plane, что позволяет независимо масштабировать управляющие и пользовательские функции.

Кроме того, архитектура 5G SA построена на принципах Service-Based Architecture (SBA), где взаимодействие между функциями осуществляется через сервисные интерфейсы.

1.2 Open RAN и программно-определяемые подходы

Если 5G SA задаёт, что именно может быть наблюдаемо внутри современной мобильной сети, то O-RAN задаёт, где и как эта наблюдаемость может быть превращена в программно-определяемый контур анализа и управления. В white paper O-RAN Alliance подчёркивается, что архитектура O-RAN строится на открытых и стандартизованных интерфейсах между декомпозированными компонентами RAN, ориентирована на виртуализацию, использование COTS-аппаратуры и внедрение встроенного интеллекта в процессы управления радиодоступом [3]. В этом смысле Open RAN важен для диссертации не как отраслевой лозунг об «открытости», а как архитектурный ответ на проблему vendor lock-in и на необходимость воспроизводимой исследовательской инфраструктуры [3].

Ключевым элементом O-RAN для тематики диссертации выступает RAN Intelligent Controller. В white paper указывается, что non-RT control functionality работает на горизонтах более 1 s, а near-RT control functions - на горизонтах менее 1 s; дополнительно уточняется, что для near-RT RIC характерны control latencies в диапазоне 10 ms–1 s [3, 8]. Non-RT RIC связывается с сервисным и policy management, RAN analytics и model training, тогда как near-RT RIC исполняет

модели и управляющие функции во времени, приближенном к реальному, взаимодействуя с CU/DU через E2 и получая сигналы policy/model distribution через A1 [3, 6, 7, 49]. Для настоящей диссертации это важно потому, что forecasting и control перестают быть внешними приложениями «поверх» сети и становятся естественными элементами архитектурного контура O-RAN.

Вместе с тем из архитектурной зрелости O-RAN не следует автоматическое решение исследовательской задачи. Большая часть документов O-RAN определяет control surfaces, временные горизонты, интерфейсные роли и принципы open programmability, но не даёт воспроизводимой end-to-end методологии, связывающей наблюдение KPI, детекцию деградации, локализацию bottleneck, прогнозирование поведения и проверку управляющего воздействия в одном и том же открытом стенде [10, 38]. Поэтому в диссертации Open RAN рассматривается как среда программируемости и вмешательства, но не как готовый метод анализа. Именно из этого возникает потребность в формализации таких понятий, как управляющее воздействие и контур управления. Под управляющим воздействием далее понимается ограниченное и воспроизводимое изменение параметров системы, способное изменить наблюдаемое состояние сети; под контуром управления - последовательность «наблюдение → интерпретация/прогноз → действие → повторная оценка». Эти определения вводятся здесь как концептуальный мост между архитектурой O-RAN и последующими исследовательскими главами.

Программно-определяемые подходы в более широком смысле - SDN, NFV, виртуализация сетевых функций и вынос управления в программный слой - в контексте диссертации имеют прежде всего методологическую ценность. Они уменьшают жёсткую связанность аппаратуры и алгоритмов, позволяют контролировать конфигурации и тем самым приближают телекоммуникационную систему к форме, пригодной для строгого экспериментального исследования. Но архитектурная программируемость сама по себе ещё не гарантирует воспроизводимость: последняя возникает только тогда, когда сценарий нагрузки, программный стек, каналные условия и телеметрия могут быть запущены повторно с приемлемо малой вариацией результатов. Именно поэтому O-RAN должен быть рассмотрен в связке с open-source testbed, а не изолированно.

Промежуточный вывод. Open RAN создаёт формальный control surface для анализа и управления состоянием сети, однако сам по себе не снимает проблему воспроизводимости и не даёт единой экспериментальной методологии. Следовательно, архитектурная программируемость должна быть дополнена воспроизводимой testbed-средой, на что и будет нацелена следующая логика главы [10, 38].

1.3 Показатели качества сети в 5G

В классическом описании сетей показатели качества часто перечисляются как набор несвязанных метрик - latency, throughput, packet loss. Для диссертационного уровня такой уровень абстракции недостаточен. В 5GS показатели качества должны рассматриваться как параметры, через которые

наблюдается состояние сети и фиксируется момент перехода от нормального функционирования к деградации. В TS 23.501 QoS monitoring включает UL/DL/round-trip packet delay, congestion information и data rate per QoS flow, причём эти параметры измеряются не как случайные внешние наблюдения, а как элементы архитектурно предусмотренного механизма мониторинга [1]. Это позволяет трактовать KPI не как произвольные индикаторы исследователя, а как нормативное ядро описания состояния системы.

Для целей настоящей работы далее используется трёхуровневое представление метрик. На первом, нормативном уровне находятся стандартизованные observables: packet delay, congestion information и data rate per QoS flow [1]. На втором, диагностическом уровне находятся метрики, необходимые для интерпретации причин деградации: показатели очередей, retransmission-related counters, загрузка вычислительных ресурсов, а также иные кросс-слойные параметры, позволяющие сопоставить изменение KPI с поведением транспортного и протокольного стека [2, 17]. На третьем, прикладном уровне находятся агрегаты, более близкие к пользовательскому качеству, например tail latency и иные сервисно-ориентированные характеристики. Такое разделение позволяет избежать частой ошибки, когда прогнозирование «среднего throughput» выдается за анализ качества сети в целом.

Особое место в этой логике занимает congestion information. Стандарт 23.501 прямо указывает, что NG-RAN может измерять и передавать в UPF information about QoS Flow congestion, то есть сама архитектура 5GS допускает явное представление перегрузки как отдельного наблюдаемого параметра [1]. Это принципиально важно для диссертации, поскольку переносит вопрос деградации из чисто постфактум-описательного режима в режим наблюдаемого процесса нарастания очередей и перегрузок. В сочетании с известной проблемой bufferbloat, понимаемой как возникновение избыточно больших и часто заполненных буферов, приводящих к лишней задержке и деградации базовых алгоритмов транспортного контроля перегрузки, это формирует естественный язык для последующей атрибуции bottleneck на уровне transport/queue rather than radio-only explanations [17].

В этой связи в первой главе вводятся следующие рабочие определения. Под наблюдаемостью понимается наличие согласованного набора KPI и диагностической телеметрии, достаточного для описания состояния системы. Под состоянием системы понимается вектор наблюдаемых параметров за заданное временное окно. Под событием деградации понимается переход этого вектора в область, где нарушается целевое качество обслуживания или baseline-функционирование. Под атрибуцией узкого места понимается причинная интерпретация деградации, связывающая её с конкретным ресурсом или слоем системы. Эти определения не подменяют последующую экспериментальную формализацию, но задают понятийный каркас, в котором KPI перестают быть «списком метрик» и становятся языком анализа состояний сети [1, 2, 17].

Промежуточный вывод. KPI/QoS в 5G SA следует рассматривать не как изолированные метрики, а как нормативно и диагностически согласованное

описание состояния сети. Такая постановка делает естественным переход к обзору методов анализа и прогнозирования, но одновременно требует критически оценить, насколько существующие методы действительно работают с наблюдаемыми состояниями и воспроизводимыми данными [1, 2, 17].

1.4 Методы анализа и прогнозирования

Литература по анализу сетевого поведения и прогнозированию трафика чрезвычайно велика, однако для настоящей диссертации важен не полный каталог моделей, а критическое сопоставление классов подходов и их ограничений. В обзоре Ferreira et al. показано, что поле network traffic forecasting охватывает как линейные статистические модели семейства ARMA/ARIMA/SARIMA, так и нелинейные и глубокие модели, включая RNN, LSTM, GRU, CNN и гибридные схемы; при этом авторы специально отмечают как рост роли deep learning, так и сохраняющуюся проблему сопоставимости работ, общих эталонных датасетов и необходимости более зрелых гибридных стратегий [37]. В более прикладной работе Saha et al. прогнозирование рассматривается в контексте proactive network management на реальной ISP-телеметрии, с многогоризонтным прогнозом и явной работой с аномальными и выбросными точками [38]. В совокупности это означает, что прогнозирование уже является зрелым самостоятельным направлением, но его зрелость ещё не означает готовность к прямому переносу в открытую 5G SA/Open RAN testbed-среду.

Сопоставление ключевых источников, на которые реально должна опираться первая глава, приведено в таблице 1. Таблица важна не как библиографический список в сжатом виде, а как средство показать, что проблема достаточно разработана по частям, но не интегрирована в единую методологию.

Таблица 1 - Сравнение ключевых источников по теме исследования

Авто р источник	Го д	Метод / фокус	База данных / эмпирическая опора	Ключевое ограничение для настоящей диссертации
3GPP TS 23.501 [1]	2025	Архитектура 5GS, QoS monitoring, service-based interfaces	Нормативная спецификация 5GS	Стандарт определяет архитектурные и QoS-принципы, однако не предоставляет воспроизводимой методологии анализа причин деградации производительнос ти

Автор / источник	Год	Метод / фокус	База данных / эмпирическая опора	Ключевое ограничение для настоящей диссертации
3GPP TS 38.300 [2]	2025	Многослойная структура NR/NG-RAN, buffering, QoS flows	Нормативная спецификация NR	Описывает кросс-слойную обработку трафика и ограничения buffering, но не предлагает механизм атрибуции bottleneck и деградации
srsRAN Project Documentation [10]	2023.11	Open-source 5G SA deployment, ZeroMQ, network namespaces	Практическая документация по развертыванию	Ориентирована на инженерный deployment и настройку testbed, но не формализует аналитическую методологию исследования KPI
Liu et al., NIST TN 2311 [9]	2024	O-RAN testbeds, automation, aggregated/disaggregated deployment scenarios	Экспериментальные O-RAN стенды NIST	Сильна в инженерии тестовых стендов, однако не объединяет деградацию, прогнозирование и управление в едином исследовательском цикле
Ericsson Mobility Report [5]	2025	Рост 5G deployment, traffic trends, коммерческий контекст	Глобальная операторская статистика и отраслевой мониторинг	Подтверждает актуальность проблемы роста нагрузки, но не решает задачу формализации деградации и прогнозирования KPI
O-RAN Alliance White Paper [3]	2018	Архитектура O-RAN, RIC, A1/E2 interfaces, control loops	Архитектурная концепция O-RAN	Формирует основу для программно-управляемых control loops, однако не

Автор / источник	Год	Метод / фокус	База данных / эмпирическая опора	Ключевое ограничение для настоящей диссертации
				содержит воспроизводимого end-to-end benchmark подхода
Jim Gettys, Kathleen Nichols [17]	2011	Bufferbloat, queue-induced latency, transport-layer degradation	Исследования сетевой буферизации и transport behavior	Концептуально объясняет деградацию очередей и задержек, но не адаптирован к 5G SA/Open RAN testbed-средам
Ferreira et al. [37]	2023	ARIMA/SARIMA, RNN, LSTM, GRU, CNN, hybrid forecasting models	Разнородные open-source сетевые датасеты	Недостаточная привязка моделей прогнозирования к унифицированной telemetry 5G SA/Open RAN
Saha et al. [38]	2024	Multi-step forecasting, outlier handling, proactive network management	Реальная ISP telemetry	Высокая практическая применимость, однако отсутствует открытый воспроизводимый 5G SA/Open RAN experimental context

Аналитически литература распадается на три большие группы. Первая группа - стандарты и архитектурные документы, которые хорошо определяют observability, interfaces и control surfaces, но не сводят их в экспериментальный протокол [1, 2, 10]. Вторая группа - инженерные документы по построению открытых стендов, которые делают воспроизводимость технически достижимой, но почти не переходят к строгому анализу состояний, причин деградации и полезности прогнозирования [10, 9]. Третья группа - работы по traffic/KPI forecasting и по transport-level degradation, которые предлагают зрелые модели и объясняют механизмы качества, но, как правило, работают либо на операторских, либо на общесетевых данных, а не на унифицированной телеметрии открытого 5G SA/Open RAN стенда [17, 37, 38]. Именно на стыке этих трёх групп и возникает свободное научное пространство для настоящей диссертации.

Современная литература убедительно показывает зрелость частных направлений - стандартизованного мониторинга, O-RAN programmability, open-source testbeds, bufferbloat analysis и traffic forecasting, но не даёт единой воспроизводимой методологии, которая связывала бы эти элементы в одном исследовательском цикле. Следовательно, глава 1 должна завершаться не констатацией «много методов существует», а чёткой формулировкой интегрального research gap [10, 9].

1.5 Проблема неконтролируемых данных и мотивация использования testbed

Ключевой методологический барьер анализа современных сетей состоит в том, что production-данные обладают высокой внешней валидностью, но низкой воспроизводимостью. В реальной операторской среде исследователь обычно не контролирует входную нагрузку, транспортные пути, скрытые политики планирования, степень перегрузки, состояние пользовательских буферов и поведение прикладных конечных точек. Именно поэтому модели, обученные на операторских или ISP-данных, нередко обладают локальной полезностью, но плохо переносятся в сравнимый экспериментальный режим; это прямо согласуется с выводами работ по forecasting, где качество прогноза чувствительно к гетерогенности датасетов, аномалиям и ограниченному объёму обучающих наблюдений [37, 38]. В этом смысле проблема «неконтролируемых данных» в диссертации - не техническая мелочь, а эпистемологическое ограничение: без повторяемого сценария трудно отличить причинную связь от случайного совпадения.

С другой стороны, полностью абстрактная симуляция устраняет часть неконтролируемости, но часто делает это ценой удаления именно тех механизмов, которые и интересуют исследование: реального поведения очередей, транспортных окон, буферизации, виртуализированных функций и взаимодействий между RAN, core и host-level ресурсами. Понятие bufferbloat, введённое Gettys и Nichols как наличие избыточно больших и часто заполненных буферов, которые создают ненужную задержку и подрывают фундаментальные механизмы congestion avoidance, важно здесь не столько как историческая ссылка, сколько как напоминание о том, что деградация нередко возникает именно в точках, которые слишком легко теряются в грубых симуляционных моделях [17]. Более того, TS 38.300 прямо допускает L2 buffer overflow в gNB, тем самым подтверждая, что проблема очередей и сбросов пакетов находится внутри нормативного описания NR, а не является внешней по отношению к 5G [2, 17].

Именно поэтому для настоящей диссертации наиболее продуктивной является промежуточная исследовательская позиция: не «чистая production-сеть» и не «чистая симуляция», а открытый программно-определяемый стенд, сохраняющий реальные программные стеки и управляемость сценариев. Документация srsRAN показывает end-to-end 5G SA setup с использованием Open5GS и ZeroMQ, причём с прямым указанием на то, что при использовании ZMQ физическое RF-оборудование не требуется; там же показана

необходимость network namespace для UE-процессов [10, 14]. NIST TN 2311 развивает эту линию в более широкую O-RAN-инфраструктуру: документ описывает both aggregated and disaggregated deployment scenarios, multi-UE ZMQ connection, use of USRP/channel-emulator variants, а также automation tool, который снижает риск human error и ускоряет deployment [9]. Следовательно, воспроизводимость сегодня является не риторическим пожеланием, а технически реализуемым требованием научного дизайна.

Логика проблемы в сжатом виде представлена в таблице 2. В отличие от таблицы 1, она сравнивает не источники, а исследовательские области как элементы единой понятийной карты. Эта таблица нужна для того, чтобы Chapter 1 перестала быть линейным перечнем тем и превратилась в целостную модель проблемы.

Таблица 2 - Карта проблемы: вклад направлений и их ограничения

Область исследования	Что даёт для исследования	Основное ограничение
5G Standalone (5G SA)	Формализует сервисно-ориентированную архитектуру сети и набор наблюдаемых показателей качества обслуживания	Не определяет методы воспроизводимого выявления причин деградации показателей качества
Open RAN / O-RAN	Предоставляет архитектурные механизмы управления и наблюдения (RIC, интерфейсы A1/E2)	Сам по себе не обеспечивает экспериментальную методологию анализа и верификации
Мониторинг QoS и KPI	Позволяет количественно описывать состояние сети через throughput, latency, packet loss и congestion	При отсутствии анализа транспортных механизмов и очередей остаётся преимущественно описательным инструментом
Методы прогнозирования	Предлагают статистические и ML/DL-подходы для краткосрочного прогнозирования состояния сети	Результаты часто основаны на закрытых, несопоставимых или невоспроизводимых наборах данных
Анализ bufferbloat и transport-level degradation	Позволяет объяснять рост задержек, перегрузку очередей и деградацию transport-level показателей	Ограниченно интегрирован с экспериментальными Open RAN / 5G SA testbed-средами
Open testbed-среды	Обеспечивают воспроизводимость экспериментов, контроль	Требуют формализованной методики перехода от

Область исследования	Что даёт для исследования	Основное ограничение
	параметров нагрузки и прозрачность программного стека	развертывания инфраструктуры к причинно-следственному анализу процессов деградации

Для выстраивания дальнейшей логики диссертации далее используются следующие рабочие определения. Повторяемость (repeatability) понимается как возможность воспроизвести один и тот же сценарий нагрузки и конфигурации с ограниченной вариацией результатов. Атрибуция узкого места (bottleneck attribution) понимается как объяснение деградации через идентификацию ресурса или слоя, изменение которого даёт наибольшее улучшение наблюдаемого KPI. Вмешательство (intervention) понимается как ограниченное и контролируемое управляющее действие над параметрами среды. Контур управления (control loop) понимается как цикл «измерение состояния → интерпретация / прогноз → вмешательство → повторная оценка». Именно эти определения позволяют объединить в одной схеме архитектуру 5G SA, programmability O-RAN, диагностические KPI, transport-level degradation и прогнозирование [1, 2, 3].

Данная схема на рисунке 1 показывает, что архитектура 5G SA задаёт наблюдаемость, O-RAN - программируемость, KPI/QoS - измеримый язык описания состояния, bufferbloat и transport bottlenecks - механизмы деградации, а testbed - условия повторяемого исследования. Отсюда и исследовательский разрыв: несмотря на зрелость каждого из указанных направлений по отдельности, в современной литературе недостаточно представлена единая открытая и воспроизводимая методология, позволяющая на одном и том же 5G SA/Open RAN стенде последовательно изучать деградацию, атрибуцию bottleneck, прогнозирование KPI и простые контуры управления [10, 17].

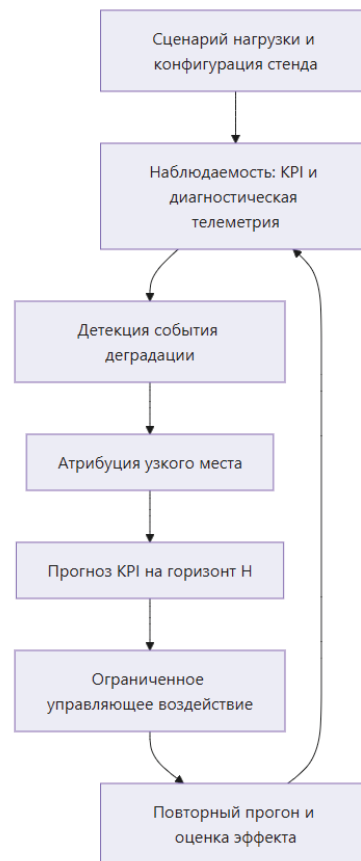


Рисунок 1 - Цикл детекции деградации, атрибуции узкого места и управляющего воздействием в тестовой среде 5G SA

Проблема неконтролируемых данных в исследовании 5G SA не решается ни ссылкой на «реальные операторские логи», ни уходом в чистую симуляцию. Наиболее научно продуктивной оказывается постановка, в которой открытый программно-определяемый testbed используется как инструмент снятия конфликта между реализмом и воспроизводимостью. Именно эта постановка и определяет переход ко второй главе, где описывается архитектура соответствующей экспериментальной среды.

Проведённый обзор показывает, что проблема исследования не является слабо разработанной; напротив, она характеризуется высокой фрагментарной зрелостью. Стандарты 3GPP детально определяют архитектуру 5GS и нормативное ядро QoS monitoring [1, 2]; O-RAN задаёт формальные контуры программно-определяемого управления и распределения моделей [3]; открытые программные стеки и testbed-blueprints доказывают техническую осуществимость воспроизводимых стендов [10, 9]; классическая и современная литература по queueing/transport effects и forecasting предоставляет зрелый теоретический и алгоритмический инструментарий [17, 37, 38].

Научная проблема, следовательно, состоит не в отсутствии знаний, а в отсутствии их воспроизводимого синтеза. Existing approaches, как правило, изолируют либо архитектурные аспекты 5G/O-RAN, либо инженерное развертывание testbed, либо отдельные методы прогнозирования, либо transport-level причины деградации. Между тем для строгого исследования поведения

программно-определяемой 5G SA системы требуется единая методология, которая в одних и тех же условиях связывает: описание состояния системы; детекцию события деградации; атрибуцию узкого места; краткосрочное прогнозирование KPI; и проверку ограниченных управляющих воздействий. Эта методология и образует основной research gap настоящей диссертации [10, 9].

Из этого непосредственно вытекает логика дальнейшей работы. Во второй главе должна быть описана воспроизводимая открытая 5G SA testbed-среда как экспериментальная основа исследования. В третьей главе - выполнен анализ состояний деградации и bottleneck-механизмов. В четвёртой главе - рассмотрены задачи прогнозирования KPI на основе телеметрии testbed-среды. В пятой главе - исследованы ограниченные контуры управления, опирающиеся на прогноз и интерпретацию состояния системы. Тем самым первая глава завершает не общий обзор области, а строгую постановку задачи, выведенную из стандарта, архитектуры и незакрытого научного разрыва.

2 АРХИТЕКТУРА ПРОГРАММНО-ОПРЕДЕЛЯЕМОЙ ТЕСТОВОЙ СРЕДЫ

2.1 Общая архитектура стенда (Open5GS + srsRAN + ZMQ)

Архитектура разработанной программно-определяемой тестовой среды базируется на интеграции решений с открытым исходным кодом, образуя полноценную инфраструктуру 5G Standalone (SA). Данный подход позволяет создать детерминированную и полностью воспроизводимую среду, необходимую для функционального тестирования и отладки стека протоколов без влияния стохастических радиопомех, присущих коммерческим сетям. Логически система разделена на три фундаментальных блока: ядро сети, сеть радиодоступа и эмулируемую радиосреду.

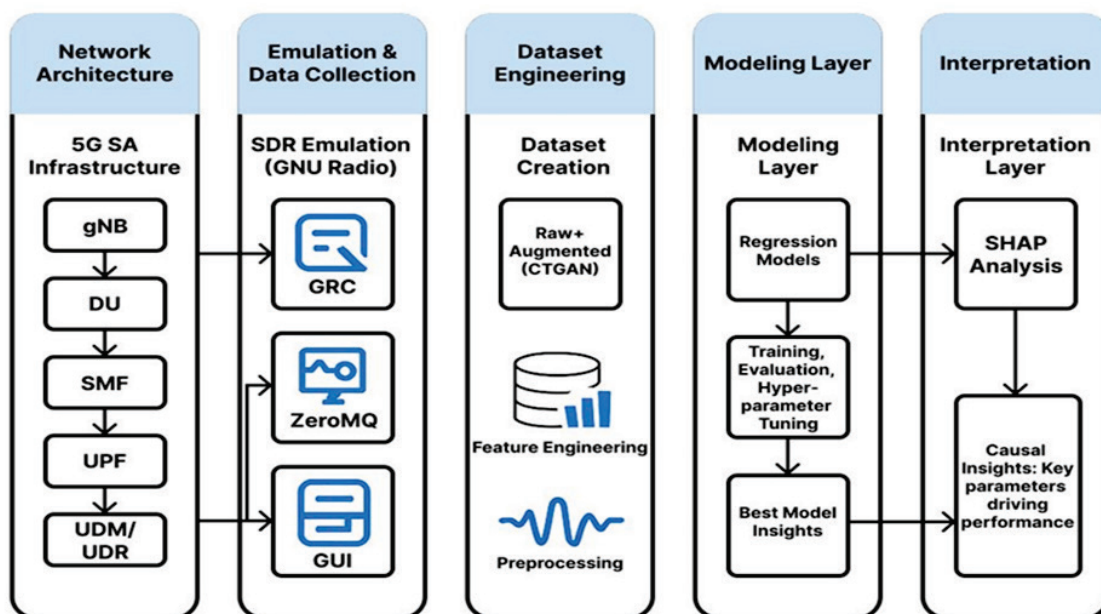


Рисунок 2 - Логическая архитектура программно-определяемой среды 5G SA, иллюстрирующая взаимодействие Open5GS, srsRAN и ZeroMQ

Архитектуры (Service-Based Architecture) в соответствии со спецификациями 3GPP. В развернутой инфраструктуре задействованы все ключевые сетевые функции, обеспечивающие централизованное управление и маршрутизацию. Функция управления доступом и мобильностью (AMF) отвечает за регистрацию абонентских устройств, в то время как функция управления сеансами (SMF) контролирует установление PDU-сессий. Обработка и пересылка пользовательского трафика возложена на высокопроизводительную функцию пользовательской плоскости (UPF). Дополнительно развернуты функция репозитория сетевых сервисов (NRF) [1, 12] и интегрированный узел управления данными абонентов с сервером аутентификации (UDM/AUSF), использующий нереляционную базу данных MongoDB для надежного хранения профилей пользователей и криптографических ключей.

Сеть радиодоступа (RAN) реализована с использованием программного комплекса srsRAN Project. Данный стек предоставляет полноценную программную реализацию базовой станции (gNodeB) и абонентского

оборудования (srsUE), инкапсулируя все уровни радиопrotocolов, включая физический (PHY), уровень управления доступом к среде (MAC), управление радиоканалом (RLC) и протокол конвергенции пакетных данных (PDCP). Взаимодействие между развернутой сетью радиодоступа и ядром Open5GS осуществляется через строго стандартизированные интерфейсы. Обмен сигнальной информацией с AMF происходит по интерфейсу N2 с использованием транспортного протокола SCTP, а маршрутизация пользовательского трафика к UPF реализуется через интерфейс N3 на базе туннелирующего протокола GTP-U.

Ключевой особенностью исследовательской установки, обеспечивающей взаимодействие компонентов без использования специализированного дорогостоящего радиочастотного оборудования (SDR, например, USRP), является применение библиотеки высокопроизводительного межпроцессного взаимодействия ZeroMQ (ZMQ). Базовая станция и эмулируемые абонентские устройства скомпилированы с поддержкой ZMQ, что позволяет осуществлять двусторонний обмен оцифрованными I/Q-сэмплами через шину сообщений непосредственно в оперативной памяти или через стандартные TCP-сокеты локальной машины. Подобная архитектура гарантирует стопроцентную изоляцию эксперимента от непредсказуемой физической радиосреды, создавая идеальные условия (over-the-air emulation) для тестирования логики протоколов. Кроме того, абонентские процессы srsUE инкапсулируются в изолированные сетевые пространства имен (network namespaces) операционной системы Linux, благодаря чему каждое эмулируемое устройство получает независимый сетевой стек и уникальный внутренний IP-адрес, полностью имитируя поведение реального физического аппарата. [10, 13, 14]

Разрабатываемая экспериментальная среда рассматривается в настоящей работе не только как инженерная конфигурация программных компонентов Open5GS, srsRAN и ZeroMQ, но и как инструмент контролируемого программного эксперимента. В отличие от production-сетей такая среда позволяет фиксировать число эмулируемых пользовательских устройств, профиль трафика, режим запуска компонентов и параметры виртуализированной радиосреды, а в отличие от абстрактной симуляции сохраняет реальный программный стек 5G Standalone. Поэтому в рамках диссертации стенд интерпретируется как исследовательская платформа, обеспечивающая компромисс между реалистичностью и воспроизводимостью и создающая основу для сопоставимого сбора KPI, анализа деградации системы и последующей проверки ограниченных управляющих воздействий. Состав стенда структурирован по трём логическим блокам — ядро сети, сеть радиодоступа и радиоинтерфейс, — каждый из которых выполняет строго определённую функциональную роль в экспериментальной методике; их программная реализация, ключевые функции и используемые интерфейсы приведены в таблице 3.

Таблица 3 - Компоненты архитектуры тестовой среды и их функциональное назначение

Логический блок	Программная реализация	Ключевые функции и протоколы	Используемые интерфейсы
Ядро сети (5G Core)	Open5GS	Маршрутизация (UPF), управление мобильностью (AMF) и сеансами (SMF), хранение профилей абонентов (MongoDB)	N2 (SCTP), N3 (GTP-U), Service-Based Interfaces (SBI)
Сеть радиодоступа (RAN)	srsRAN Project	Реализация стека PHY, MAC, RLC, PDCP для программной базовой станции (gNodeB) и абонентского терминала (srsUE)	N2, N3, Uu (эмулированный радиоинтерфейс)
Радиоинтерфейс	ZeroMQ	Эфирная эмуляция, передача оцифрованных I/Q-сэмплов на физическом уровне без использования аппаратных SDR-модулей	Локальные TCP-сокеты, IPC (межпроцессное взаимодействие)

2.2 Модуль эмуляции пользователей (multi-UE, namespaces)

Для проведения комплексного нагрузочного тестирования и оценки алгоритмов распределения ресурсов в разработанной среде реализована масштабируемая эмуляция множественного пользовательского оборудования (multi-UE). Возможности программного стека srsRAN позволяют запускать несколько независимых процессов абонентских терминалов в пределах одного физического вычислительного узла. Подобный подход необходим для моделирования реалистичных сценариев эксплуатации, при которых базовая станция обслуживает параллельные потоки данных от нескольких клиентов, что позволяет исследовать поведение планировщика MAC-уровня и влияние многопользовательской конкуренции на сквозные метрики сети. В рамках проводимых экспериментов архитектура среды позволяет детерминированно масштабировать количество абонентов, например, от единичного тестового устройства, не создающего перегрузок, до группы из пяти параллельно работающих терминалов, генерирующих интенсивный трафик.

Критически важным инженерным аспектом одновременной эмуляции нескольких устройств на одной хост-машине является обеспечение строгой сетевой изоляции генерируемого трафика. Эта задача решается за счет применения механизма сетевых пространств имен (network namespaces) на уровне ядра операционной системы Linux. При инициализации тестового сценария каждый эмулируемый процесс абонентского устройства помещается в выделенное изолированное пространство имен, внутри которого формируется

собственный виртуальный сетевой интерфейс и независимая таблица маршрутизации. Применение данной технологии виртуализации на уровне ОС предотвращает конфликты адресации на уровне хост-машины и полностью исключает перекрестное влияние абонентских процессов друг на друга на транспортном и сетевом уровнях. [10, 14]

Благодаря изолированным сетевым пространствам каждое виртуальное абонентское устройство способно устанавливать абсолютно независимые параллельные подключения к ядру сети 5G. В процессе регистрации в сети каждый эмулируемый терминал выполняет процедуру случайного доступа, устанавливает собственное RRC-соединение с базовой станцией и инициирует создание индивидуальной PDU-сессии с функцией пользовательской плоскости (UPF). В результате ядро сети (через функцию SMF) динамически выделяет каждому пространству имён уникальный IP-адрес из заданного пула, маршрутизируя трафик через соответствующие GTP-туннели. Это даёт возможность запускать независимые экземпляры клиентских приложений (например, iPerf3) внутри каждого пространства имён, детерминированно нагружая транспортную и радиосеть параллельными потоками данных для точного измерения индивидуальной пропускной способности и задержек каждого пользователя. Соответствие между идентификаторами устройств, сетевыми пространствами имён, процессами эмуляции и адресацией плоскости пользователя, а также участие каждого устройства в сценариях нагрузки систематизированы в таблице 4.

Таблица 4 - Логическая структура сетевой изоляции при эмуляции параллельных подключений

Идентификатор устройства	Пространство имен Linux (netns)	Процесс в среде эмуляции	IP-адресация плоскости пользователя (пример)	Участие сценариях нагрузки
UE 1	ue1	srsue (instance 1)	10.45.0.2	Базовая нагрузка (1 UE, параллельная (5 UE))
UE 2	ue2	srsue (instance 2)	10.45.0.3	Исключительная параллельная нагрузка (multi-UE)
UE 3	ue3	srsue (instance 3)	10.45.0.4	Исключительная параллельная нагрузка (multi-UE)
UE 4	ue4	srsue (instance 4)	10.45.0.5	Исключительная параллельная нагрузка (multi-UE)
UE 5	ue5	srsue (instance 5)	10.45.0.6	Исключительная параллельная нагрузка (multi-UE)

С методологической точки зрения последующее экспериментальное исследование опирается на явное разделение управляемых факторов и наблюдаемых выходных показателей. К управляемым факторам относятся число активных пользовательских устройств, уровень предлагаемой нагрузки, профиль трафика, состояние радиосреды и режим управления. Такое представление позволяет интерпретировать каждый сценарий как комбинацию заранее заданных условий, а регистрируемые KPI - как отклик системы на контролируемое воздействие. При этом для корректного академического описания принципиально важно различать экспериментальный замысел и степень его текущей операциональной реализации. Рабочие значения каждого фактора, достигнутый уровень их операционализации и рекомендуемые формулировки для описания в тексте работы приведены в таблице 5.

Таблица 5 - Управляемые факторы эксперимента и степень их операционализации

Фактор	Рабочие значения для главы 2	Текущая операционализация	Как корректно формулировать в тексте
Число UE	1 UE и 5 UE в основном корпусе результатов; расширение масштаба предусмотрено архитектурой стенда	Задаётся напрямую числом запускаемых экземпляров UE	Фактор детерминированно варьируется в сценариях
Уровень нагрузки	low / medium / high	Представлен как сценарный уровень предлагаемой нагрузки; для строгой верификации должен быть явно связан с параметрами генератора трафика	Использовать формулировку «сценарный уровень offered load»
Профиль трафика	DL / UL / Symmetric	Реализуется выбором направления и режима генерации пользовательского трафика	Фактор операционализирован корректно
Состояние радиосреды	good / moderate / poor	Детально описывается в разделе 2.3; часть режимов требует отдельной настройки pathloss/noise	Формулировать как программно задаваемый радиофактор с различной степенью автоматизации
Режим управления	базовый режим / статическая	В базовых сценариях доступны	Не выдавать внешний контур за

Фактор	Рабочие значения для главы 2	Текущая операционализация	Как корректно формулировать в тексте
	конфигурация / внешнее расширение управления	стандартный и статический режимы; внешнее интеллектуальное управление относится к расширению стенда	полностью встроенный элемент б

2.3 Эмуляция радиосреды (pathloss, noise, fading)

Моделирование физической радиосреды в условиях отсутствия аппаратного высокочастотного тракта осуществляется программно-математическими методами с использованием библиотеки ZeroMQ и инструментов эмуляции на базе GNU Radio. Данный подход позволяет транслировать I/Q-сэмплы в виде оцифрованных потоков данных, полностью сохраняя логику работы стека протоколов физического уровня. Главным преимуществом такой архитектуры является создание детерминированной, строго контролируемой среды, абсолютно свободной от непредсказуемого стохастического радиочастотного шума и неконтролируемой интерференции, которые неизбежно присутствуют в реальных коммерческих сетях. [10, 13]

Уровень принимаемого сигнала и общие условия виртуального канала связи в рамках тестовой платформы регулируются программным путем. Ключевым механизмом для изменения качества радиопокрытия выступает целенаправленное и систематическое варьирование параметра затухания сигнала (pathloss) индивидуально для каждого эмулируемого абонентского устройства. Подобный контроль над параметрами эфира позволяет детерминированно изменять условия радиоканала, моделируя различное удаление абонентов от базовой станции или наличие физических преград на пути распространения сигнала.

В ходе проведения многопользовательских экспериментальных исследований платформа позволяет эмулировать дискретные градации качества радиоканала, которые классифицируются как хорошее (good), умеренное (moderate) и плохое (poor) радиопокрытие. Искусственное манипулирование качеством эфира, включая внедрение программного шума и снижение пропускной способности, является критически важным инструментом для оценки эффективности встроенных алгоритмов уровня управления доступом к среде (MAC). Программно деградируя условия канала, исследователи получают возможность в лабораторных условиях наблюдать за реакцией планировщика ресурсов (например, алгоритма Proportional Fair) и механизмами адаптивной модуляции и кодирования, изолируя при этом влияние случайных внешних факторов. Детальная характеристика каждого из трёх эмулируемых состояний радиоканала - включая описание среды, перечень применяемых сценариев

тестирования и наблюдаемое влияние на пропускную способность нисходящего канала - представлена в таблице 6.

Таблица 6 - Классификация эмулируемых условий радиосреды в сценариях нагрузочного тестирования

Состояние радиоканала	Описание эмулируемой среды	Применяемые сценарии тестирования	Влияние на пропускную способность (DL)
Хорошее (Good)	Минимальное затухание (low pathloss), отсутствие программно-индуцированного шума, имитация прямой видимости (LoS)	S1, S2, S3, S4, S7, S8, S11	Стабильная, достигает ~27.5 Mbps при высокой загрузке
Умеренное (Moderate)	Среднее затухание сигнала, эмуляция типичной городской застройки или умеренной дистанции до gNodeB	S9	Сопоставима с хорошим покрытием (~27.5 Mbps), но влияет на буферизацию
Плохое (Poor)	Высокое затухание (high pathloss), деградация сигнала, имитация нахождения абонента на краю соты (Cell Edge)	S10	Поддерживается на уровне ~27.7 Mbps за счет агрессивного планирования MAC

2.4 Процесс запуска и сценарии работы системы

Процесс проведения экспериментальных исследований в программно-определяемой среде строго регламентирован и опирается на принципы автоматизации генерации трафика и абсолютной повторяемости условий. Отсутствие физического эфира исключает стохастические радиопомехи, что позволяет детерминированно оценивать влияние конфигурации сети на итоговую производительность.

Для всесторонней оценки производительности и выявления узких мест была разработана матрица тестовых сценариев (S1–S11), охватывающая различные уровни интенсивности трафика (низкая, средняя, высокая). В качестве основных инструментов для генерации пользовательской нагрузки применяются стандартизированные сетевые утилиты: iPerf3 для оценки пропускной способности TCP-соединений (нисходящий канал DL, подключение к порту 5201) и утилита ping для измерения сквозных сетевых задержек (стандартная отправка 300 ICMP-пакетов на каждое эмулируемое устройство до шлюза 10.45.0.1). В зависимости от конкретного сценария в сети одновременно эмулируется работа от одного до пяти пользовательских устройств (UE). Для

получения референсных значений минимальной сетевой задержки предусмотрен специальный "холостой" (idle) сценарий (S7), в котором активный трафик iPerf полностью отсутствует, позволяя изолировать задержки буферизации от чистой задержки радиоинтерфейса. [15, 16]

Исследовательская платформа предоставляет оператору полный контроль над переменными эксперимента. В процессе тестирования осуществляется целенаправленное управление качеством радиоканала (через эмуляцию хороших, умеренных и плохих условий приема), а также варьирование объемов генерируемых данных. Дополнительно реализована возможность управления логической конфигурацией сети, например, активация так называемого статического контроля (static control) в сценарии S11. При включении этого режима сетевые параметры фиксируются и перестают адаптироваться к текущей динамике трафика, что преднамеренно создает дополнительные задержки в очередях (queuing delay) и позволяет исследователям доказать эффективность адаптивных механизмов управления.

Для обеспечения строгой повторяемости условий и исключения человеческого фактора процесс оркестрации экспериментов был полностью автоматизирован посредством специально разработанного программного скрипта (код алгоритма приведен в Приложении А). Данный инструмент инкапсулирует логику управления жизненным циклом всех узлов программно-определяемой сети. В процессе работы алгоритм осуществляет детерминированную последовательную инициализацию компонентов ядра Open5GS, базовой станции srsRAN и эмулятора радиосреды GNU Radio. После подтверждения готовности инфраструктуры скрипт разворачивает заданное количество экземпляров абонентского оборудования (srsUE), распределяя их по изолированным сетевым пространствам имен. Вся матрица тестируемых конфигураций жестко зафиксирована во внутренних структурах скрипта, что позволяет в автоматическом режиме циклически переключать сценарии, динамически изменяя параметры затухания радиоканала, объемы генерируемой нагрузки, направления потоков данных и режимы управления (включая активацию статических политик или интеллектуального контроллера RIC) [23, 50]. Каждая итерация эксперимента сопровождается принудительной полной очисткой сетевых буферов и перезапуском радиоинтерфейса, что гарантирует абсолютную независимость результатов каждого прогона.

Для обеспечения воспроизводимости и сопоставимости результатов каждый эксперимент проводился в рамках заранее определённого сценария с фиксированной комбинацией управляемых параметров. Варьируемыми факторами выступали число эмулируемых пользовательских устройств, интенсивность предлагаемой нагрузки, состояние радиоэфира и режим управления, тогда как цель каждого сценария формулировалась исходя из конкретной исследовательской задачи. От измерения базовых задержек в штатном режиме до выявления деградации при многопользовательской нагрузке и плохих условиях канала. Перечень базовых сценариев с указанием управляемых параметров и целей эксперимента приведён в таблице 7.

Таблица 7 - Базовые сценарии нагрузочного тестирования и управляемые параметры

Сценарий	Кол-во эмулируемых UE	Интенсивность нагрузки	Состояние радиоэфира	Режим управления Control
S1	1	Низкая (low)	Хорошее (good)	Адаптивный
S3	1	Высокая (high)	Хорошее (good)	Адаптивный
S4	5	Средняя (medium)	Хорошее (good)	Адаптивный
S7	5	Отсутствует (idle)	Хорошее (good)	Адаптивный
S10	5	Средняя (medium)	Плохое (poor)	Адаптивный
S11	5	Средняя (medium)	Хорошее (good)	Статический (static)

Одного перечисления сценариев недостаточно для обеспечения воспроизводимости экспериментов. Для сопоставимости результатов необходим

единый протокол выполнения каждого прогона, в котором фиксируются подготовка среды, порядок запуска компонентов, окно наблюдения и правила материализации результатов. В рассматриваемой тестовой среде протокол прогона не зависит от конкретного сценария и задаёт общий каркас проведения серии экспериментов.

Воспроизводимость результатов обеспечивается строгим соблюдением единого протокола выполнения каждого экспериментального прогона, исключая случайные различия в состоянии среды между сценариями. Протокол охватывает полный жизненный цикл эксперимента от подготовки и валидации компонентов стенда до генерации трафика, сбора KPI и материализации данных и применяется без исключений ко всем сценариям таблицы 7. Последовательность этапов и их содержание приведены в таблице 8.

Таблица 8 - Протокол выполнения одного экспериментального прогона

Этап	Содержание
Подготовка среды	Остановка ранее запущенных процессов, очистка временных файлов и освобождение занятых портов
Валидация ядра	Проверка готовности и запуск сервисов Open5GS
Подъём сети радиодоступа	Запуск gNB и ожидание признака готовности к обработке трафика
Подъём RF-эмуляции	Запуск GNU Radio / ZeroMQ-конвейера
Подъём пользовательских устройств	Запуск требуемого числа UE в отдельных пространствах имён и проверка получения IP-адресов
Проверка связности	Параллельный ping для подтверждения базовой доступности пользовательской плоскости
Генерация трафика	Запуск DL-, UL- или двунаправленного трафика в соответствии со сценарием
Окно наблюдения	Выдерживание фиксированной длительности прогона при неизменных условиях
Парсинг результатов	Извлечение KPI из журналов iperf, ping и gNB
Материализация данных	Запись результатов в master CSV и в CSV, относящийся к конкретному сценарию

Следует различать концептуальную постановку эксперимента и текущую реализацию оркестратора. На данном этапе автоматизированы инициализация Open5GS, gNB, GNU Radio, запуск группы UE, базовая проверка связности, генерация трафика и запись структурированного CSV. Одновременно часть факторов, прежде всего строгая runtime-операционализация offered load, отдельные режимы радиодеградаций и расширенные контуры управления,

реализована не в одинаковой степени полноты. Поэтому корректно характеризовать стенд как воспроизводимую исследовательскую платформу, а не как завершённый факторный стенд с равномерно автоматизированными уровнями управления.

2.5 Сбор и хранение KPI (метрики, pipeline, CSV)

Процедура сбора данных глубоко интегрирована во все уровни программно-определяемой сетевой архитектуры и направлена на формирование масштабного и репрезентативного набора данных для последующего статистического анализа. Для проведения комплексной оценки генерируются сотни файлов с результатами (например, анализ 297 независимых файлов с логами), которые агрегируются в единые структурированные датасеты. Основной формат хранения всей собираемой телеметрии - машиночитаемые CSV-файлы что обеспечивает их прямую совместимость с фреймворками обработки данных и алгоритмами машинного обучения.

В процессе проведения экспериментов система непрерывно фиксирует ключевые метрики производительности в разрезе каждого отдельного пользовательского устройства (UE). Измерения пропускной способности (Throughput) осуществляются посредством логирования результатов утилиты iPerf, где фиксируются средние (Avg), минимальные (Min) и максимальные (Max) значения битрейта нисходящего (DL) и восходящего (UL) каналов. При этом посекундный сбор данных позволяет фиксировать специфический характер трафика: кратковременные пики пропускной способности (TCP bursts), возникающие при увеличении транспортного окна (congestion window), а также периодические провалы скорости до нуля, свидетельствующие о переполнении буферов и срабатывании таймаутов TCP на уровне сети.

Параллельно осуществляется гранулярный сбор метрик сетевой задержки (Latency) посредством отправки фиксированного пакета ICMP-запросов (по 300 пакетов на каждое устройство до шлюза). Для объективной оценки качества обслуживания (QoS) логируются не только средние значения (Avg Ping), но и минимальные (Min), максимальные (Max) задержки, среднее квадратичное отклонение (Mdev), а также 95-й перцентиль (P95) [20, 48]. Сбор метрики потерь пакетов (Packet Loss Rate) выступает критическим индикатором TCP-ретрансмиссий и нарушений последовательности пакетов при перегрузке очередей.

Процедура агрегации разрозненных телеметрических данных и формирования итогового набора данных также автоматизирована в рамках единого скрипта-оркестратора (Приложение А). По завершении заданного временного окна генерации трафика алгоритм инициирует внутренние процедуры парсинга текстовых логов утилит iPerf и Ping. С помощью встроенных регулярных выражений извлекаются посекундные значения пропускной способности, расчетные показатели средней задержки, значения 95-го перцентиля и доля потерянных пакетов. Одновременно скрипт осуществляет лексический анализ системного журнала базовой станции (gNodeB) для извлечения низкоуровневых параметров физического уровня, включая

статистику распределения физических ресурсных блоков (PRB) и частоту ошибочных блоков (BLER). Вся извлеченная и нормализованная информация объединяется в единую структуру и экспортируется в сводный CSV-файл, формируя готовый датасет для последующего применения алгоритмов машинного обучения и статистического анализа.

Структура датасета позволяет проводить глубокий временной анализ динамики сети. Благодаря посекундному сбору и нумерации пакетов, итоговый набор данных отчетливо демонстрирует фазовые переходы в поведении задержек при активации пользовательского трафика. В частности, данные измерений ICMP-пакетов четко структурируются на три временные фазы, отражающие жизненный цикл переполнения сетевых буферов (bufferbloat) в эмулируемой среде [17, 18].

Формируемый в ходе экспериментов датасет охватывает три категории метрик, каждая из которых ориентирована на измерение отдельного аспекта производительности сети: пропускной способности, сквозной задержки и надёжности доставки пакетов. Совокупность этих показателей обеспечивает достаточную наблюдаемость для выявления деградации системы, анализа буферного переполнения и последующего применения методов машинного обучения. Состав измеряемых параметров и цели их сбора систематизированы в таблице 9.

Таблица 9 - Структура собираемых метрик и их характеристика в формируемом датасете

Категория метрик	Измеряемые параметры	Цель сбора данных в датасете
Пропускная способность (DL/UL)	Avg DL, Min DL, Max DL (Мбит/с)	Оценка пропускной способности, фиксация TCP bursts и периодов нулевой передачи (buffer overflow)
Сквозная задержка (Ping)	Avg Ping, Min/Max Ping, P95, Mdev (мс)	Оценка сетевых задержек, расчет вариации задержки (джиттера) и выявление критических перегрузок
Надежность доставки	Количество потерянных пакетов (из 300 отправленных)	Идентификация масштабов отбрасывания пакетов (packet drops) и TCP-ретрансмиссий при пиковой нагрузке

Для корректной интерпретации формируемого датасета необходимо учитывать неоднородность временной структуры собираемых KPI. В текущей реализации посекундная детализация обеспечена прежде всего для показателей пропускной способности, тогда как метрики задержки и потерь пакетов вычисляются как агрегаты по прогону, а PRB и BLER извлекаются из журнала

gNB как моментальные значения. Следовательно, итоговый CSV следует трактовать как смешанно-агрегированный набор наблюдений, а не как полностью синхронизированный многомерный временной ряд. Источники данных, способы их получения и фактическая временная гранулярность каждого KPI приведены в таблице 10.

Таблица 10 - KPI, источники и фактическая временная гранулярность в текущем pipeline

KPI	Источник данных	Способ получения текущей реализации	Фактическая гранулярность
DL throughput	iperf log	Парсинг интервалов передачи последующей агрегацией каждого UE	Посекундная
UL throughput	iperf log	Парсинг интервалов передачи последующей агрегацией каждого UE	Посекундная
Средняя задержка (RTT average)	ping log	Усреднение значений RTT в пределах одного прогона эксперимента	По прогону
P95 задержки	ping log	Расчёт 95-го перцентиля RTT в пределах одного прогона эксперимента	По прогону
Packet loss	ping summary	Извлечение итоговой доли потерянных пакетов после завершения прогона	По прогону
PRB utilization	gNB log	Извлечение значений загрузки PRB из текстовых журналов gNB	Моментальное значение / snapshot
BLER	gNB log	Извлечение значений BLER из текстовых журналов gNB	Моментальное значение / snapshot

Поэтому последующий анализ должен различать собственно временные ряды throughput и агрегированные характеристики других KPI. Такое уточнение не уменьшает аналитическую ценность датасета, но задаёт корректные границы для статистической интерпретации, выявления деградации и последующего использования методов машинного обучения.

Таблица 11 - Динамическая структура данных задержки (на примере сценария S9 с 5 UE)

Временная фаза	Пакеты (в датасете)	Диапазон задержки (мс)	Характеристика сетевого поведения
Фаза 1 - Стабильность	#1 – #11	13 – 40 мс	Сеть свободна, iPerf еще не создал критическую очередь в буферах
Фаза 2 - Рост и насыщение	#12 – #120	128 – 5870 мс	Буферы активно заполняются генерируемым трафиком, задержка растет линейно
Фаза 3 - Осцилляции	#121 – #300	4000 – 5870 мс	Генерация трафика iPerf завершается, наблюдается дренирование (опустошение) очередей

Описанная трёхфазная динамика наглядно прослеживается на графике временного ряда RTT, построенного по данным сценария S9. График отражает последовательный переход от стабильно низких задержек в начале прогона к стремительному росту по мере заполнения буферов, устойчивым осцилляциям в диапазоне 4000–5870 мс в фазе насыщения и резкому спаду после завершения генерации трафика iPerf. Данная визуализация служит прямым эмпирическим подтверждением эффекта *bufferbloat* в условиях многопользовательской нагрузки и обосновывает необходимость перцентильных метрик (P95) наряду со средними значениями RTT, как показано на рисунке 3.

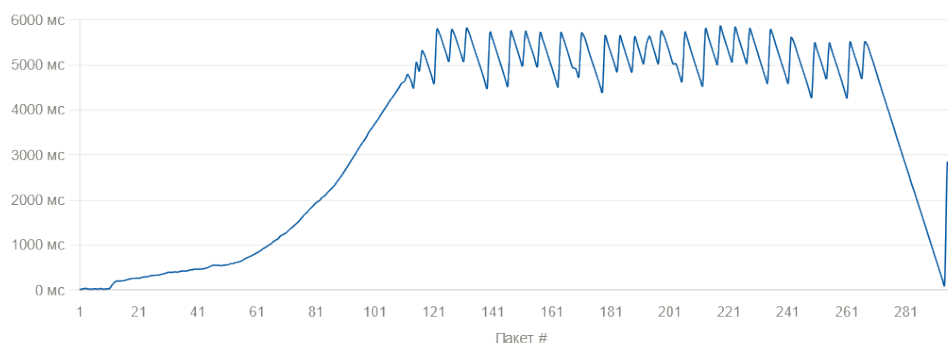


Рисунок 3 - Динамика изменения задержки Ping в сценарии S9 (5 UE), демонстрирующая фазы роста, насыщения и осцилляций при переполнении буферов

В диапазоне пакетов #12–#120 происходит переход системы из стабильного состояния в режим перегрузки. После достижения насыщения наблюдаются осцилляции задержки, обусловленные периодическим дренированием очередей и повторным заполнением буферов.

На рисунке 4 наблюдаются кратковременные пики пропускной способности (TCP bursts), возникающие вследствие расширения congestion window транспортного протокола TCP. Одновременно фиксируются периоды падения пропускной способности до нулевых значений, что указывает на переполнение буферов и срабатывание механизмов TCP timeout.

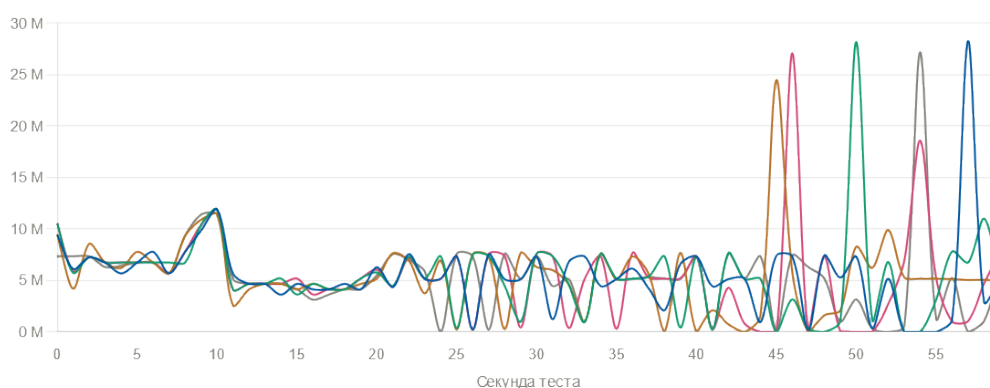


Рисунок 4 - Посекундная динамика пропускной способности DL для нескольких UE в сценарии многопользовательской нагрузки

Совокупность наблюдаемых явлений — фазовый рост задержки, осцилляции RTT при насыщении буферов, TCP bursts и периодическое падение пропускной способности до нуля — формирует целостную картину деградации производительности в многопользовательской среде 5G SA под средней нагрузкой. Эти эффекты не являются артефактами эмуляции: они воспроизводимо фиксируются в рамках единого протокола прогона и отражают реальное поведение транспортного стека при буферном переполнении. Тем самым собранный датасет обеспечивает достаточную эмпирическую базу для перехода к количественному анализу деградации, атрибуции узких мест и построению прогностических моделей KPI в последующих разделах работы.

2.6 Ограничения и особенности тестовой среды

Несмотря на полное архитектурное соответствие стандартам 3GPP и успешную реализацию механизмов QoS flows и 5QI, разработанная программно-определяемая тестовая среда обладает рядом ограничений, обусловленных особенностями программной реализации на стандартном серверном оборудовании общего назначения (COTS).

2.6.1 Ограничения программно-определяемой реализации

Фундаментальным ограничением разработанной среды является отсутствие специализированного аппаратного ускорения физического уровня. В отличие от коммерческих телекоммуникационных систем, использующих FPGA, DSP или ASIC-ускорители, обработка PHY- и MAC-уровней в рамках программной реализации выполняется центральным процессором под управлением операционной системы общего назначения Linux.

Вследствие этого на характеристики системы существенное влияние оказывают:

- планировщик задач Linux
- переключения контекста процессов
- задержки обработки пакетов в kernel/user space
- вычислительная нагрузка на CPU

Данные особенности приводят к появлению дополнительного джиттера и нестабильности задержек даже при относительно низкой нагрузке.

Кроме того, программная реализация не использует физическое SDR-оборудование и не взаимодействует с реальным радиочастотным эфиром. Взаимодействие между компонентами осуществляется через ZeroMQ и передачу I/Q-сэмплов по локальным TCP-сокетах, что обеспечивает высокую воспроизводимость экспериментов, однако ограничивает возможность моделирования реальных физических эффектов распространения радиоволн.

Проведённые эксперименты также показали, что ограничивающим фактором производительности системы выступает не радиоканал, а вычислительные ресурсы хост-системы, прежде всего способность UPF и сетевого стека обрабатывать пользовательский трафик при высокой нагрузке.

В условиях многопользовательских сценариев планировщик MAC-уровня фактически распределяет между пользователями не только радиоресурс, но и доступное процессорное время вычислительной платформы.

2.6.2 Ограничения измерений и влияние транспортных механизмов

Особенности программной архитектуры тестовой среды напрямую влияют на характер собираемых измерений и интерпретацию результатов.

Использование ZeroMQ и отсутствие реального радиоэфира ограничивают возможность получения достоверных физических радиометрик, характерных для коммерческих сетей. В частности, показатели BLER и PRB в рамках программной эмуляции имеют ограниченную физическую достоверность и не отражают влияние многолучевого распространения, интерференции и внешних радиопомех.

В ходе экспериментов было установлено, что существенное влияние на показатели качества оказывают механизмы транспортного уровня и системной буферизации.

При сценариях с одним пользовательским устройством средняя задержка остаётся относительно стабильной и составляет порядка 600 мс даже при активной генерации трафика. Однако при переходе к multi-UE сценариям наблюдается резкий рост задержки до значений свыше 3000–5000 мс.

При этом «холостой» сценарий без генерации пользовательского трафика показывает задержку порядка 31 мс, что указывает на отсутствие ограничений со стороны радиоинтерфейса.

Следовательно, основной причиной роста задержек являются:

- переполнение сетевых буферов
- рост очередей
- эффекты bufferbloat
- TCP timeout и retransmissions

Посекундный анализ трафика также показывает наличие кратковременных TCP bursts, при которых пропускная способность резко возрастает вследствие расширения congestion window, а затем периодически падает до нуля при переполнении буферов и возникновении TCP timeout.

Дополнительное ухудшение показателей наблюдается при использовании статических конфигураций управления (static control), при которых система не адаптируется к изменению нагрузки, что приводит к дополнительному росту queuing delay.

Полученные результаты отражают поведение прежде всего программной реализации 5G Standalone среды и требуют осторожной интерпретации при переносе выводов на реальные коммерческие сети.

2.6.3 Угрозы валидности экспериментальной методики

Помимо технических ограничений программно-определяемой среды, для корректной интерпретации результатов необходимо выделить и методологические угрозы валидности.

К угрозам конструктивной валидности относится неполная операционализация части факторов эксперимента. Для отдельных параметров постановка эксперимента в настоящий момент опережает степень автоматизации оркестратора: это относится прежде всего к строгой runtime-реализации уровней предлагаемой нагрузки, к отдельным режимам деградации радиосреды и к расширенным режимам управления. Поэтому при интерпретации результатов необходимо различать факторы, задаваемые полностью автоматически, и факторы, представленные на уровне сценарной декларации либо частичной настройки.

Измерительная валидность

К угрозам измерительной валидности относятся смешанная временная гранулярность KPI, зависимость части парсинга от текстового формата журналов и использование реконструируемой временной метки при сборке итогового CSV. Эти особенности не исключают применение сформированного датасета для анализа, однако требуют осторожности при синхронизации разнотипных метрик и при построении причинно-следственных интерпретаций временной динамики системы.

Внешняя валидность

Внешняя валидность результатов ограничивается тем, что разработанный стенд воспроизводит поведение программно-определяемой 5G Standalone среды в контролируемых лабораторных условиях. Использование ZeroMQ, GNU Radio и программной эмуляции радиосреды обеспечивает высокую повторяемость прогонов, но не охватывает всего спектра эффектов реального радиочастотного окружения и эксплуатационных особенностей коммерческих сетей. По этой причине полученные выводы следует интерпретировать как валидные прежде всего для открытого исследовательского testbed-класса систем.

Сформулированные угрозы валидности не снижают значимости проведённых экспериментов, но определяют корректные границы их обобщения и тем самым повышают научную строгость диссертационного текста.

Выводы по главе 2

В результате проведённой работы была разработана воспроизводимая программно-определяемая экспериментальная среда 5G Standalone, предназначенная для контролируемого исследования поведения сети при различных уровнях нагрузки и конфигурации системы. Архитектура стенда построена на базе Open5GS, srsRAN и ZeroMQ и обеспечивает полноценную реализацию ключевых компонентов 5G SA, включая ядро сети, радиодоступ и эмулируемую радиосреду.

В рамках разработанной среды реализована масштабируемая multi-UE эмуляция пользовательских устройств с использованием механизмов network namespaces, обеспечивающих независимую сетевую изоляцию каждого эмулируемого абонента. Это позволило проводить воспроизводимые многопользовательские эксперименты и исследовать влияние конкурентной нагрузки на показатели качества сети.

Разработанная система эмуляции радиосреды позволила программно изменять параметры радиоканала, включая уровень затухания сигнала и условия распространения, обеспечивая контролируемое моделирование различных сценариев радиопокрытия без использования физического SDR-оборудования.

Для проведения исследований был реализован автоматизированный механизм запуска сценариев и сбора телеметрии, обеспечивающий повторяемость экспериментов и исключение влияния человеческого фактора. Сформирован набор сценариев нагрузочного тестирования, включающий различные комбинации количества пользователей, интенсивности трафика, состояния радиосреды и режимов управления.

Организован процесс сбора, агрегации и хранения экспериментальных данных в формате CSV. В ходе экспериментов осуществлялся посекундный сбор метрик throughput, latency, packet loss, а также вспомогательных показателей транспортного и физического уровней. Это позволило сформировать структурированный датасет, пригодный для дальнейшего статистического анализа и применения методов машинного обучения.

Проведённый анализ показал, что в условиях программной реализации ключевым ограничивающим фактором системы выступают механизмы транспортного уровня и буферизации, а не радиоинтерфейс. В

многопользовательских сценариях наблюдаются выраженные эффекты bufferbloat, рост очередей, TCP timeout и осцилляции задержек, существенно влияющие на показатели качества обслуживания. [17]

Также установлено, что программная архитектура среды обладает рядом ограничений, связанных с отсутствием специализированного аппаратного ускорения РНУ-уровня, влиянием планировщика Linux, вычислительных ограничений CPU и отсутствием физического радиоэфира. Это требует осторожной интерпретации результатов при переносе выводов на реальные коммерческие сети.

3 АНАЛИЗ ПОКАЗАТЕЛЕЙ КАЧЕСТВА СЕТИ

В рамках проведённого исследования был сформирован экспериментальный датасет, предназначенный для анализа поведения сети 5G Standalone в условиях изменяющейся нагрузки и последующего прогнозирования показателей качества функционирования системы. [1, 39]

3.1 Формирование датасета

Формирование датасета выполнялось на основе воспроизводимой экспериментальной среды, построенной с использованием компонентов Open5GS, srsRAN Project и ZeroMQ. Экспериментальная инфраструктура обеспечивала возможность проведения контролируемых сценариев с фиксированными параметрами нагрузки, количеством пользователей и условиями функционирования сети. [10, 11]

Структура обучающего набора данных представляется в виде зависимости:

$$X \rightarrow Y \quad (1)$$

где: X- совокупность входных параметров и наблюдаемых характеристик системы; Y- выходные показатели качества сети.

В качестве входных параметров Xиспользовались следующие группы признаков:

- число активных пользователей (UE);
- уровень генерируемой нагрузки (low, medium, high);
- параметры радиосреды (good, moderate, poor);
- режим функционирования системы управления;
- характеристики сетевого трафика;
- агрегированные показатели текущего состояния сети.

Дополнительно в состав признаков включались параметры временной динамики системы:

- lag-признаки;
- rolling mean;
- rolling standard deviation;
- производные временных рядов.

Формирование датасета выполнялось на основе посекундной агрегации логов: iperf; ping; gNB telemetry.

На этапе предобработки осуществлялись:

- синхронизация временных меток;
- очистка аномальных значений;
- удаление пропусков;
- формирование временных рядов;
- исключение признаков, содержащих data leakage.

Для обеспечения корректности прогнозирования использовалось временное разделение выборки:

$$\text{train} \rightarrow \text{past}, \text{test} \rightarrow \text{future} \quad (2)$$

что исключало попадание будущих наблюдений в процесс обучения модели.

В качестве выходных переменных рассматривались:

- throughput (DL/UL);
- latency (mean, p95);
- packet loss;
- интегральный показатель состояния сети state_score;
- класс состояния системы.

Прогнозирование выполнялось для горизонтов: $\Delta t \in \{1,2,3,5,10,15,20\}$ секунд

Сформированный датасет отражает как текущее состояние сети, так и динамику переходных процессов деградации, что позволяет использовать его для задач анализа, прогнозирования и исследования устойчивости функционирования 5G SA системы.

3.2 Выбор ключевых показателей качества

В качестве основных показателей качества функционирования сети были выбраны:

- задержка передачи пакетов (latency);
- пропускная способность (throughput);
- потери пакетов (packet loss).

Выбор указанных параметров обусловлен тем, что они непосредственно характеризуют состояние сети и позволяют оценивать влияние изменения нагрузки на качество обслуживания пользователей.

Показатель throughput отражает способность сети обеспечивать передачу пользовательского трафика и характеризует эффективность использования доступных ресурсов системы.

Показатель packet loss позволяет выявлять критические режимы функционирования, связанные с перегрузкой, переполнением очередей и нарушением стабильности передачи данных.

Наиболее чувствительным индикатором деградации сети рассматривается задержка (latency). В ходе экспериментов было установлено, что при увеличении нагрузки рост задержки возникает раньше значительного падения throughput, что делает latency ранним признаком формирования перегрузки системы.

Дополнительно анализировались:

- среднее значение задержки;
- 95-й перцентиль задержки (p95).

Использование p95 позволяет учитывать редкие аномально высокие значения RTT, отражающие нестабильность и формирование очередей в системе.

Для интегральной оценки состояния сети был введён агрегированный показатель: state_score(t), который формируется на основе совокупности нормализованных KPI и используется для анализа переходов между состояниями: stable; loaded; degraded; critical.

Подобный подход позволяет исследовать деградацию сети как непрерывный переходный процесс, а не как набор независимых событий отказа.

3.3 Формализация исследуемых зависимостей

В рамках настоящего исследования анализ экспериментальных данных рассматривается как задача выявления закономерностей между параметрами функционирования программно-определяемой 5G Standalone системы и наблюдаемыми показателями качества обслуживания сети. В отличие от традиционного подхода, при котором KPI анализируются изолированно как отдельные метрики, в данной работе показатели качества рассматриваются как отклик системы на контролируемое изменение параметров нагрузки, конфигурации сети и условий функционирования.

Исследуемая система интерпретируется как динамическая многопараметрическая среда, состояние которой определяется совокупностью входных факторов и наблюдаемых показателей качества. В общем виде данная зависимость может быть представлена следующим образом:

$$KPI = f(X) \quad (3)$$

где:

- (X) - совокупность входных параметров системы;
- (KPI) - наблюдаемые показатели качества сети.

В рамках проведённых экспериментов в качестве входных параметров рассматриваются:

- число активных пользовательских устройств;
- уровень генерируемой нагрузки;
- параметры радиосреды;
- загрузка радиоресурсов;
- режим управления системой;
- параметры транспортного поведения сети.

В качестве основных наблюдаемых показателей качества используются:

- latency;
- throughput;
- packet loss;
- показатели загрузки радиоресурсов;
- параметры деградации системы.

Подобная постановка позволяет рассматривать экспериментальную среду не просто как средство генерации трафика, а как управляемую исследовательскую систему, в которой изменение отдельных факторов вызывает наблюдаемое изменение состояния сети. Это принципиально важно для последующего анализа причин деградации, выявления bottleneck-механизмов и применения методов машинного обучения.

Следует отметить, что исследуемые зависимости не предполагаются заранее линейными. Напротив, результаты предварительного анализа показывают, что поведение программно-определяемой 5G Standalone среды характеризуется выраженной нелинейностью, наличием пороговых эффектов и переходов между различными режимами функционирования. При увеличении

нагрузки незначительное изменение параметров системы может приводить к резкому ухудшению показателей качества, росту очередей и развитию эффектов bufferbloat. [17, 18]

В связи с этим основной задачей анализа является не только определение факта зависимости между параметрами, но и исследование:

- характера изменения KPI;
- областей устойчивого функционирования;
- моментов перехода системы в режим деградации;
- влияния отдельных факторов на поведение сети;
- взаимосвязей между различными показателями качества.

Особое значение имеет исследование воспроизводимости выявляемых закономерностей. Поскольку эксперименты проводятся в контролируемой testbed-среде, появляется возможность многократного повторения одинаковых сценариев с минимальным влиянием внешних факторов. Это позволяет рассматривать наблюдаемые зависимости как устойчивые свойства исследуемой системы, а не как случайные артефакты единичных измерений.

Для систематизации дальнейшего анализа в работе формализуется набор исследуемых зависимостей, отражающих основные направления экспериментального исследования.

Таблица 12 - Формализованные исследуемые зависимости в экспериментальной 5G Standalone среде

Исследуемая зависимость	Формализация	Что исследуется	Практический смысл
Load → Latency	$(\text{Latency} = f(\text{load}, \text{UE}))$	Рост задержек при увеличении нагрузки	Анализ деградации сети
Load → Throughput	$(\text{Throughput} = f(\text{load}, \text{UE}))$	Насыщение системы и ограничение throughput	Определение пределов сети
Load → Packet Loss	$(\text{Loss} = f(\text{load}))$	Потери пакетов при перегрузке	Выявление критических состояний
UE Count → KPI	$(\text{KPI} = f(\text{UE}))$	Изменение KPI при росте числа пользователей	Анализ масштабируемости сети
PRB Utilization → KPI	$(\text{KPI} = f(\text{PRB}))$	Влияние загрузки радиоресурсов	Анализ эффективности использования PRB

Исследуемая зависимость	Формализация	Что исследуется	Практический смысл
Radio Condition $\rightarrow$ KPI	$(KPI=f(\text{radio}))$	Влияние качества радиоканала	Анализ устойчивости сети
KPI $\rightarrow$ Network State	$(\text{State}=f(KPI))$	Переход системы к деградации	Выделение режимов работы сети
KPI во времени	$(KPI_t=f(t))$	Динамика изменения показателей качества	Основа временного прогнозирования
Scenario Repeatability	$(KPI_{\{run1\}} \approx KPI_{\{run2\}})$	Повторяемость результатов экспериментов	Доказательство воспроизводимости среды
Комплекс параметров $\rightarrow$ KPI	$(KPI=f(\text{load}, \text{PRB}, \text{radio}, U_E, \dots))$	Совместное влияние факторов	Основа многомерных ML-моделей

Представленные зависимости формируют основу дальнейшего экспериментального анализа. Их исследование позволяет перейти от описательного наблюдения отдельных метрик к системному анализу поведения программно-определяемой 5G Standalone среды как динамической многопараметрической системы.

Кроме того, формализация подобных зависимостей создаёт основу для применения методов машинного обучения, поскольку позволяет интерпретировать задачу прогнозирования как задачу аппроксимации отображения между входными параметрами системы и наблюдаемыми показателями качества.

Таким образом, предложенный подход связывает архитектуру экспериментального стенда, процедуру сбора телеметрии, анализ деградации, прогнозирование KPI и исследование методов управления в рамках единой исследовательской модели.

3.4 Анализ зависимостей показателей качества

Анализ экспериментальных данных показывает наличие выраженной зависимости показателей качества сети от числа пользователей, интенсивности генерируемого трафика и состояния транспортных механизмов системы. Полученные результаты демонстрируют нелинейный характер деградации сети при увеличении нагрузки и позволяют выявить особенности поведения программно-определяемой 5G Standalone среды в многопользовательских сценариях.

Зависимость latency от числа пользователей

Экспериментальные исследования показывают, что увеличение количества одновременно обслуживаемых пользовательских устройств (UE) приводит к существенному росту сетевой задержки.

Установлено, что:

- при 1 UE средняя задержка составляет около ~600 мс
- при 5 UE задержка возрастает до ~2800–3200 мс
- при увеличении нагрузки задержка может достигать ~5000 мс и более

Таким образом, увеличение числа пользователей приводит к выраженному нелинейному росту latency.

Для интерпретации причин роста задержек был проведён сравнительный анализ сценариев с активной генерацией трафика и без неё. В специальном idle-сценарии при пяти пользовательских устройствах и отсутствии нагрузки iPerf средняя задержка составляет около ~31 мс, что соответствует базовой задержке программно-определяемой сети без существенного влияния очередей и буферизации.

Однако при запуске параллельного TCP-трафика наблюдается резкое увеличение задержек до нескольких секунд. Это указывает на то, что основным источником деградации являются механизмы буферизации и накопления пакетов в очередях (bufferbloat), а не ограничения радиоинтерфейса. [17, 18]

Дополнительное влияние оказывает транспортный протокол TCP. В процессе экспериментов наблюдаются кратковременные всплески пропускной способности (TCP bursts), связанные с динамикой congestion window, а также периодические падения throughput до нулевых значений вследствие переполнения буферов и TCP timeout. В результате происходит накопление пакетов в очередях, что приводит к значительному увеличению времени ожидания и росту задержек.

На рисунке 5 представлено сравнение средних задержек (Avg Ping) и значений 95-го перцентиля (P95) для различных сценариев нагрузочного тестирования. Полученные результаты демонстрируют выраженный нелинейный рост задержек при переходе от одного пользовательского устройства к многопользовательским сценариям. Для сценариев S1–S3 с одним UE средняя задержка сохраняется на уровне около 600 мс даже при увеличении интенсивности нагрузки, что указывает на отсутствие критической перегрузки системы.

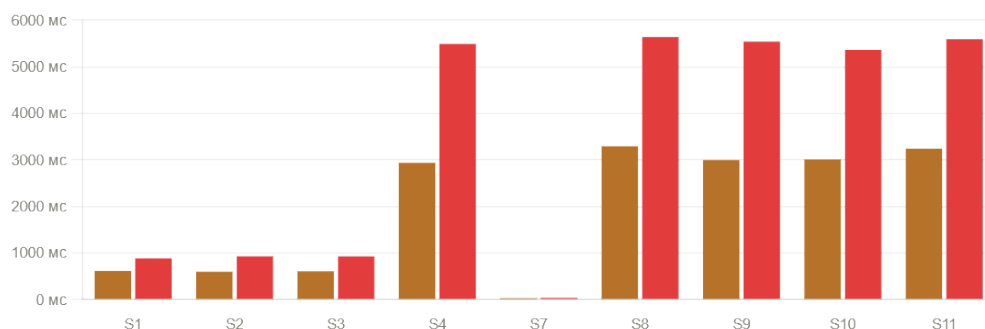


Рисунок 5 - Сравнение средних задержек (Avg Ping) и 95-го перцентиля (P95) для различных сценариев нагрузочного тестирования, демонстрирующее нелинейный рост latency при увеличении количества пользовательских устройств и эффект bufferbloat

Однако при увеличении количества одновременно работающих пользовательских устройств до пяти (сценарии S4, S8–S11) наблюдается резкий рост задержек до 2800–3200 мс, а значения 95-го перцентиля превышают 5500 мс. Подобное поведение свидетельствует о накоплении пакетов в очередях и развитии эффекта bufferbloat в условиях интенсивного TCP-трафика.

Особое значение имеет сценарий S7, в котором генерация пользовательского трафика отсутствует. В данном случае средняя задержка составляет около 31 мс, что соответствует базовой задержке программно-определяемой сети. Это подтверждает, что основной вклад в деградацию вносят не ограничения радиointерфейса, а механизмы буферизации и транспортного уровня.

Дополнительно видно, что изменение качества радиоканала (S8–S10) оказывает значительно меньшее влияние на задержку по сравнению с ростом количества параллельных TCP-потоков. Это подтверждает смещение ограничивающего фактора системы с радиосреды на вычислительные и транспортные механизмы программно-определяемой инфраструктуры.

Зависимость throughput от нагрузки

Анализ пропускной способности показывает наличие эффекта насыщения системы при увеличении количества пользователей.

Экспериментально установлено, что:

- при 1 UE throughput составляет ~23–25 Mbps
- при 5 UE суммарная пропускная способность возрастает до ~27 Mbps
- дальнейшее увеличение нагрузки не приводит к существенному росту throughput

Полученные результаты показывают, что увеличение числа пользователей действительно повышает суммарную загрузку системы, однако итоговая пропускная способность ограничивается внутренними ресурсами программно-определяемой платформы.

При этом в многопользовательских сценариях наблюдаются:

- кратковременные пики throughput
- неравномерное распределение полосы между пользователями
- периодические провалы скорости передачи данных

Подобное поведение связано с работой транспортных механизмов TCP и перегрузкой очередей в системе.

Влияние параметров радиосреды

В ходе экспериментов выполнялось изменение параметров эмулируемой радиосреды путём варьирования уровня pathloss и программной деградации канала.

Рассматривались три состояния радиосреды: good, moderate, poor.

Анализ результатов показывает, что изменение параметров радиоканала не приводит к существенному изменению throughput и не оказывает доминирующего влияния на задержку.

Даже при ухудшении условий радиосреды суммарная пропускная способность сохраняется на сопоставимом уровне, а основные деградационные эффекты проявляются в механизмах транспортного уровня и системной буферизации.

Полученные результаты позволяют сделать вывод о том, что в рамках рассматриваемой программно-определяемой тестовой среды ограничивающим фактором выступает не радиоканал, а вычислительные и транспортные механизмы системы, включая обработку пакетов, очереди и TCP-буферизацию.

3.5 Анализ узких мест системы (bottleneck)

На основе анализа данных выявлено наличие узкого места системы. Проведённые эксперименты показывают, что ограничивающим фактором исследуемой программно-определяемой среды выступает не радиоканал, а вычислительные ресурсы серверной платформы, прежде всего способность пользовательской плоскости (UPF) и сетевого стека обрабатывать интенсивный многопользовательский трафик.

При росте числа пользовательских устройств происходит насыщение вычислительных ресурсов CPU, увеличение времени обработки пакетов и накопление данных в очередях. В результате планировщик MAC-уровня фактически распределяет между пользователями не только радиоресурс, но и доступное процессорное время вычислительной платформы.

Таким образом, в условиях программной реализации происходит смещение bottleneck с радиоинтерфейса на вычислительную инфраструктуру системы.

Выявление точки деградации

Переход от 1 к 5 пользователям приводит к:

- резкому росту задержки
- появлению неустойчивости
- насыщению системы

Данный переход можно рассматривать как порог деградации.

Насыщение системы

При дальнейшем увеличении нагрузки:

- система переходит в режим перегрузки
- показатели качества ухудшаются
- рост задержки становится неконтролируемым

Аномальные режимы

В ряде сценариев наблюдаются:

- резкие скачки задержки (до десятков секунд)
- падение throughput

Такие режимы соответствуют критическим состояниям системы.

3.6 Причины деградации

Анализ экспериментальных данных показывает, что деградация показателей качества сети в рассматриваемой программно-определяемой 5G Standalone среде связана преимущественно не с ограничениями радиоинтерфейса, а с особенностями работы транспортного уровня, механизмов буферизации и обработки трафика в условиях многопользовательской нагрузки.

Bufferbloat

Одной из основных причин роста задержек является эффект переполнения сетевых буферов (bufferbloat).

При увеличении интенсивности пользовательского трафика и числа одновременно обслуживаемых UE происходит накопление пакетов в очередях сетевого стека и транспортного уровня. В результате время ожидания пакетов существенно возрастает, что приводит к резкому увеличению latency.

Экспериментальные результаты показывают, что при отсутствии активного пользовательского трафика средняя задержка составляет около ~31 мс, тогда как при запуске многопользовательского TCP-трафика задержка возрастает до нескольких секунд. Это указывает на то, что основной вклад в деградацию вносит именно буферизация и длительное нахождение пакетов в очередях.

Очереди и перегрузка системы

При увеличении нагрузки наблюдается формирование устойчивых очередей на уровне сетевого стека и пользовательской плоскости передачи данных.

Рост количества одновременно передаваемых TCP-потоков приводит к:

- увеличению времени обработки пакетов
- накоплению данных в буферах
- росту времени ожидания
- нестабильности задержек

В ряде сценариев система переходит в состояние насыщения, при котором дальнейшее увеличение нагрузки практически не приводит к росту throughput, однако существенно увеличивает latency и вариацию задержек.

Влияние TCP и retransmission

Дополнительным фактором деградации выступают особенности работы транспортного протокола TCP.

Посекундный анализ throughput показывает наличие кратковременных всплесков пропускной способности (TCP bursts), возникающих вследствие динамического изменения congestion window. Одновременно наблюдаются

периодические провалы скорости передачи данных до нулевых значений, связанные с переполнением буферов и срабатыванием TCP timeout [21].

При критических перегрузках фиксируются:

- повторные передачи пакетов (TCP retransmission)
- потеря пакетов
- нестабильность throughput
- рост задержек до нескольких секунд

Подобное поведение указывает на то, что транспортный уровень становится ключевым источником деградации системы при многопользовательской нагрузке.

Влияние статической конфигурации

Дополнительное ухудшение показателей наблюдается при использовании статической конфигурации управления (static control).

Отключение адаптивных механизмов приводит к увеличению времени нахождения пакетов в очередях и дальнейшему росту задержек. В соответствующих сценариях средняя задержка превышает ~3200 мс, а значения 95-го перцентиля достигают более ~5500 мс.

Результаты экспериментов показывают, что ключевым ограничивающим фактором исследуемой системы является не радиоканал, а механизмы транспортного уровня, системная буферизация и обработка очередей при интенсивной многопользовательской нагрузке.

Полученные результаты подтверждают наличие выраженных эффектов bufferbloat и показывают существенное влияние TCP-механизмов на показатели качества сети в программно-определяемой 5G Standalone среде.

3.7 Динамика поведения системы

Одним из ключевых результатов проведённого исследования является выявление характерной динамики перехода 5G Standalone системы из устойчивого состояния в режим деградации при увеличении нагрузки и росте числа одновременно обслуживаемых пользовательских устройств. Анализ временных рядов задержки, пропускной способности и состояния очередей показал, что деградация системы развивается не мгновенно, а представляет собой последовательный переход между несколькими фазами функционирования.

В рамках проведённых экспериментов было установлено, что даже при отсутствии выраженной деградации радиоканала система способна переходить в состояние перегрузки вследствие накопления очередей и роста задержек на транспортном уровне. Это подтверждает кросс-слойный характер деградации в программно-определяемой 5G SA среде.

Для описания наблюдаемого поведения далее выделяются четыре характерные фазы динамики системы.

Как видно на рисунке 6, в экспериментах поведение системы поддаётся формализации в виде дискретной модели состояний, выделяющей четыре режима функционирования сети в зависимости от уровня нагрузки и состояния

буферов. В стабильном режиме задержки минимальны, очереди пусты, пропускная способность соответствует номинальной; по мере роста нагрузки система последовательно переходит через режим нарастающей загрузки и деградации к режиму осцилляций, характеризующемуся нестабильными задержками и циклическим дренированием буферов. Данная модель переходов, эмпирически подтверждённая результатами экспериментальных прогонов.

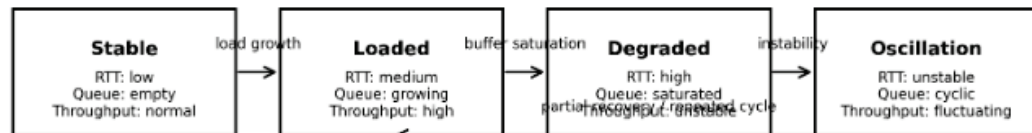


Рисунок 6 - Переходы между режимами функционирования 5G SA системы при увеличении нагрузки

Фаза 1 - стабильное состояние (stable state)

На начальном этапе функционирования система находится в устойчивом режиме работы, при котором число активных пользовательских устройств и уровень генерируемого трафика не приводят к насыщению сетевых ресурсов.

Для данной фазы характерны:

- низкие значения RTT и queueing delay;
- отсутствие переполнения буферов;
- стабильная пропускная способность;
- минимальная вариативность временных рядов;
- отсутствие выраженных oscillation effects.

В idle-сценариях минимальная наблюдаемая задержка составляла порядка 31 мс, что соответствует базовой задержке экспериментальной среды при отсутствии значимой пользовательской нагрузки. В условиях одного активного пользовательского устройства система сохраняла относительно стабильное состояние даже при увеличении интенсивности трафика, а рост задержек носил ограниченный и управляемый характер.

С точки зрения динамики временных рядов данная фаза характеризуется малой дисперсией наблюдаемых значений и отсутствием долгосрочного накопления очередей.

Фаза 2 - рост нагрузки и накопление очередей (loaded state)

При увеличении числа активных пользовательских устройств и интенсивности генерируемого трафика система начинает переходить в состояние повышенной нагрузки.

Для данной фазы характерны:

- постепенный рост RTT;
- увеличение queue occupancy;
- накопление пакетов в буферах;

- рост jitter;
- появление кратковременных локальных всплесков задержки.

На данном этапе деградация ещё не носит критического характера, однако система начинает демонстрировать признаки приближения к saturation region. Анализ временных рядов показывает появление выраженной инерционности задержки: после кратковременного роста нагрузки RTT перестаёт быстро возвращаться к исходному уровню вследствие накопления данных в очередях.

Наблюдаемое поведение соответствует эффектам bufferbloat, при которых увеличение объёма буферов временно предотвращает потери пакетов, но приводит к существенному росту задержки.

Важной особенностью является то, что на данном этапе throughput системы ещё остаётся относительно стабильным. Это означает, что деградация первоначально проявляется именно через рост задержек, а не через мгновенное снижение пропускной способности.

Переходы между режимами функционирования, показанные на рисунке 6, обусловлены единым механизмом накопления и циклического сброса очередей. Поступающий трафик последовательно вызывает заполнение буферов, рост задержки и в конечном счёте освобождение очереди; при сохраняющейся нагрузке этот цикл повторяется, порождая наблюдаемые осцилляции RTT. Концептуальная схема данного механизма представлена на рисунке 7.

Рисунок 7 - Концептуальный механизм накопления и циклического сброса очередей в перегруженной системе

Фаза 3 - насыщение системы (degraded / saturation state)

При дальнейшем увеличении нагрузки система переходит в режим насыщения, в котором ресурсы транспортного уровня и механизмы буферизации начинают выступать ограничивающим фактором производительности.

Для данной фазы характерны:

- резкий нелинейный рост RTT;
- значительное увеличение времени ожидания в очередях;
- нестабильность throughput;
- появление packet loss;
- рост числа retransmission events;
- деградация responsiveness системы.
- В multi-UE сценариях наблюдался существенный рост задержек:
- при умеренной нагрузке RTT достигал сотен миллисекунд;
- при высоком уровне загрузки задержки возрастали до 2800–3200

мс;

- в отдельных сценариях наблюдались значения порядка 5000 мс.

Полученные результаты показывают, что переход системы в degraded-state носит выраженный пороговый характер. До достижения определённого уровня нагрузки система способна поддерживать относительно стабильные показатели качества, однако после превышения критического уровня начинается резкое ухудшение состояния сети.

При этом важным результатом является то, что throughput в ряде сценариев сохранялся на сравнительно высоком уровне даже при сильной деградации задержек. Это указывает на то, что основным bottleneck-фактором становятся механизмы queueing и buffering, а не исключительно радиоканал.

Таким образом, деградация системы проявляется прежде всего через рост latency и нестабильность транспортного поведения.

Фаза 4 - осцилляции и циклическая деградация (oscillation state)

В наиболее нагруженных сценариях система переходила в режим выраженных колебательных процессов.

Для данной фазы характерны:

- периодические колебания RTT;
- циклическое накопление и сброс очередей;
- burst-поведение throughput;
- нестабильность временных рядов;
- резкие кратковременные изменения состояния системы.

Наблюдаемая динамика объясняется взаимодействием нескольких механизмов:

- TCP congestion control;
- queue draining;
- retransmission behavior;
- scheduling dynamics;
- transport-level feedback loops.

Временные ряды демонстрировали чередование фаз накопления очередей и их частичного опустошения, после чего цикл деградации повторялся вновь. Подобное поведение соответствует нелинейной динамике перегруженных сетевых систем и свидетельствует о наличии внутренних feedback-механизмов, влияющих на устойчивость функционирования.

На рисунке 8 представлены временные ряды RTT для различных уровней нагрузки и числа активных пользовательских устройств. Полученные результаты демонстрируют наличие выраженных фаз деградации, включающих стабильный режим, накопление очередей, насыщение системы и осциллирующее поведение задержек. Наблюдаемые пики RTT соответствуют моментам насыщения буферов и нарушения устойчивости транспортного уровня.

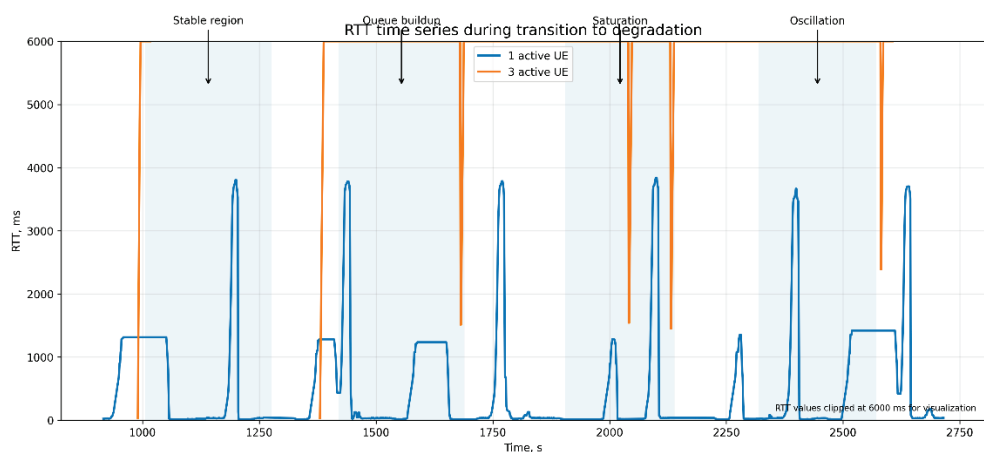


Рисунок 8 - Временной ряд RTT при переходе системы в состояние деградации

На рисунке 9 представлена тепловая карта зависимости RTT p95 от уровня нагрузки и числа активных пользовательских устройств. Полученные результаты показывают наличие выраженной нелинейной области деградации, возникающей при увеличении offered load и росте числа одновременно обслуживаемых UE.

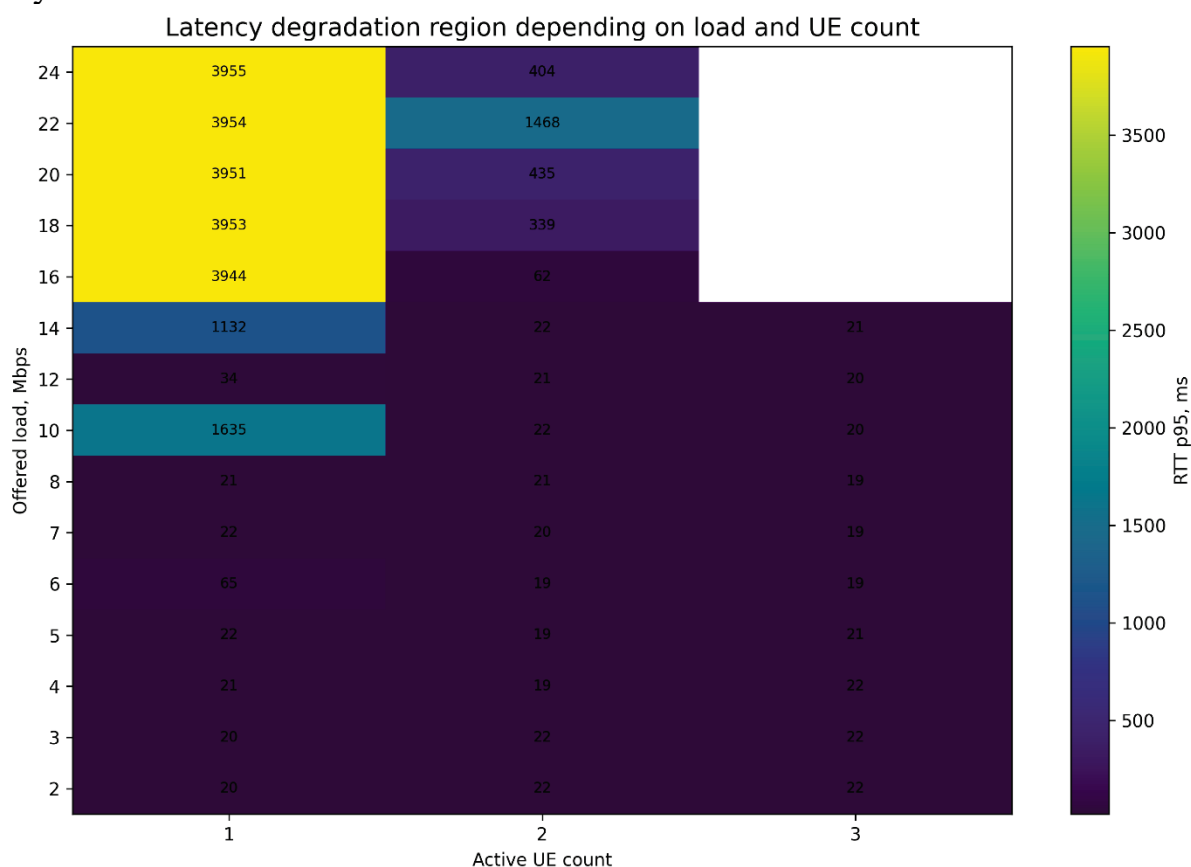


Рисунок 9 - Области деградации RTT в зависимости от offered load и числа активных UE

Совокупность полученных результатов подтверждает, что деградация производительности в исследуемой 5G SA среде носит системный и воспроизводимый характер: переход от стабильного режима к осциллирующей перегрузке детерминированно обусловлен комбинацией уровня нагрузки и числа активных пользовательских устройств. Нелинейная область деградации,

выявленная на тепловой карте RTT p95, задаёт практически значимые пороговые условия, при достижении которых стандартные механизмы TCP и буферизации утрачивают способность обеспечивать приемлемое качество обслуживания. Эти результаты формируют эмпирическую основу для перехода к следующему этапу исследования – атрибуции узких мест и построению прогностических моделей KPI.

Выводы по динамике поведения системы

Проведённый анализ показал, что деградация 5G Standalone системы развивается как последовательный переход между несколькими режимами функционирования: *stable* → *loaded* → *degraded* → *oscillation*.

Полученные результаты позволяют сделать следующие выводы:

- деградация системы носит нелинейный и пороговый характер;
- рост задержек начинается раньше существенного снижения throughput;
- ключевую роль в деградации играют механизмы буферизации и queueing behavior;
- транспортный уровень способен становиться основным bottleneck системы;
- временные ряды KPI обладают выраженной динамической структурой и инерционностью;
- состояние системы демонстрирует частичную краткосрочную предсказуемость.

Последний вывод является особенно важным, поскольку наличие устойчивой временной структуры в поведении KPI создаёт предпосылки для применения методов прогнозирования состояния сети, что рассматривается в следующей главе диссертационной работы.

3.8 Модель состояний системы

На основе проведённого анализа временных рядов KPI, поведения задержек, динамики пропускной способности и механизмов деградации была сформирована модель состояний 5G Standalone системы, позволяющая описывать функционирование сети на уровне дискретных режимов работы.

Предлагаемая модель рассматривает сеть как динамическую систему, состояние которой изменяется под воздействием нагрузки, числа активных пользовательских устройств и механизмов буферизации. В отличие от анализа отдельных мгновенных значений KPI, модель состояний позволяет интерпретировать поведение сети как последовательность переходов между устойчивыми и деградирующими режимами функционирования.

В рамках исследования выделены четыре основных состояния системы.

3.8.1 Состояние Normal

Состояние Normal соответствует устойчивому режиму функционирования сети, при котором система обслуживает пользовательский трафик без признаков перегрузки.

Для данного состояния характерны:

- низкие значения RTT;
- стабильная пропускная способность;
- отсутствие выраженного queue buildup;
- минимальная вариативность временных рядов;
- отсутствие packet loss и retransmission spikes.

В данном режиме сеть сохраняет устойчивость даже при кратковременных изменениях нагрузки, а транспортные механизмы функционируют в штатном режиме.

С точки зрения временной динамики состояние Normal характеризуется высокой предсказуемостью поведения KPI и отсутствием длительных эффектов накопления очередей.

3.8.2 Состояние Loaded

Состояние Loaded соответствует режиму повышенной нагрузки, при котором система ещё сохраняет работоспособность, однако начинает демонстрировать признаки приближения к saturation region.

Для данного состояния характерны:

- рост RTT;
- увеличение queue occupancy;
- накопление пакетов в буферах;
- увеличение jitter;
- появление локальных всплесков задержки.

На данном этапе throughput системы может оставаться сравнительно стабильным, однако транспортный уровень начинает работать в условиях повышенной нагрузки.

Именно в состоянии Loaded начинают проявляться bufferbloat-эффекты, связанные с накоплением пакетов в очередях и ростом времени ожидания обработки данных.

Как показано на рисунке 10, увеличение offered load приводит к постепенному росту задержек и переходу системы из устойчивого режима функционирования в состояние повышенной нагрузки. При этом throughput остаётся относительно стабильным, тогда как RTT начинает демонстрировать признаки инерционного роста.

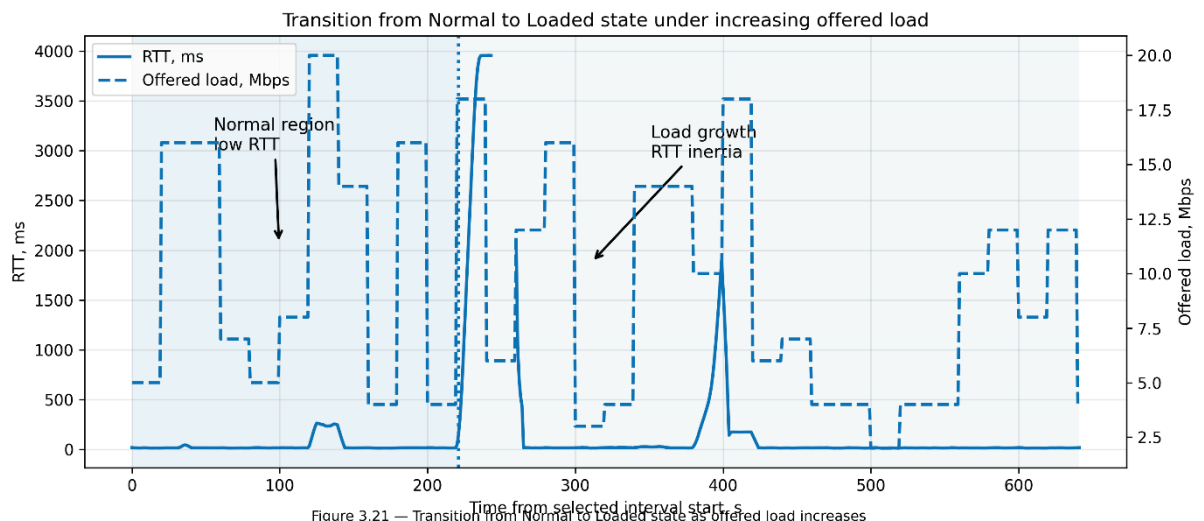


Figure 3.21 — Transition from Normal to Loaded state as offered load increases

Рисунок 10 - Переход системы из состояния Normal в состояние Loaded при увеличении offered load

3.8.3 Состояние Degraded

Состояние Degraded соответствует режиму выраженной деградации системы, возникающему при насыщении транспортных механизмов и буферов.

Для данного состояния характерны:

- высокий RTT;
- нестабильность throughput;
- saturation queue behavior;
- packet loss;
- retransmission spikes;
- рост времени ожидания в очередях.

В ходе экспериментов было установлено, что переход в состояние Degraded носит выраженный пороговый характер. До достижения критического уровня нагрузки система сохраняет относительную устойчивость, однако после насыщения буферов происходит резкий нелинейный рост задержек.

При этом деградация системы проявляется прежде всего через ухудшение latency-характеристик, а не через мгновенное снижение throughput.

Полученные результаты указывают на доминирующее влияние queueing и buffering mechanisms в процессе деградации системы.

На рисунке 11 показано, что рост RTT начинается раньше существенного снижения throughput. Это указывает на то, что деградация системы первоначально формируется на уровне queue buildup и buffering effects, а снижение пропускной способности возникает на более поздних стадиях перегрузки.

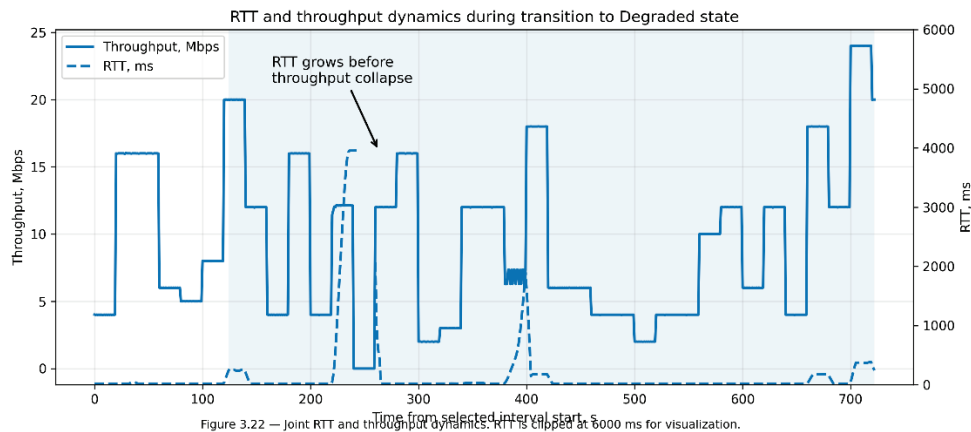


Рисунок 11 - Динамика RTT и throughput при переходе системы в состояние Degraded

3.8.4 Состояние Breakdown

Состояние Breakdown соответствует критическому режиму функционирования системы, при котором сеть теряет устойчивость и начинает демонстрировать выраженное oscillation behavior.

Для данного состояния характерны:

- экстремально высокие значения RTT;
- циклические колебания задержек;
- нестабильность throughput;
- периодические queue release events;
- высокая вариативность KPI;
- потеря устойчивости временных рядов.

В данном режиме транспортный уровень функционирует в условиях постоянной перегрузки, а система начинает переходить между кратковременными фазами частичного восстановления и повторной деградации.

Поведение соответствует нелинейной динамике перегруженных распределённых систем с выраженными feedback-механизмами.

Как показано на рисунке 12, в состоянии Breakdown система начинает демонстрировать циклическое поведение задержек, обусловленное периодическим накоплением и частичным сбросом очередей. Подобное поведение указывает на потерю устойчивости транспортного уровня и переход системы в режим нестабильного функционирования.

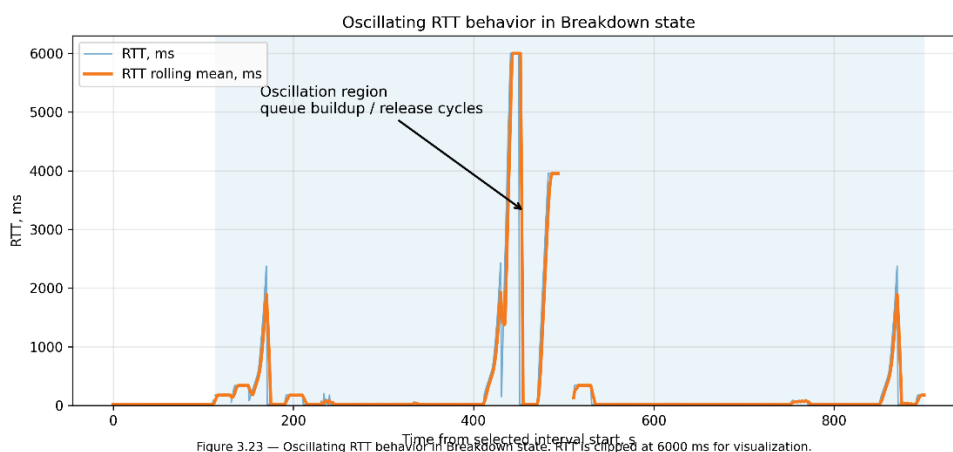


Figure 3.23 — Oscillating RTT behavior in Breakdown state. RTT is clipped at 6000 ms for visualization.

Рисунок 12 - Осциллирующее поведение RTT в состоянии Breakdown

3.8.5 Переходы между состояниями

Переходы между состояниями системы определяются совокупным влиянием нескольких факторов:

- числа активных пользовательских устройств;
- интенсивности offered load;
- состояния радиосреды;
- уровня буферизации;
- транспортных механизмов обработки трафика.

В рамках проведённого исследования установлено, что ключевым фактором деградации являются механизмы queue buildup и saturation behavior транспортного уровня.

При этом возможны кратковременные частичные возвраты системы в менее деградированное состояние вследствие временного уменьшения нагрузки или частичного освобождения очередей.

На рисунке 13 представлена итоговая модель состояний исследуемой системы. Предлагаемая модель позволяет интерпретировать поведение 5G SA сети на уровне дискретных режимов функционирования и описывать процессы деградации как последовательность переходов между состояниями системы.

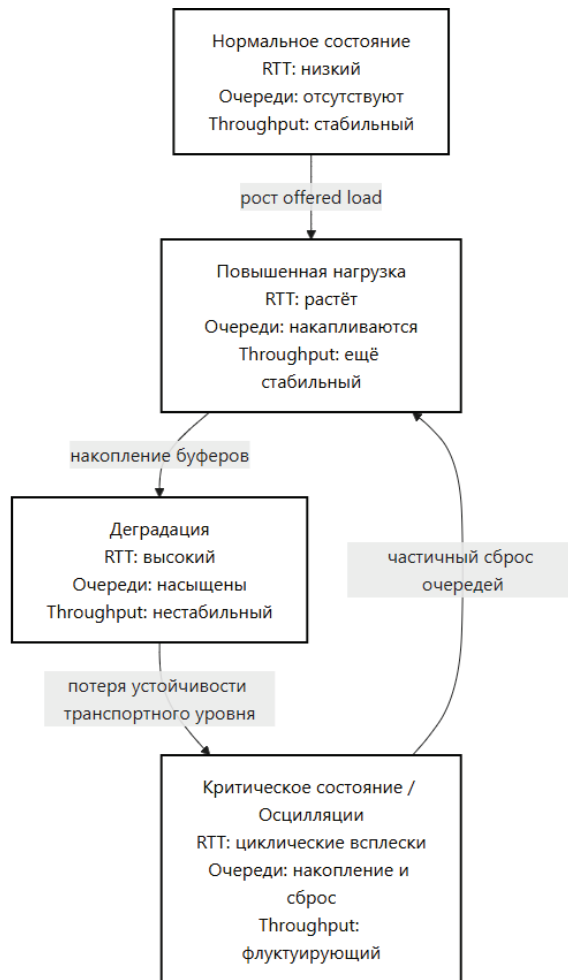


Рисунок 13 - Модель переходов между состояниями 5G SA системы

3.9 Выводы по главе 3

В третьей главе выполнен анализ экспериментальных данных, полученных в контролируемой 5G Standalone среде, и исследовано поведение показателей качества сети при изменении нагрузки, числа активных пользовательских устройств и условий функционирования системы. Основное внимание было уделено не отдельным мгновенным значениям KPI, а динамике перехода системы между различными режимами работы.

Проведённый анализ показал, что поведение 5G SA системы в условиях многопользовательской нагрузки имеет выраженный нелинейный характер. При малых и умеренных значениях нагрузки система сохраняет устойчивость: значения задержки остаются низкими или умеренными, пропускная способность стабильна, а признаки накопления очередей выражены слабо. Однако при дальнейшем увеличении нагрузки наблюдается пороговый переход, при котором незначительное изменение входных условий приводит к резкому ухудшению показателей качества. Таким образом, зависимость между параметрами нагрузки и состоянием сети не может быть описана как простая линейная функция.

Отдельно установлено, что одним из наиболее чувствительных индикаторов деградации является задержка RTT. В отличие от throughput, который в ряде сценариев сохраняет сравнительно высокие значения даже при ухудшении состояния сети, задержка начинает расти раньше и отражает

начальные признаки накопления очередей. Это позволяет рассматривать latency-метрики, включая RTT и RTT p95, как ранние признаки перехода системы из устойчивого состояния в нагруженное или деградирующее состояние.

Анализ временных рядов показал, что деградация системы развивается не мгновенно, а проходит несколько характерных фаз: стабильное состояние, рост нагрузки, насыщение и осциллирующее поведение. Такая динамика соответствует последовательности: Normal→Loaded→Degraded→Breakdown.

или в более общем виде:

stable → loaded → degraded → oscillation

Данная последовательность показывает, что ухудшение качества сети является переходным процессом, а не единичным случайным событием. Это особенно важно для последующего прогнозирования, поскольку наличие устойчивой временной структуры означает, что признаки будущей деградации могут появляться до наступления критического состояния.

В ходе анализа также установлено, что основным механизмом деградации в рассматриваемых условиях выступают процессы буферизации и накопления очередей. При увеличении offered load система первоначально пытается сохранять пропускную способность за счёт накопления пакетов в буферах, однако это приводит к росту RTT, увеличению jitter и появлению осцилляций. В критических режимах наблюдается циклическое поведение: накопление очередей сменяется их частичным сбросом, после чего процесс повторяется. Это подтверждает, что деградация имеет кросс-слойный характер и не может быть объяснена только параметрами радиоканала.

Сравнение сценариев с различными условиями радиосреды показало, что в рассматриваемой экспериментальной конфигурации радиоканал не является основным ограничивающим фактором. Даже при изменении условий радиосреды пропускная способность в ряде случаев сохранялась на сопоставимом уровне, тогда как задержка и устойчивость временных рядов ухудшались преимущественно вследствие queue buildup и транспортных эффектов. Следовательно, bottleneck системы в данных экспериментах связан прежде всего с механизмами буферизации, TCP-поведением и обработкой трафика на транспортном уровне.

На основе анализа поведения KPI была предложена модель состояний системы, включающая четыре режима: Normal, Loaded, Degraded и Breakdown. Состояние Normal соответствует устойчивой работе сети; Loaded - режиму повышенной нагрузки и начального накопления очередей; Degraded - выраженному ухудшению показателей качества; Breakdown - критическому состоянию с осцилляциями, всплесками RTT и нестабильным throughput. Такая модель позволяет перейти от анализа отдельных метрик к описанию состояния сети как дискретного режима функционирования. В текущем тексте уже зафиксирована модель переходов между этими состояниями и показано, что частичный сброс очередей может возвращать систему из критического режима в нагруженное состояние, но не обязательно восстанавливает её до Normal.

Сформированная модель состояний имеет методологическое значение для всей дальнейшей работы. Она позволяет интерпретировать деградацию не как

набор разрозненных отклонений в RTT, throughput или packet loss, а как структурированный переход между режимами функционирования системы. Тем самым создаётся основа для классификации состояния сети, построения интегрального показателя качества и последующего краткосрочного прогнозирования.

В результате анализа также обоснована необходимость использования интегрального агрегированного показателя состояния сети. Отдельные KPI отражают разные стороны поведения системы: throughput характеризует способность передавать данные, RTT и RTT p95 отражают задержку и накопление очередей, packet loss фиксирует более поздние признаки нарушения устойчивости. Поэтому оценка состояния сети только по одному показателю является недостаточной. Интегральный показатель позволяет объединить несколько характеристик качества в единую шкалу и использовать её для анализа, классификации и прогнозирования состояния сети. Данная логика соответствует общей постановке работы, где состояние 5G SA сети оценивается не по одному KPI, а через агрегированное представление нескольких характеристик.

Таким образом, в главе 3 получены следующие основные результаты:

1. Показано, что показатели качества 5G SA сети зависят от числа активных пользовательских устройств, уровня offered load и режима функционирования системы.
2. Установлено, что зависимость между нагрузкой и качеством сети носит нелинейный характер и включает пороговые эффекты деградации.
3. Выявлено, что рост RTT и RTT p95 является более ранним признаком деградации, чем снижение throughput. Показано, что в рассматриваемых условиях основным bottleneck-фактором являются механизмы буферизации, накопления очередей и транспортного уровня, а не только радиоканал.

=Установлено, что деградация системы развивается как переходный процесс, включающий фазы устойчивого состояния, роста нагрузки, насыщения и осцилляций.

Предложена модель состояний 5G SA системы: Normal, Loaded, Degraded и Breakdown, описывающая переходы между режимами функционирования сети.

12. Обоснована необходимость перехода от анализа отдельных KPI к интегральному агрегированному показателю состояния сети.

13. Показано, что временные ряды KPI обладают выраженной динамической структурой, что создаёт предпосылки для краткосрочного прогнозирования будущего состояния системы.

Полученные в главе 3 результаты формируют основу для следующего этапа исследования. Если в данной главе было показано, как система переходит между состояниями и какие признаки сопровождают деградацию, то в следующей главе рассматривается вопрос о том, насколько эти переходы могут быть предсказаны заранее на основе экспериментальных данных и методов машинного обучения.

4 ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ КАЧЕСТВА

В предыдущей главе было показано, что поведение 5G Standalone системы носит динамический и нелинейный характер, а деградация сети развивается как последовательный переход между различными режимами функционирования. Анализ временных рядов KPI показал наличие выраженных фазовых переходов, эффектов накопления очередей, saturation behavior и oscillation dynamics. Задача прогнозирования сетевых KPI активно исследуется в современной литературе [37, 38].

Полученные результаты позволяют сформулировать задачу прогнозирования состояния сети как задачу предсказания будущих значений показателей качества на основе текущего состояния системы и предшествующей динамики наблюдений.

4.1 Постановка задачи прогнозирования

В рамках настоящей работы прогнозирование рассматривается не как абстрактная задача временных рядов, а как задача оценки будущего состояния программно-определяемой 5G SA системы в условиях изменяющейся нагрузки и возможного перехода в режим деградации. [37, 38]

В отличие от традиционного подхода, в котором прогнозирование часто выполняется только на основе отдельных KPI, в настоящей работе учитываются выявленные в главе 3 закономерности поведения системы:

- зависимость состояния сети от числа активных пользовательских устройств;
- нелинейный характер роста RTT;
- наличие пороговых эффектов деградации;
- влияние queue buildup и buffering effects;
- различие между устойчивыми и деградирующими режимами функционирования;
- наличие временной структуры переходов между состояниями системы.

Таким образом, задача прогнозирования формулируется как задача оценки будущего состояния сети на основе наблюдаемого состояния системы в текущий момент времени.

Формально задача прогнозирования может быть представлена следующим образом:

$$Y(t+\Delta t)=F(X(t),X(t-1),...,Y(t),Y(t-1),...)$$

где:

- (X) - набор входных параметров системы;
- (Y) - наблюдаемые показатели качества;
- (Δt) - горизонт прогнозирования;
- (F) - функция отображения состояния системы.

В качестве прогнозируемых показателей в работе рассматриваются:

- RTT;
- RTT p95;

- throughput;
- интегральный показатель состояния системы (state score);
- класс состояния сети (stable / loaded / degraded / critical).

Как показано на рисунке 14, задача прогнозирования рассматривается как отображение текущего состояния сети в прогнозируемое будущее состояние. При этом модель должна учитывать не только текущие значения KPI, но и динамику изменения состояния системы во времени.

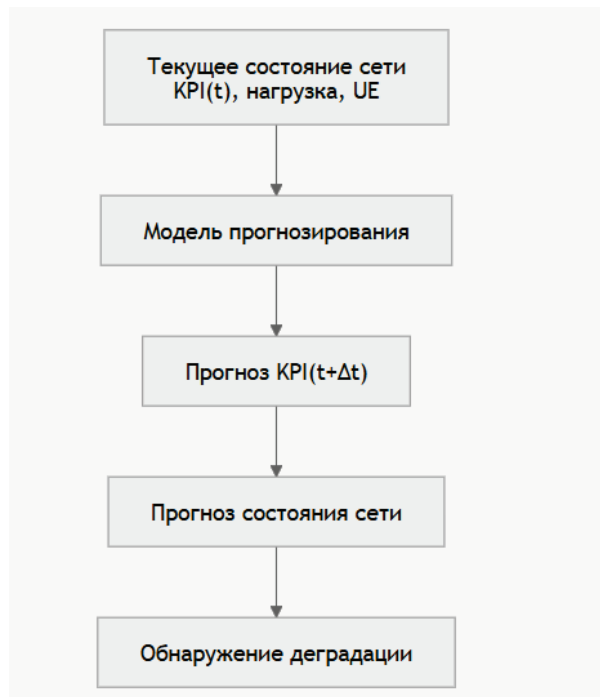


Рисунок 14 - Общая схема прогнозирования состояния сети на основе наблюдаемого состояния системы

Особенностью рассматриваемой задачи является необходимость прогнозирования поведения системы вблизи областей деградации. В таких режимах поведение сети становится существенно менее устойчивым, а зависимость между параметрами нагрузки и KPI приобретает выраженный нелинейный характер.

Дополнительную сложность создаёт наличие oscillation behavior и queue buildup dynamics, при которых система может быстро переходить между различными режимами функционирования.

В рамках работы рассматриваются горизонты прогнозирования:

$$\Delta t \in \{1, 2, 3, 5, 10, 15, 20\} \text{ секунд}$$

Подобный диапазон горизонтов позволяет исследовать:

- краткосрочное прогнозирование;
- устойчивость моделей при увеличении горизонта;
- скорость деградации качества прогноза;
- предсказуемость различных KPI.

В работе рассматриваются две принципиально различные постановки задачи прогнозирования.

Первая постановка предполагает использование только текущего наблюдаемого состояния системы без явного использования lag-признаков и rolling statistics.

Вторая постановка использует временную память в виде lag-based и rolling-based признаков, позволяющих модели учитывать предшествующую динамику временных рядов.

Подобное разделение позволяет исследовать, насколько текущее состояние сети само по себе содержит информацию о будущем переходе системы в деградацию.

В отличие от операторских production-сетей, используемая экспериментальная среда позволяет выполнять прогнозирование в условиях контролируемых и воспроизводимых сценариев. Это существенно снижает влияние скрытых внешних факторов и делает возможным исследование причинно-следственных зависимостей между состоянием системы и будущими значениями KPI.

Таким образом, задача прогнозирования в настоящей работе рассматривается как задача оценки будущего состояния 5G SA системы на основе наблюдаемой динамики KPI и параметров нагрузки в контролируемой экспериментальной среде.

4.2 Подготовка данных

Для решения задачи прогнозирования была выполнена подготовка экспериментальных данных, полученных в разработанной 5G SA testbed-среде.

Основной целью подготовки данных являлось формирование согласованного набора временных рядов, пригодных для последующего применения методов машинного обучения и анализа динамики состояния системы.

Исходные данные были сформированы на основе результатов нагрузочных экспериментов, описанных в предыдущих главах, и включали:

- показатели RTT;
- RTT p95;
- throughput;
- packet loss;
- jitter;
- параметры offered load;
- число активных пользовательских устройств;
- параметры радиосреды;
- признаки состояния системы.

Экспериментальные данные представляли собой последовательности наблюдений, фиксируемых через равные интервалы времени.

4.2.1 Формирование временных рядов

На первом этапе подготовки данных исходные журналы наблюдений были преобразованы в структурированные временные ряды.

Каждая запись соответствовала фиксированному моменту времени и содержала:

- параметры состояния сети;
- значения KPI;
- конфигурацию сценария;
- параметры нагрузки;
- признаки активности пользовательских устройств.

Для каждого сценария формировались последовательности вида:

latency(t)
throughput(t)
packet loss(t)
state score(t)

Подобное представление позволяет интерпретировать сеть как динамическую систему, состояние которой изменяется во времени.

Как показано на рисунке 15, подготовка данных включает последовательное преобразование экспериментальных журналов в структурированные временные ряды и последующее формирование обучающих выборок.

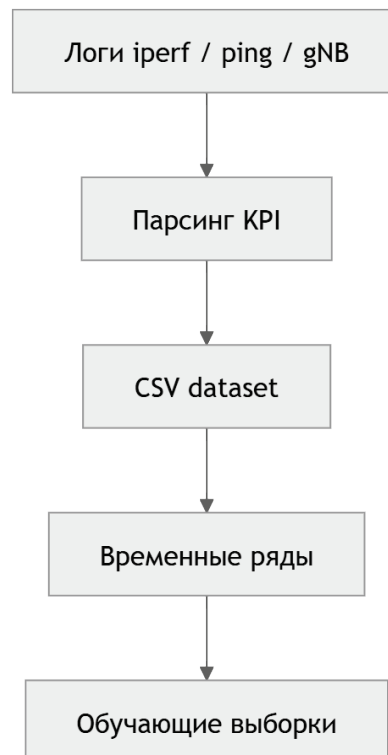


Рисунок 15 - Преобразование экспериментальных журналов в временные ряды для прогнозирования

В работе использовались данные с временным разрешением порядка одной секунды, что позволило фиксировать быстрые изменения состояния системы и переходы между режимами деградации.

Особое внимание уделялось сохранению временной структуры данных. В отличие от классических задач табличного машинного обучения, временные

ряды обладают зависимостью между соседними наблюдениями, поэтому нарушение временного порядка может приводить к data leakage и переоценке качества моделей.

4.2.2 Нормализация данных

Поскольку используемые параметры имеют различную физическую природу и различную размерность, перед обучением моделей выполнялась нормализация признаков.

В частности:

- throughput измеряется в Mbps;
- RTT - в миллисекундах;
- packet loss - в процентах;
- state score - в нормированном диапазоне.

Для обеспечения сопоставимости признаков использовались процедуры масштабирования и стандартизации данных.

В рамках работы применялись:

- StandardScaler;
- MinMaxScaler.

Использование нормализации позволяет [28, 47]:

- повысить устойчивость обучения;
- уменьшить влияние масштаба признаков;
- ускорить сходимость моделей;
- снизить влияние выбросов.

Перед обучением моделей машинного обучения выполнялась нормализация признаков. Это необходимо, поскольку различные показатели качества сети имеют существенно отличающиеся диапазоны значений и физическую размерность.



Рисунок 16 - Приведение признаков к сопоставимому масштабу перед обучением моделей

Как показано на рисунке 16, до нормализации отдельные признаки доминируют по масштабу значений, что может приводить к нестабильности обучения моделей. После применения StandardScaler признаки приводятся к сопоставимому диапазону без потери структуры временных зависимостей.

4.2.3 Формирование обучающих выборок

После формирования структурированного временного dataset следующим этапом является преобразование последовательности наблюдений в обучающие и тестовые выборки для моделей прогнозирования. В отличие от задач статической классификации, где порядок строк может быть произвольно перемешан, в задаче прогнозирования состояния сети временная структура данных имеет принципиальное значение.

Каждая строка исходного dataset соответствует состоянию 5G SA системы в некоторый момент времени (t). В строке фиксируются параметры текущего состояния сети: уровень offered load, число активных UE, направление трафика, protocol, pathloss, throughput, RTT, RTT p95, jitter, packet loss, broken UE ratio, а также производные признаки состояния системы. В скрипте формирования dataset каждая фаза эксперимента задаётся параметрами нагрузки, протокола, направления передачи, pathloss и активности UE, после чего данные ping и iperf агрегируются в посекундные записи.

Для построения задачи прогнозирования каждая запись преобразуется в пару «вход - целевое значение». Входной вектор ($X(t)$) описывает наблюдаемое состояние системы в текущий момент времени, а целевое значение ($Y(t+\Delta t)$) соответствует будущему значению прогнозируемого показателя через заданный горизонт времени.

Формально обучающий пример имеет вид:

$$X(t) \rightarrow Y(t + \Delta t)$$

где:

$$\Delta t \in \{1, 2, 3, 5, 10, 15, 20\}$$

В качестве входных признаков использовались следующие группы параметров:

Таблица 13 - Группы признаков, используемых для прогнозирования

Группа признаков	Примеры признаков	Назначение
Параметры нагрузки	offeredloadmbps, parallel_streams	описание интенсивности генерируемого трафика
Параметры UE	ue1active, ue2active, ue3active, activeue_count	описание числа и состава активных пользователей

Группа признаков	Примеры признаков	Назначение
Параметры канала	pathloss	описание условий радиосреды
Параметры трафика	protocol, direction, traffic_pattern	описание характера передачи данных
Текущие KPI	RTT, RTT p95, throughput, jitter, packet loss	описание текущего состояния сети
Признаки отказа измерений	brokenueratio, measurement_broken	фиксация неполных или нарушенных измерений

Целевыми переменными в задаче прогнозирования выступали:

- будущий RTT;
- будущий RTT p95;
- будущий throughput;
- будущий state_score;
- будущий класс состояния сети.

Для каждого горизонта прогнозирования формировалась отдельная целевая переменная:

$$\text{target}(t + \Delta t)$$

Например, для горизонта 5 секунд модель обучалась предсказывать состояние системы в момент (t+5) на основе информации, доступной в момент (t).



Рисунок 17 - Формирование обучающих пар для задачи прогнозирования

Как показано на рисунке 4.5, каждая обучающая запись формируется как соответствие между текущим состоянием сети и будущим значением прогнозируемого показателя. Такой подход позволяет рассматривать задачу прогнозирования как задачу отображения текущего состояния системы в её будущее состояние.

Особое внимание при формировании выборок уделялось исключению утечки данных. Под утечкой данных в данной работе понимается ситуация, при которой модель получает во входных признаках информацию, прямо или косвенно связанную с будущим целевым значением. Это может привести к завышенной оценке качества модели и сделать результаты непригодными для реальной задачи прогнозирования.

Поэтому из входных признаков исключались:

- future targets;
- phase_id;
- phase_second;
- global_second;
- признаки, напрямую кодирующие положение внутри сценария;
- будущие значения KPI;
- целевые переменные, сформированные с помощью сдвига во времени.

В no-leakage варианте эксперимента модель использует только признаки, которые доступны в текущий момент времени. Такой подход позволяет оценивать реальную способность модели предсказывать будущее состояние сети, а не запоминать структуру экспериментального сценария. В загруженном скрипте прогнозирования явно указано, что из признаков исключаются временные счётчики, phaseid, future targets, lag/rolling признаки, а также statescore и state_class как входные признаки для соответствующих задач.

Разделение данных на обучающую и тестовую части выполнялось с сохранением временного порядка. Это означает, что первая часть временного ряда использовалась для обучения модели, а последующая часть - для проверки качества прогноза. Такой подход имитирует реальную ситуацию, в которой модель обучается на ранее наблюдавшихся данных и затем применяется к будущим состояниям сети.

Схематично это можно представить следующим образом:

$$\underbrace{t_1, t_2, \dots, t_k}_{\text{train}} \rightarrow \underbrace{t_{k+1}, \dots, t_n}_{\text{test}}$$

При этом случайное перемешивание строк не использовалось, поскольку оно нарушило бы временную структуру данных и могло бы привести к попаданию будущих состояний сети в обучающую выборку.

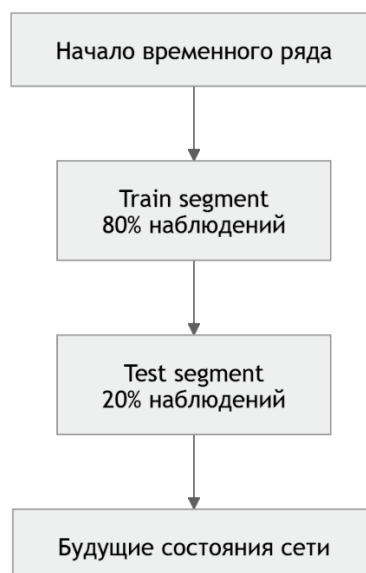


Рисунок 18 - Temporal split обучающей и тестовой выборки без перемешивания наблюдений

Использование temporal split позволяет сохранить причинную структуру задачи прогнозирования: модель обучается только на прошлых наблюдениях и проверяется на последующих состояниях сети. Это делает оценку качества более строгой по сравнению со случайным разбиением данных.

В результате подготовки данных для каждого горизонта прогнозирования формировался отдельный набор обучающих примеров. Это позволило оценить, как изменяется качество прогноза при увеличении горизонта от 1 до 20 секунд. Такой подход важен для определения рабочего горизонта прогнозирования, в пределах которого состояние сети остаётся достаточно предсказуемым для практического применения.

Итоговая логика формирования обучающей выборки может быть описана следующим образом:



Рисунок 19 - Логика формирования обучающей выборки

Таким образом, формирование обучающих выборок обеспечивает переход от экспериментальных измерений к корректной задаче машинного обучения, в которой модель должна предсказывать будущее состояние 5G SA системы только на основе информации, доступной в текущий момент времени.

4.3 Модели прогнозирования

Для прогнозирования состояния сети и будущих значений показателей качества в работе использовались методы машинного обучения и анализа временных рядов. Основной целью являлось исследование возможности краткосрочного прогнозирования деградации 5G SA системы на основе текущего состояния сети и наблюдаемой динамики KPI.

В отличие от классических задач анализа временных рядов, рассматриваемая задача характеризуется высокой нелинейностью, наличием резких переходов между состояниями системы и выраженными фазами деградации. Как было показано в предыдущей главе, поведение сети включает области стабильной работы, накопления очередей, saturation behavior и oscillation dynamics. Поэтому задача прогнозирования рассматривалась не только как задача оценки будущих значений отдельных KPI, но и как задача прогнозирования перехода системы в режим деградации.

В рамках работы исследовались два основных направления прогнозирования:

- прогнозирование будущих значений KPI;
- прогнозирование перехода системы в деградирующее состояние.

В качестве прогнозируемых показателей рассматривались:

- RTT;

- RTT p95;
 - throughput;
 - state score;
 - класс состояния сети.
-

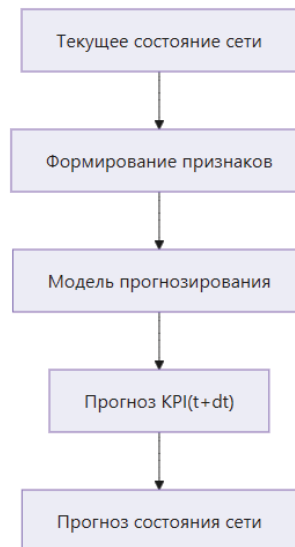


Рисунок 20 - Общая схема прогнозирования будущего состояния 5G SA системы

Для решения задачи прогнозирования использовались как классические ансамблевые модели машинного обучения, так и последовательные нейросетевые архитектуры. В частности, исследовались:

- Ridge Regression [25];
- Random Forest [24];
- Extra Trees [30];
- HistGradientBoosting [27, 44, 31];
- Logistic Regression [26, 46];
- GRU [35, 36];
- LSTM;
- CNN-GRU [32].

Классические ансамблевые методы [33] использовались для анализа табличных признаков и оценки зависимости между текущим состоянием сети и будущими значениями KPI. Последовательные нейросетевые модели [41] применялись для исследования временной структуры деградации и накопительных эффектов.

Особое внимание уделялось предотвращению data leakage. Входные признаки формировались только из параметров, доступных в текущий момент времени. Из набора признаков исключались:

- future targets;
- phase identifiers;
- временные счётчики сценариев;
- будущие значения KPI;

- признаки, напрямую кодирующие положение внутри сценария.

Это позволяло исследовать реальную способность модели прогнозировать будущее состояние системы, а не запоминать структуру экспериментальных сценариев. В используемых no-leakage pipeline явно исключались future target columns, lag-признаки будущих наблюдений и phase identifiers.

В качестве входных признаков использовались:

- offeredloadmbps;
- activeuecount;
- protocol;
- direction;
- traffic_pattern;
- pathloss;
- throughput;
- RTT;
- RTT p95;
- jitter;
- packet loss;
- state score;
- measurement flags.

Дополнительно формировались lag-based и rolling-based признаки, отражающие временную динамику системы:

$$x(t-1), x(t-2), x(t-5)$$

$$\text{rollmean}(x, w), \text{rollstd}(x, w)$$

Rolling-окна вычислялись только по предыдущим наблюдениям:

$$x_{\text{roll}}(t) = f(x(t-1), x(t-2), \dots)$$

что исключало использование будущей информации при обучении моделей.

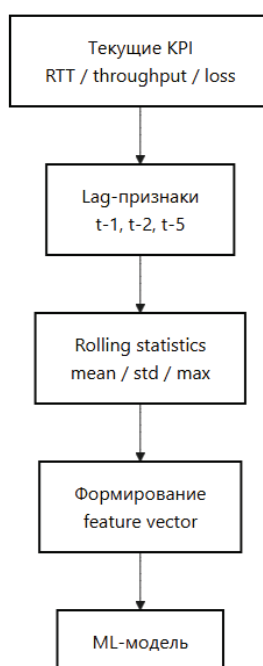


Рисунок 21 - Схема формирования признаков для моделей прогнозирования

Одной из основных задач являлось прогнозирование будущих значений RTT. Для уменьшения влияния экстремальных всплесков задержки использовалось логарифмическое преобразование целевой переменной:

$$y = \log(1 + \text{RTT})$$

Подобный подход позволял уменьшить влияние редких аномально больших значений RTT и повысить устойчивость обучения моделей. В используемых regression pipeline прогнозирование выполнялось для горизонтов от 1 до 20 секунд.

Для исследования временной структуры деградации применялись последовательностные модели GRU и LSTM [34]. Входом модели являлась последовательность предыдущих наблюдений:

$$X = [x(t - n), \dots, x(t - 1), x(t)]$$

где длина последовательности составляла:

- 30 секунд;
- 60 секунд;
- 120 секунд.

Использование последовательностных архитектур позволяло учитывать накопительный характер деградации, временную инерцию процессов буферизации и oscillation behavior системы.

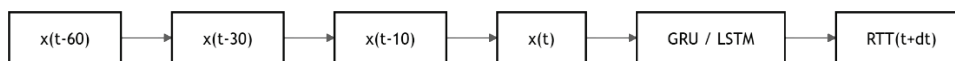


Рисунок 22 - Формирование последовательностей для GRU/LSTM моделей

Помимо задачи регрессии RTT рассматривалась задача раннего предупреждения деградации сети. Формировалась бинарная целевая переменная:

$$Y = \begin{cases} 1, & \text{если через } \Delta t \text{ сеть перейдет в degraded/critical} \\ 0, & \text{иначе} \end{cases}$$

Таким образом модель должна была определить вероятность перехода системы в деградирующее состояние до фактического наступления критического режима. В используемом early warning pipeline обучение выполнялось только на наблюдениях, в которых текущее состояние сети ещё оставалось stable или loaded.

Для оценки качества прогнозирования использовались следующие метрики:

Для задач регрессии:

- MAE;
- RMSE;
- (R^2) .

Для задач классификации [29, 45]:

- Accuracy;
- Precision;
- Recall;
- F1-score;
- ROC-AUC;
- PR-AUC.

Разделение данных на обучающую и тестовую выборки выполнялось с сохранением временной структуры:

train→past, test→future

Случайное перемешивание наблюдений не использовалось, поскольку это нарушало бы причинную структуру задачи прогнозирования и приводило бы к попаданию будущих состояний системы в обучающую выборку.

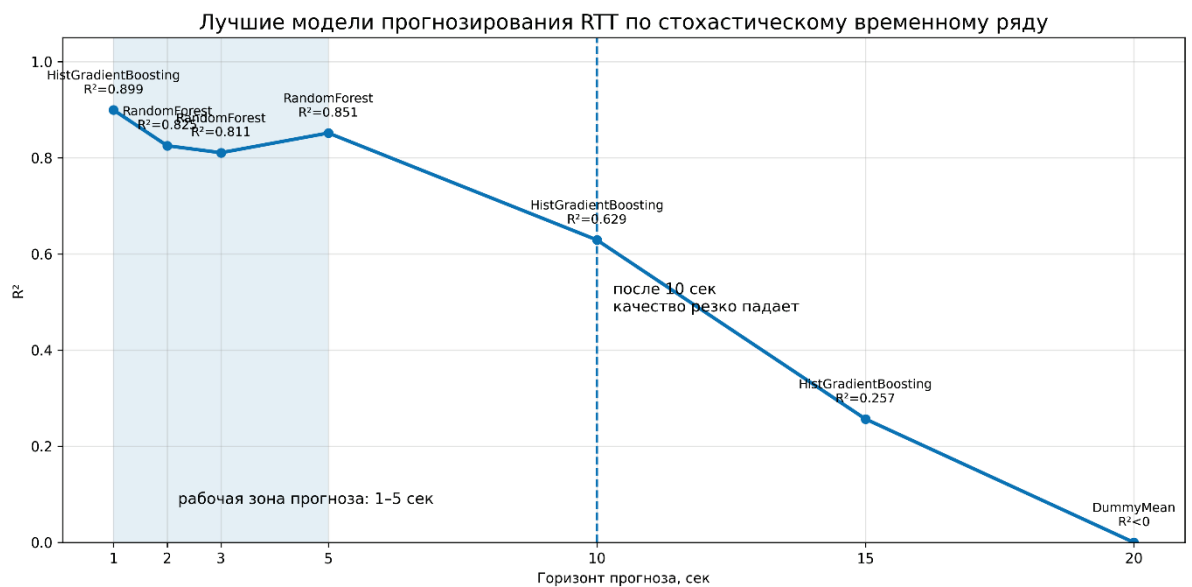


Рисунок 23 - Изменение качества прогнозирования RTT при увеличении горизонта прогноза

Полученные модели позволяют исследовать не только точность прогнозирования отдельных KPI, но и предсказуемость переходов между состояниями системы. Это особенно важно для задач раннего предупреждения деградации и потенциального использования прогнозирования в механизмах адаптивного управления сетью.

4.4 Результаты прогнозирования по фазово-структурированным данным

Для первоначального исследования прогнозируемости состояния 5G SA системы использовались фазово-структурированные экспериментальные данные, в которых изменение нагрузки и параметров сети происходило в контролируемой последовательности сценариев.

В отличие от полностью стохастического режима, фазово-структурированные эксперименты характеризуются управляемыми переходами

между состояниями системы. Это позволяет исследовать предсказуемость деградации в условиях, когда динамика изменения нагрузки имеет выраженную структуру и повторяемость.

Основной целью данного этапа являлось:

- исследование возможности краткосрочного прогнозирования KPI;
- оценка устойчивости моделей при контролируемых переходах

между состояниями сети;

- определение рабочих горизонтов прогнозирования;
- анализ влияния деградации на качество прогноза.

В рамках экспериментов прогнозировались:

- RTT;
- RTT p95;
- throughput;
- state score;
- класс состояния системы.
- Прогнозирование выполнялось для горизонтов:
– $\Delta t \in \{1, 2, 3, 5, 10, 15, 20\}$ секунд

При этом использовались модели:

- Random Forest;
- Extra Trees;
- HistGradientBoosting;
- Ridge Regression;
- GRU/LSTM [34].



Рисунок 24 - Фазово-структурированный экспериментальный временной ряд

На рисунке 24 показано, что качество прогнозирования постепенно снижается с увеличением горизонта прогноза. Наиболее устойчивые результаты наблюдаются для горизонтов 1–5 секунд, тогда как при дальнейшем увеличении горизонта точность моделей начинает существенно ухудшаться.

Подобное поведение связано с накоплением неопределённости во временных рядах и ростом влияния нелинейных эффектов деградации системы.

Особенно важно отметить, что снижение качества прогноза происходит не линейно. В области малых горизонтов система демонстрирует выраженную краткосрочную предсказуемость, однако после достижения определённого временного порога точность моделей начинает снижаться значительно быстрее.

Это соответствует результатам анализа динамики деградации, полученным в предыдущей главе, где было показано, что поведение сети содержит области устойчивости и области нелинейного перехода между состояниями.

Наиболее устойчивыми к увеличению горизонта прогнозирования оказались ансамблевые методы:

- Extra Trees;
- Random Forest.

Данные модели лучше адаптировались к нелинейной структуре зависимости между параметрами нагрузки и будущими значениями KPI.



Рисунок 25 - Сравнение фактического и прогнозируемого RTT

Как показано на рисунке 25, модель способна воспроизводить основную динамику изменения RTT, включая участки роста задержек и переходы между состояниями системы. При этом наибольшие ошибки возникают в областях резких переходов и oscillation behavior.

Анализ результатов показал, что различные KPI обладают различной степенью прогнозируемости.

Наиболее устойчивыми к прогнозированию оказались:

- RTT;
- RTT p95;
- state score.

В то же время throughput демонстрировал существенно более сложное поведение. Это связано с тем, что throughput сильнее зависит от burst behavior транспортного уровня, TCP congestion control и кратковременных queue release events.

Таким образом, latency-based показатели отражают состояние деградации системы более устойчиво, чем throughput-based показатели.

Можно сделать bar-chart:

Таблица 14 - Сравнение качества прогнозирования по KPI

KPI	Mean R ²
RTT	high
RTT p95	high
state_score	medium/high

KPI	Mean R ²
throughput	lower

Дополнительно была исследована задача прогнозирования перехода сети в деградирующее состояние.

В данной постановке модель должна была определить вероятность перехода системы в состояния:

- degraded;
- breakdown.

до фактического наступления деградации.

Результаты показали, что признаки будущей деградации начинают проявляться заранее через:

- рост RTT;
- увеличение jitter;
- накопление очередей;
- изменение temporal structure временных рядов.

Это позволяет рассматривать прогнозирование как механизм раннего предупреждения деградации сети.

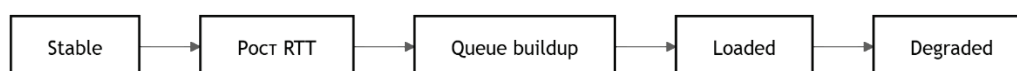


Рисунок 26 - Раннее обнаружение перехода системы в состояние деградации

Проведённый анализ показывает, что фазово-структурированные данные обладают выраженной краткосрочной прогнозируемостью. Это связано с тем, что система проходит через управляемые и относительно устойчивые переходы между состояниями, а временные ряды KPI содержат выраженную динамическую структуру.

В то же время области деградации и oscillation behavior остаются наиболее сложными для прогнозирования. Именно в этих режимах наблюдается максимальное накопление ошибок и снижение устойчивости моделей.

Полученные результаты показывают, что:

- состояние 5G SA системы обладает краткосрочной предсказуемостью;
- наиболее информативными признаками являются latency-based KPI;
- рабочий горизонт прогнозирования в фазово-структурированном режиме составляет порядка 1–5 секунд;
- деградация сети может быть обнаружена до наступления критического состояния.

Таким образом, фазово-структурированные эксперименты подтверждают возможность краткосрочного прогнозирования состояния сети в контролируемой 5G SA среде и создают основу для дальнейшего исследования прогнозирования в более сложных стохастических режимах.

4.5 Результаты прогнозирования по стохастическим данным

После исследования прогнозируемости системы в фазово-структурированном режиме был выполнен анализ прогнозирования по стохастическим экспериментальным данным.

В отличие от предыдущего режима, где изменение нагрузки происходило по заранее заданной фазовой структуре, в стохастическом режиме параметры системы изменялись случайным образом в пределах допустимых сценариев. Это позволило приблизить поведение экспериментальной среды к условиям реальной сети, в которой изменение нагрузки, активности пользователей и состояния системы носит существенно менее детерминированный характер.

Основной целью данного этапа являлось:

- исследование устойчивости моделей в условиях случайной динамики нагрузки;
- анализ деградации качества прогноза при усложнении поведения системы;
- определение рабочего горизонта прогнозирования в стохастическом режиме;
- оценка предсказуемости деградации сети при отсутствии жёсткой фазовой структуры.

Стохастический режим включал:

- случайное изменение offered load;
- динамическое изменение числа активных UE;
- изменение traffic patterns;
- случайные изменения параметров передачи;
- чередование различных состояний системы.

В результате временные ряды KPI приобретали существенно более сложную структуру по сравнению с фазово-структурированным режимом.

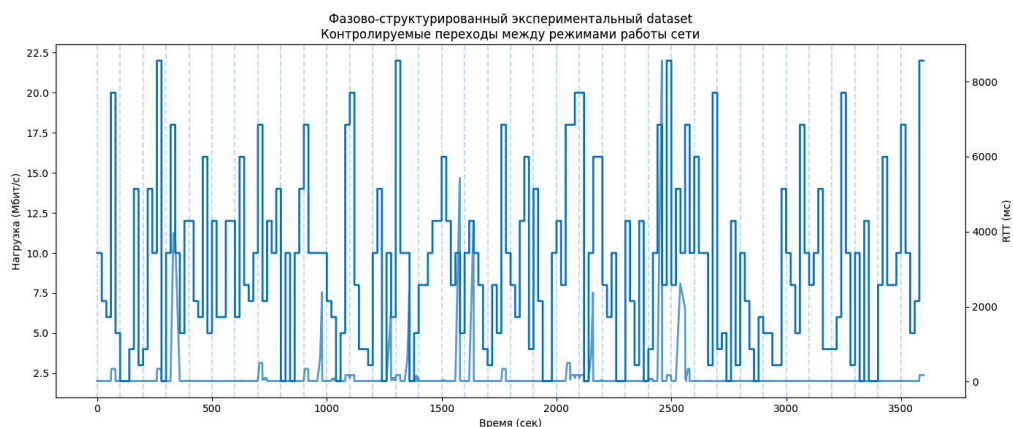


Рисунок 27 - Стохастический временной ряд состояния сети

Как показано на рисунке 27, стохастический режим характеризуется отсутствием жёсткой фазовой структуры и более сложной динамикой изменения KPI. Переходы между состояниями системы происходят менее предсказуемо, а

временные ряды содержат значительно больше локальных колебаний и кратковременных всплесков.

Для прогнозирования использовались те же модели, что и в фазово-структурированном режиме:

- Random Forest;
- Extra Trees;
- HistGradientBoosting;
- Ridge Regression;
- GRU/LSTM [34].

Прогнозирование выполнялось для горизонтов:

$\Delta t \in \{1, 2, 3, 5, 10, 15, 20\}$ секунд

Основное внимание уделялось прогнозированию RTT и интегрального показателя состояния сети.

Полученные результаты показали, что стохастический режим существенно усложняет задачу прогнозирования. Хотя модели сохраняют способность прогнозировать краткосрочное поведение системы, качество прогноза начинает снижаться значительно быстрее по сравнению с фазово-структурированным режимом.

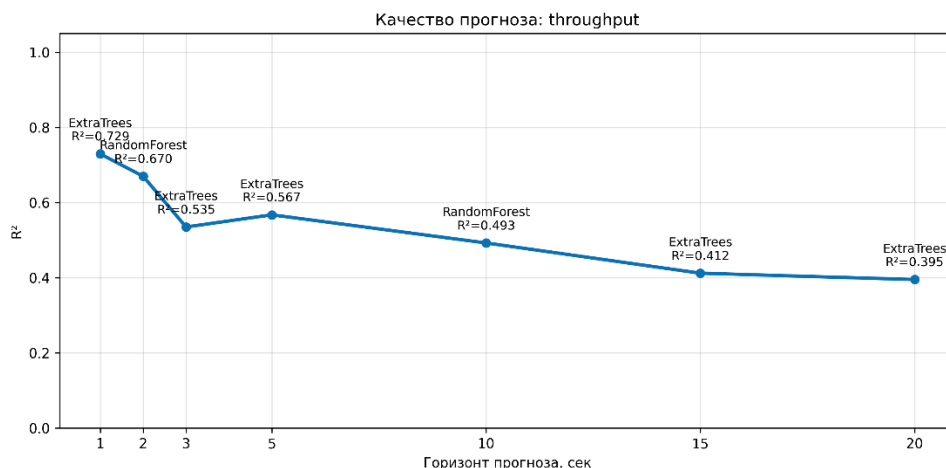


Рисунок 28 - Качество прогнозирования RTT по стохастическому временному ряду

На рисунке 28 показано, что качество прогнозирования постепенно ухудшается с ростом горизонта прогноза. Наиболее устойчивые результаты наблюдаются для горизонтов 1–3 секунды, тогда как после 5–10 секунд точность моделей начинает резко снижаться.

Подобное поведение связано с ростом неопределённости временного ряда и увеличением влияния случайных переходов между состояниями системы.

В отличие от фазово-структурированных данных, в стохастическом режиме временные ряды содержат существенно меньше устойчивых переходов между состояниями. Это приводит к тому, что модель быстрее теряет информацию о будущем состоянии сети.

Особенно заметное ухудшение качества наблюдается в областях:

- резкого роста нагрузки;
- queue buildup;
- oscillation behavior;
- burst traffic;
- переходов в degraded/critical state.

В данных режимах система начинает демонстрировать выраженную нелинейность и высокую чувствительность к кратковременным изменениям параметров нагрузки.

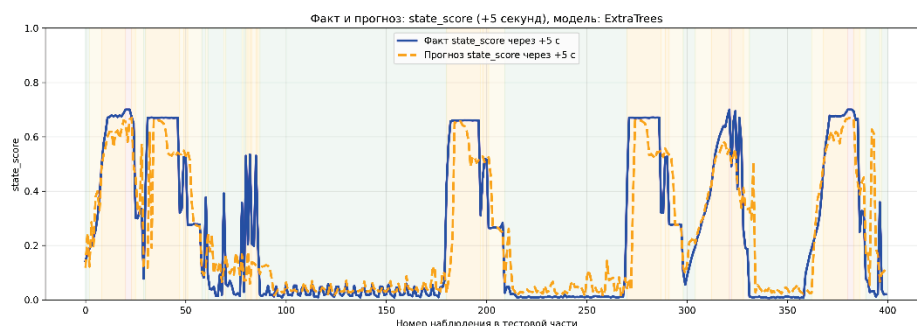


Рисунок 29 - Сравнение фактического и прогнозируемого RTT в стохастическом режиме

Как показано на рисунке 29, модель сохраняет способность воспроизводить общую тенденцию изменения RTT, однако точность прогнозирования существенно ухудшается в областях резких переходов между состояниями системы и при возникновении oscillation behavior.

Наибольшие ошибки наблюдаются вблизи областей деградации, где поведение системы становится существенно менее устойчивым и хуже описывается предыдущей динамикой временного ряда.

Несмотря на ухудшение качества прогноза, результаты показывают, что даже в стохастическом режиме система сохраняет ограниченную краткосрочную прогнозируемость.

Наиболее устойчивыми к стохастическим изменениям оказались:

- RTT;
- RTT p95;
- state score.

В то же время throughput демонстрировал наиболее нестабильное поведение и хуже прогнозировался на всех горизонтах времени.

Подобный результат согласуется с выводами предыдущей главы, где было показано, что throughput существенно сильнее зависит от кратковременных burst effects, TCP congestion behavior и queue release events.

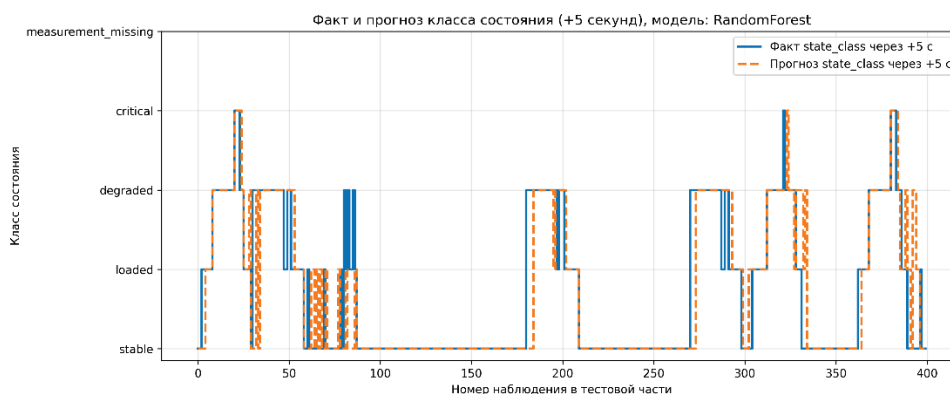


Рисунок 30 - Рабочие зоны прогнозирования в стохастическом режиме

На рисунке 30 показано, что рабочий горизонт прогнозирования в стохастическом режиме существенно сокращается по сравнению с фазово-структурированным режимом. Наиболее информативными остаются прогнозы на горизонтах 1–5 секунд, тогда как после 10 секунд качество прогноза становится ограниченным.

Дополнительно был проведён анализ причин ухудшения качества прогнозирования.

Основными источниками ошибок являются:

- нелинейность поведения системы;
- резкие переходы между состояниями;
- queue buildup;
- oscillation behavior;
- burst traffic;
- влияние TCP congestion control;
- высокая чувствительность системы к кратковременным изменениям нагрузки.

В отличие от фазово-структурированных данных, стохастический режим значительно сильнее приближен к условиям реальной эксплуатации сети, поэтому снижение качества прогноза в данном случае является ожидаемым результатом.

При этом особенно важным является то, что даже в условиях стохастической динамики система сохраняет признаки краткосрочной предсказуемости. Это означает, что временные ряды KPI содержат информацию о будущем состоянии сети даже при отсутствии жёсткой фазовой структуры.

Полученные результаты позволяют сделать следующие выводы:

- стохастический режим существенно усложняет задачу прогнозирования;
- качество прогноза снижается быстрее по сравнению с фазово-структурированными данными;
- наиболее устойчивыми к прогнозированию остаются latency-based показатели;
- throughput обладает наиболее сложной динамикой;
- рабочий горизонт прогнозирования сокращается до 1–5 секунд;

— области деградации и oscillation behavior являются наиболее сложными для моделей прогнозирования.

Таким образом, результаты стохастических экспериментов подтверждают, что прогнозирование состояния 5G SA сети возможно даже в условиях сложной динамики нагрузки, однако рабочий горизонт прогноза существенно зависит от характера поведения системы и степени случайности временного ряда.

4.6 Анализ ошибок и ограничений прогнозирования

Несмотря на полученные положительные результаты прогнозирования, проведённый анализ показал, что качество моделей существенно зависит от характера динамики системы, режима нагрузки и текущего состояния сети.

Наиболее высокое качество прогнозирования наблюдается в устойчивых режимах функционирования системы, где временные ряды KPI обладают сравнительно плавной и предсказуемой структурой. Однако при переходе системы в области деградации точность моделей начинает снижаться, а ошибки прогнозирования существенно возрастают.

Подобное поведение связано с тем, что деградация 5G SA системы носит выраженный нелинейный характер и сопровождается сложными переходными процессами.

Основными источниками ошибок прогнозирования являются:

- нелинейность поведения системы;
- burst traffic;
- queue buildup;
- oscillation behavior;
- saturation effects;
- TCP congestion dynamics;
- резкие переходы между состояниями сети.

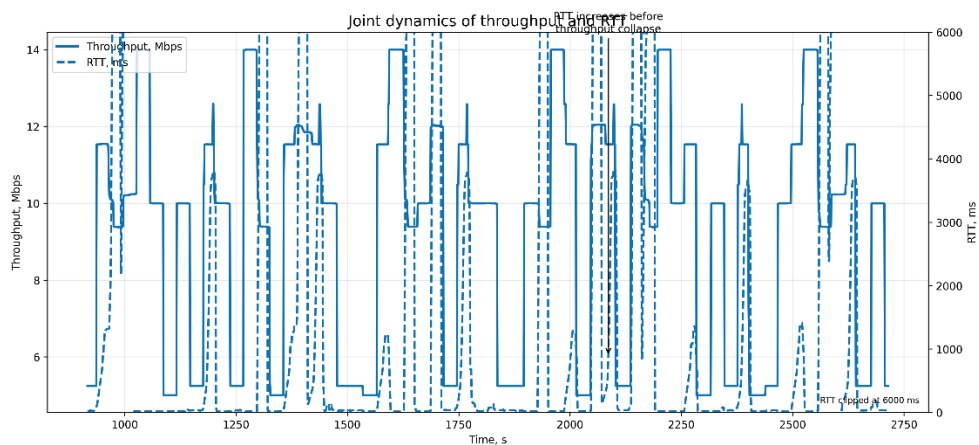


Рисунок 31 - Области устойчивого и неустойчивого прогнозирования

Как показано на рисунке 31, в устойчивых режимах функционирования временные ряды обладают сравнительно высокой предсказуемостью. Однако при переходе системы в состояние деградации структура временного ряда

становится существенно менее устойчивой, что приводит к росту ошибок прогнозирования.

Одной из ключевых особенностей исследуемой системы является наличие пороговых эффектов деградации. При достижении определённого уровня нагрузки система может резко переходить из состояния устойчивого функционирования в область queue buildup и saturation behavior.

Подобные переходы значительно усложняют задачу прогнозирования, поскольку небольшие изменения входных параметров могут приводить к существенным изменениям будущего состояния сети.

В ходе экспериментов было установлено, что модели наиболее чувствительны к следующим событиям:

- резкому увеличению числа активных UE;
- burst traffic;
- накоплению очередей;
- появлению oscillation behavior;
- переходу в degraded/critical state.

В данных режимах временные ряды теряют локальную устойчивость, а будущее состояние системы начинает существенно зависеть от кратковременных и трудно прогнозируемых процессов транспортного уровня.

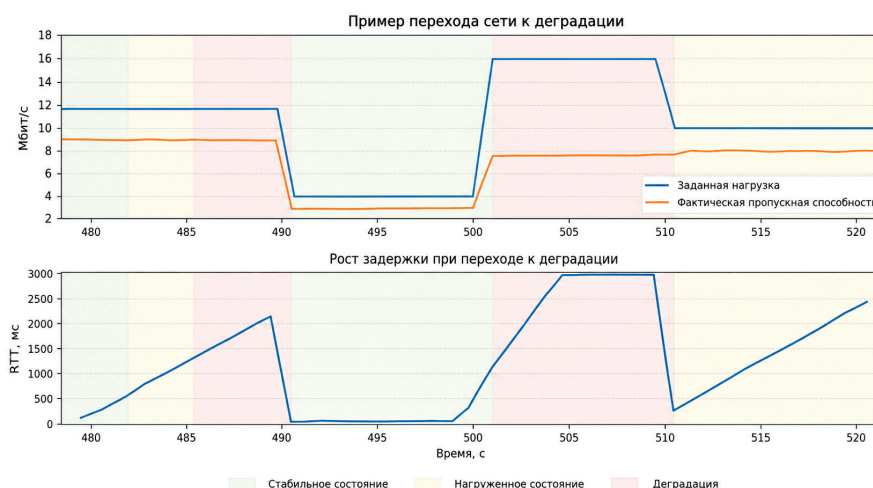


Рисунок 32 - Рост ошибки прогнозирования при переходе системы в состояние деградации

На рисунке 32 показано, что наибольшие ошибки прогнозирования возникают в областях нелинейного перехода между состояниями системы. Особенно выраженный рост ошибки наблюдается при возникновении queue buildup и oscillation behavior.

Существенное влияние на качество прогнозирования оказывает burst behavior транспортного уровня.

В отличие от плавного изменения нагрузки, burst traffic приводит к кратковременным всплескам RTT и throughput, которые плохо описываются предыдущей динамикой временного ряда. Подобные события возникают вследствие:

- TCP congestion control;
- packet retransmission;
- queue release events;
- кратковременных перегрузок буферов.

В результате система начинает демонстрировать локально хаотичное поведение, существенно снижающее предсказуемость будущих состояний сети.

Дополнительной особенностью является различная степень прогнозируемости отдельных KPI.

В ходе исследования установлено, что:

- RTT и RTT p95 обладают сравнительно высокой устойчивостью к прогнозированию;
- state score сохраняет устойчивую краткосрочную прогнозируемость;
- throughput демонстрирует наиболее сложную динамику;
- oscillation regions являются наиболее сложными для моделей прогнозирования.

Подобное различие связано с физической природой рассматриваемых показателей. RTT отражает накопление очередей и процессы деградации более инерционно, тогда как throughput сильнее зависит от кратковременных burst effects и TCP dynamics.

В результате latency-based показатели оказываются более устойчивыми для краткосрочного прогнозирования.

Отдельно необходимо отметить ограничения используемой экспериментальной среды.

Несмотря на воспроизводимость экспериментов и контролируемый характер среды, используемая архитектура на базе ZeroMQ и программной эмуляции не воспроизводит в полном объеме реальные физические эффекты радиоканала, включая:

- multipath fading;
- interference;
- mobility effects;
- реальные RF fluctuations.

Поэтому полученные результаты в большей степени отражают динамику транспортного уровня, queue buildup и поведения программно-определяемой части системы, чем реальные эффекты физического радиоканала.

Тем не менее подобный подход обладает важным преимуществом: он позволяет изолировать влияние отдельных факторов и исследовать причинно-следственные зависимости между параметрами нагрузки, состоянием системы и деградацией KPI.

Дополнительно следует учитывать, что увеличение числа UE и усложнение сценариев приводит к росту вычислительной нагрузки на сервер экспериментальной среды. В результате часть наблюдаемой деградации связана не только с поведением самой сети, но и с ограничениями вычислительных ресурсов тестовой платформы.

Несмотря на перечисленные ограничения, проведённое исследование показывает, что:

- краткосрочное прогнозирование состояния сети возможно;
- predictability существенно зависит от режима функционирования системы;
- наиболее сложными для прогнозирования являются области нелинейной деградации;
- burst traffic и oscillation behavior формируют зоны локальной непредсказуемости;
- latency-based показатели обладают наиболее устойчивой временной структурой.

Полученные результаты позволяют сделать важный вывод: поведение 5G SA системы обладает ограниченной краткосрочной прогнозируемостью, однако степень этой прогнозируемости определяется характером динамики сети и текущим режимом функционирования системы.

4.7 Анализ прогнозируемости деградации сети

Одним из наиболее важных результатов проведённого исследования является установление факта краткосрочной прогнозируемости деградации 5G SA системы.

В предыдущих разделах было показано, что деградация сети развивается не мгновенно, а представляет собой переходный процесс, включающий:

- рост RTT;
- накопление очередей;
- saturation behavior;
- oscillation dynamics;
- потерю устойчивости временных рядов.

Подобная структура деградации означает, что переход системы в критическое состояние сопровождается появлением предварительных признаков, наблюдаемых до наступления полной деградации сети.

В рамках настоящей работы прогнозируемость деградации рассматривается как способность модели обнаруживать признаки будущего ухудшения состояния системы на основе текущих значений KPI и динамики временных рядов.

Основное внимание уделялось следующим вопросам:

- существуют ли признаки приближающейся деградации;
- насколько заранее возможно обнаружить переход системы в degraded state;
- какие KPI являются наиболее информативными;
- как изменяется predictability при усложнении поведения системы.

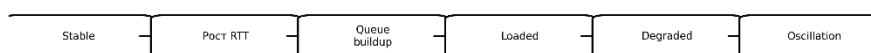


Рисунок 33 - Последовательность развития деградации сети

Как показано на рисунке 33, деградация системы развивается как последовательный переход между состояниями сети. При этом ухудшение состояния сопровождается накоплением предварительных признаков деградации, что создаёт предпосылки для краткосрочного прогнозирования будущего состояния системы.

Проведённый анализ показал, что наиболее ранними признаками приближающейся деградации являются:

- рост RTT;
- увеличение RTT p95;
- накопление очередей;
- рост jitter;
- изменение temporal structure временных рядов;
- увеличение вариативности latency-based KPI.

Особенно важно отметить, что throughput в ряде случаев сохраняет сравнительно стабильные значения даже после начала деградации системы, тогда как RTT начинает расти значительно раньше.

Это подтверждает выводы предыдущих глав о том, что latency-based показатели являются более чувствительными индикаторами деградации сети.

В ходе исследования было установлено, что система обладает выраженной краткосрочной прогнозируемостью на горизонтах 1-5 секунд.

На данных горизонтах модели сохраняют способность обнаруживать приближение деградации до фактического перехода сети в critical state.

При этом качество early warning существенно зависит от характера динамики системы.

Для фазово-структурированных данных прогнозируемость деградации остаётся сравнительно высокой даже на горизонтах порядка 10 секунд, поскольку переходы между состояниями системы обладают более устойчивой временной структурой.

В стохастическом режиме рабочий горизонт предупреждения существенно сокращается вследствие:

- случайных изменений нагрузки;
- burst behavior;
- oscillation dynamics;
- локальной потери устойчивости временных рядов.

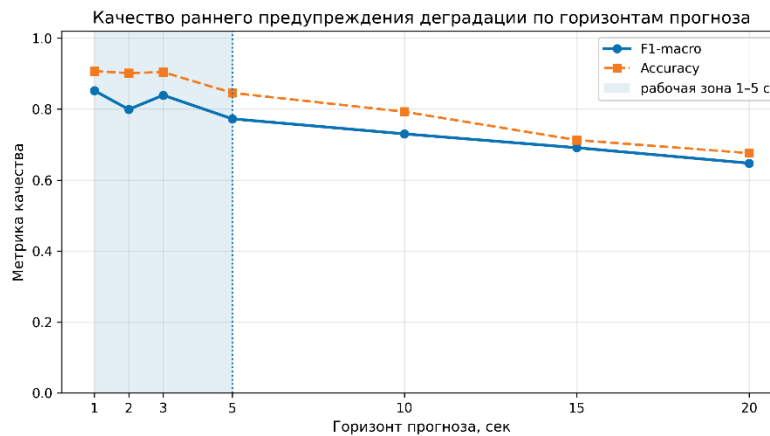


Рисунок 34 - Изменение качества раннего предупреждения деградации по горизонтам прогноза

На рисунке 34 показано, что качество раннего предупреждения постепенно снижается при увеличении горизонта прогноза. Наиболее устойчивые результаты достигаются на горизонтах 1–5 секунд, тогда как при дальнейшем увеличении горизонта неопределённость поведения системы существенно возрастает.

Особое значение для прогнозируемости деградации имеют процессы queue buildup [19].

В ходе анализа было установлено, что именно накопление очередей формирует основную временную инерцию деградации системы. Поскольку queue buildup развивается постепенно, временные ряды RTT начинают демонстрировать признаки ухудшения состояния до возникновения полного saturation behavior.

Это позволяет использовать latency-based показатели в качестве признаков раннего предупреждения деградации сети.

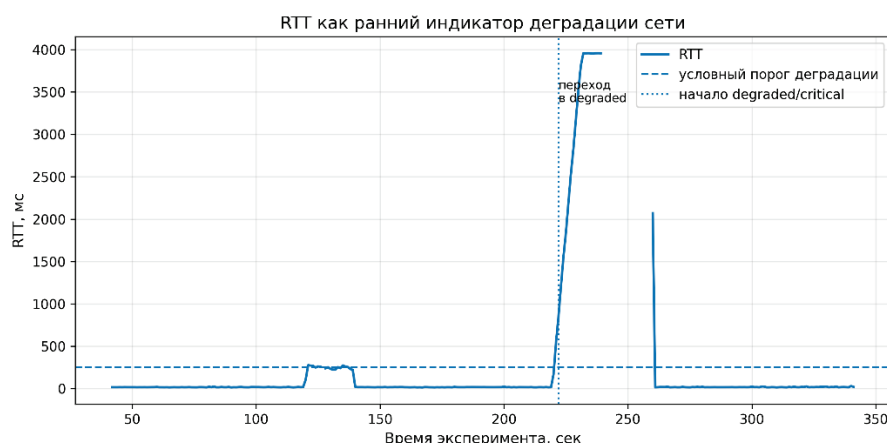


Рисунок 35 - RTT как ранний индикатор деградации сети

Как показано на рисунке 35, рост RTT начинается раньше полного перехода системы в состояние деградации. Это позволяет рассматривать latency-based показатели как ранние признаки ухудшения состояния сети.

Дополнительно проведенный анализ показывает, что predictability деградации определяется не только качеством модели машинного обучения, но и внутренней динамикой самой системы.

В устойчивых режимах временные ряды обладают сравнительно высокой инерцией и предсказуемостью. Однако при возникновении oscillation behavior система начинает демонстрировать локально хаотичное поведение, что существенно снижает возможность долгосрочного прогнозирования.

Таким образом, существует фундаментальное ограничение predictability, связанное с нелинейной динамикой деградации системы.

Несмотря на это ограничение, результаты исследования показывают, что даже краткосрочное прогнозирование обладает практической значимостью.

Возможность обнаружения деградации за несколько секунд до перехода системы в critical state потенциально позволяет:

- выявлять начало перегрузки сети;
- обнаруживать накопление очередей;
- формировать механизмы раннего предупреждения;
- оценивать устойчивость состояния системы;
- использовать прогнозирование как дополнительный источник observability [40].

Концептуальная схема раннего предупреждения деградации сети

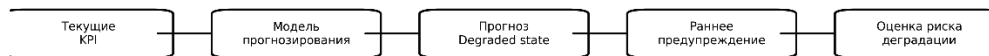


Рисунок 36 - Концептуальная схема раннего предупреждения деградации сети

Важным результатом работы является то, что predictability деградации была продемонстрирована именно в контролируемой воспроизводимой среде.

Это означает, что выявленные зависимости между поведением KPI и будущим состоянием системы являются воспроизводимыми и могут использоваться для дальнейших исследований механизмов observability и adaptive monitoring в программно-определяемых сетях.

Таким образом, проведенное исследование показывает, что:

- деградация 5G SA системы обладает предварительными признаками;
- состояние сети демонстрирует краткосрочную прогнозируемость;
- наиболее информативными являются latency-based показатели;
- queue buildup формирует временную структуру деградации;
- рабочий горизонт раннего предупреждения составляет порядка 1–5 секунд;
- predictability существенно снижается в областях oscillation behavior и burst traffic.

Полученные результаты подтверждают, что прогнозирование состояния сети может рассматриваться как инструмент анализа динамики деградации и раннего обнаружения ухудшения функционирования 5G SA системы.

4.8 Выводы по главе 4

В четвёртой главе выполнено исследование методов прогнозирования состояния 5G SA системы на основе экспериментальных данных, полученных в контролируемой воспроизводимой среде. Основное внимание было уделено анализу краткосрочной прогнозируемости поведения сети, исследованию устойчивости моделей при различных режимах функционирования системы и выявлению ограничений прогнозирования в условиях деградации.

Проведённый анализ показал, что состояние 5G SA системы обладает выраженной краткосрочной прогнозируемостью. Временные ряды KPI содержат информацию о будущем состоянии сети, что позволяет прогнозировать поведение системы на малых горизонтах времени.

В ходе исследования были рассмотрены как фазово-структурированные, так и стохастические режимы функционирования сети.

Для фазово-структурированных данных установлено, что контролируемые переходы между состояниями системы обладают устойчивой временной структурой, что обеспечивает сравнительно высокое качество прогнозирования. Наиболее устойчивые результаты были получены на горизонтах 1-5 секунд.

При этом отдельные модели сохраняли приемлемое качество прогноза и на более длительных горизонтах времени.

Для стохастических данных качество прогнозирования оказалось существенно ниже. Это связано с отсутствием жёсткой фазовой структуры, более сложной динамикой временных рядов и высокой чувствительностью системы к случайным изменениям нагрузки.

Тем не менее даже в стохастическом режиме было показано, что система сохраняет ограниченную краткосрочную прогнозируемость, особенно для latency-based показателей.

Одним из наиболее важных результатов исследования является установление того, что различные KPI обладают различной степенью прогнозируемости.

Наиболее устойчивыми к прогнозированию оказались:

- RTT;
- RTT p95;
- интегральный показатель состояния сети (state score).

В то же время throughput продемонстрировал существенно более сложную динамику и хуже прогнозировался на всех исследуемых горизонтах времени.

Проведённый анализ показал, что основными причинами ухудшения качества прогнозирования являются:

- нелинейность поведения системы;
- burst traffic;
- saturation behavior;
- oscillation dynamics;

- queue buildup;
- TCP congestion effects;
- резкие переходы между состояниями сети.

Особенно сложными для прогнозирования оказались области деградации и oscillation behavior, в которых временные ряды теряют локальную устойчивость и начинают демонстрировать выраженную нелинейную динамику.

В рамках исследования было показано, что деградация сети развивается как переходный процесс и сопровождается появлением предварительных признаков ухудшения состояния системы. Наиболее ранними индикаторами деградации являются:

- рост RTT;
- увеличение RTT p95;
- накопление очередей;
- рост jitter;
- изменение temporal structure временных рядов.

Полученные результаты подтверждают, что latency-based показатели могут использоваться как признаки раннего предупреждения деградации сети.

Важным результатом работы является установление рабочего горизонта прогнозирования. Для исследуемой экспериментальной среды наиболее устойчивое качество прогноза наблюдается на горизонтах 1-5 секунд.

При увеличении горизонта прогнозирования неопределённость поведения системы существенно возрастает, а качество моделей начинает быстро ухудшаться.

Проведённый анализ также показал, что predictability определяется не только используемой моделью машинного обучения, но и внутренней динамикой самой системы. В устойчивых режимах функционирования временные ряды обладают сравнительно высокой инерцией и предсказуемостью, тогда как в областях деградации и oscillation behavior система демонстрирует локально хаотичное поведение, ограничивающее возможность долгосрочного прогнозирования.

Несмотря на ограничения, результаты исследования подтверждают возможность использования прогнозирования как инструмента анализа состояния 5G SA системы и раннего обнаружения деградации сети.

В результате проведённого исследования получены следующие основные выводы:

14. Показано, что временные ряды KPI 5G SA системы обладают краткосрочной прогнозируемостью.

15. Установлено, что качество прогнозирования существенно зависит от характера динамики системы и режима функционирования сети.

16. Для фазово-структурированных данных получены более устойчивые результаты прогнозирования благодаря наличию управляемых переходов между состояниями системы.

17. Для стохастических данных рабочий горизонт прогнозирования существенно сокращается вследствие роста неопределённости временных рядов.

18. Наиболее устойчивыми к прогнозированию являются latency-based показатели: RTT, RTT p95 и state score.

19. Throughput демонстрирует наиболее сложную динамику и обладает меньшей устойчивостью к прогнозированию.

20. Основными источниками ошибок прогнозирования являются burst traffic, queue buildup, saturation behavior и oscillation dynamics.

21. Показано, что деградация системы сопровождается предварительными признаками, наблюдаемыми до наступления критического состояния сети.

22. Установлено, что рабочий горизонт раннего предупреждения деградации составляет порядка 1–5 секунд.

23. Подтверждена возможность использования прогнозирования для анализа состояния сети и раннего обнаружения деградации в контролируемой 5G SA экспериментальной среде.

Таким образом, результаты главы 4 подтверждают, что поведение 5G SA системы обладает ограниченной, но устойчивой краткосрочной прогнозируемостью. Полученные результаты завершают исследование динамики деградации сети и показывают возможность использования методов прогнозирования для анализа будущего состояния программно-определяемых 5G SA систем.

ЗАКЛЮЧЕНИЕ

В диссертационной работе решена взаимосвязанная научно-техническая задача, включающая разработку контролируемой воспроизводимой 5G SA экспериментальной среды, исследование динамики деградации состояния сети и анализ возможности краткосрочного прогнозирования состояния сети и показателей качества на основе данных, полученных из функционирующего программного стека 5G Standalone.

Разработанная экспериментальная среда построена на базе Open5GS, srsRAN Project и ZeroMQ и обеспечивает проведение воспроизводимых многопользовательских экспериментов с управляемым изменением параметров нагрузки и условий функционирования сети. Использование Linux namespaces позволило реализовать multi-UE конфигурацию с независимыми пользовательскими потоками и организовать автоматизированный сбор телеметрии из различных источников: ping, iperf и gNB logs. В результате сформирован структурированный датасет временных рядов, содержащий показатели RTT, RTT p95, throughput, jitter, packet loss и производные характеристики состояния сети.

Проведённые эксперименты показали, что деградация 5G SA сети в исследуемой среде не является мгновенным событием, а развивается как переходный процесс. Установлено, что ухудшение состояния сопровождается последовательным ростом RTT и RTT p95, накоплением очередей, saturation behavior и последующим переходом к oscillation behavior и потере устойчивости временных рядов. При этом throughput в ряде сценариев сохраняет сравнительно стабильные значения даже после начала деградации, тогда как latency-based показатели демонстрируют ухудшение значительно раньше.

На основе анализа временных рядов выявлены характерные переходы между состояниями сети:

stable → loaded → degraded → oscillation.

Для описания данных переходов предложена интерпретируемая модель состояний сети, позволяющая формализовать динамику деградации и использовать её для маркировки данных и постановки задач прогнозирования.

Для перехода от анализа отдельных KPI к интегральному описанию поведения системы предложен агрегированный показатель состояния сети state_score, объединяющий latency-based показатели, jitter, packet loss и throughput efficiency. Использование state_score позволило представить многомерное состояние сети в виде единого интерпретируемого скаляра и использовать его как целевой показатель прогнозирования состояния системы.

В работе реализован причинно-корректный no-leakage pipeline прогнозирования с temporal split и исследованы модели Ridge Regression, Random Forest, Extra Trees, HistGradientBoosting, а также рекуррентные нейросетевые архитектуры GRU/LSTM [34]. Исследование выполнялось на горизонтах прогнозирования 1, 2, 3, 5, 10, 15 и 20 секунд. [37, 38]

Полученные результаты показали наличие краткосрочной прогнозируемости состояния сети. Наиболее устойчивые результаты достигнуты

для latency-based показателей. Для фазово-структурированного режима RTT и RTT p95 сохраняют высокое качество прогнозирования на горизонтах до 5–10 секунд, тогда как throughput демонстрирует существенно более низкую предсказуемость вследствие нелинейного характера saturation behavior и burst traffic dynamics.

Показано, что рабочий горизонт прогнозирования определяется характером динамики сети. В фазово-структурированном режиме система обладает более устойчивой временной структурой, что обеспечивает сохранение приемлемого качества прогнозирования вплоть до 10–15 секунд. В стохастическом режиме вследствие случайных изменений нагрузки, burst traffic и oscillation behavior качество прогнозирования существенно снижается, а рабочий горизонт сокращается до 1–5 секунд.

Проведённый анализ показывает, что predictability деградации определяется не только используемой моделью машинного обучения, но и внутренней динамикой самой системы. Наличие queue buildup и временной инерции деградации формирует область краткосрочной предсказуемости, тогда как при переходе к oscillation behavior система начинает демонстрировать локально хаотичное поведение, ограничивающее возможность долгосрочного прогнозирования.

Практическая значимость работы заключается в создании воспроизводимой экспериментальной платформы для исследования поведения 5G SA сетей, формировании структурированного корпуса экспериментальных данных и демонстрации возможности раннего обнаружения деградации состояния сети. Полученные результаты показывают, что прогнозирование может рассматриваться как дополнительный инструмент observability и краткосрочного раннего предупреждения ухудшения состояния сети.

Научная новизна работы заключается в объединении воспроизводимой 5G SA экспериментальной среды, модели состояний сети и методов краткосрочного прогнозирования в единую исследовательскую схему, позволяющую анализировать динамику деградации сети как наблюдаемый переходный процесс с режимно-зависимой краткосрочной прогнозируемостью.

Таким образом, основной результат диссертационной работы состоит в том, что разработанная контролируемая воспроизводимая 5G SA среда позволила экспериментально показать: деградация сети имеет характер переходного процесса, обладает наблюдаемыми предварительными признаками и демонстрирует ограниченную, но практически значимую краткосрочную прогнозируемость, зависящую от характера динамики системы.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 1 ETSI. (2023). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (3GPP TS 23.501). ETSI.
- 2 ETSI. (2023). Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description (3GPP TS 38.300). ETSI.
- 3 O-RAN Alliance. (2018). O-RAN: Towards an Open and Smart RAN. O-RAN Alliance.
- 4 Trivedi, A., & Prasad, R. (2025). Open RAN: A Concise Overview. IEEE Communications Magazine, 63(5), 80–85.
- 5 Ericsson Mobility Report, November 2025. Ericsson, 2025.
- 6 O-RAN Alliance. (2021). Near-Real-Time RAN Intelligent Controller Architecture.
- 7 Rimedo Labs. (2021). O-RAN Near-Real-Time RIC.
- 8 O-RAN Software Community. (2023). O-RAN Architecture Overview.
- 9 NIST. (2022). 5G O-RAN Testbed – A Blueprint for Deploying Advanced Wireless Testbeds (NIST Technical Note 2311).
- 10 srsRAN Project. (2022). 5G SA End-to-End Application Note.
- 11 Open5GS. (2023). Quickstart Guide: Control and User Plane Separation.
- 12 Open5GS. (2023). Quickstart Guide: Service-Based Architecture.
- 13 ZeroMQ. (2021). The ZeroMQ Messaging Library.
- 14 Linux Foundation. (2020). ip-netns – Network Namespace Management.
- 15 iPerf3 Project. (2023). iPerf3: A TCP, UDP, and SCTP Network Bandwidth Measurement Tool.
- 16 TechTarget. (2019). What Is Ping and How Does It Work?
- 17 Bufferbloat.net. (2020). What Is Bufferbloat?
- 18 Nichols, K., Jacobson, V., & McKenzie, A. (2018). Controlled Delay (CoDel) Active Queue Management. RFC 8289.
- 19 Hoiland-Jorgensen, T., et al. (2018). FQ-CoDel: A Mixed Flow Scheduler/AQM. IETF Draft.
- 20 Kentik. (2023). Understanding Latency, Packet Loss and Jitter.
- 21 Cardwell, N., et al. (2021). BBR Congestion Control. IETF Internet-Draft.
- 22 Verizon. (2020). What Is 5G Network Slicing?
- 23 Celona. (2022). Network Slicing: What It Is & Why It Matters.
- 24 Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
- 25 Hoerl, A., & Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67.
- 26 Great Learning. (2022). Lasso Regression: A Complete Guide.
- 27 GeeksforGeeks. (2023). Gradient Boosting Algorithm.
- 28 GeeksforGeeks. (2023). Stochastic Gradient Descent.
- 29 GeeksforGeeks. (2023). F1 Score vs Accuracy.
- 30 mljar. (2021). Extra Trees vs Random Forest.

- 31 Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).
- 32 Microsoft Research. (2017). LightGBM Documentation.
- 33 Yandex Research. (2018). CatBoost: Gradient Boosting with Categorical Features.
- 34 Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- 35 Zhang, A., Lipton, Z., & Li, M. (2020). Dive into Deep Learning: Gated Recurrent Units.
- 36 Cho, K., et al. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation.
- 37 Boda, A., et al. (2023). Forecasting Network Traffic: A Survey and Tutorial with Open-Source Comparative Evaluation. *IEEE Access*, 11, 1–45.
- 38 Sadykova, A., et al. (2024). Multi-Step Internet Traffic Forecasting Models with Variable Forecast Horizons for Proactive Network Management. *Sensors*, 24(6), 1–21.
- 39 ETSI. (2023). Management and Orchestration; 5G Performance Measurements (3GPP TS 28.541). ETSI.
- 40 3GPP. (2025). Network Automation Enablers in 5GS.
- 41 CAIDA. (2020). LSTM Framework for Network Traffic Prediction.
- 42 ShareTechnote. (n.d.). 5G Core Network Architecture.
- 43 Flolive. (2023). 5G Core Network Architecture Explained.
- 44 GeeksforGeeks. (2023). Gradient Boosting in Machine Learning.
- 45 GeeksforGeeks. (2023). F1 Score in Machine Learning.
- 46 Great Learning. (2024). A Complete Understanding of LASSO Regression.
- 47 GeeksforGeeks. (2023). Stochastic Gradient Descent (SGD).
- 48 Kentik. (2023). Understanding Latency, Packet Loss and Jitter in Network Performance.
- 49 Rimedo Labs. (2022). Near-RT RIC in O-RAN.
- 50 Celona. (2022). Network Slicing: What It Is & Why It Matters.