

На правах рукописи

Данг Ван Тханг

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ МОДЕЛЕЙ И МЕТОДОВ ИНТЕГРАЦИИ
ТУМАННЫХ ВЫЧИСЛЕНИЙ И ФЕДЕРАТИВНОГО ОБУЧЕНИЯ В СЕТЯХ
СВЯЗИ ШЕСТОГО ПОКОЛЕНИЯ**

2.2.15. Системы, сети и устройства телекоммуникаций

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2026

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» на кафедре сетей связи и передачи данных.

Научный руководитель: доктор технических наук,
Волков Артём Николаевич

Официальные оппоненты: **Самуйлов Константин Евгеньевич**,
доктор технических наук, профессор,
Российский университет дружбы народов имени Патриса
Лумумбы, кафедра теории вероятностей и
кибербезопасности, заведующий кафедрой

Маркелов Олег Александрович,
кандидат технических наук доцент,
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ» им. В.И.
Ульянова (Ленина), кафедра радиотехнических систем,
заведующий кафедрой

Ведущая организация: Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»,
г. Москва

Защита состоится 09 декабря 2026 года в 11.00 на заседании диссертационного совета 55.2.004.01, созданном на базе федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича», по адресу: Санкт-Петербург, пр. Большевиков, д. 22, корп. 1, ауд. 554/1.

С диссертацией можно ознакомиться в библиотеке СПбГУТ по адресу Санкт-Петербург, пр. Большевиков, д. 22, корп. 1 и на сайте www.sut.ru.

Автореферат разослан 19 октября 2026 года.

Ученый секретарь
диссертационного совета 55.2.004.01,
канд. техн. наук, доцент

А.Г. Владыко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. После старта внедрения сетей пятого поколения (IMT-2020/5G) и достижения ими определённых результатов мировое научное сообщество сосредоточило усилия на исследовании и формировании сетей шестого поколения. Сети шестого поколения (IMT-2030/6G) представляют собой естественное развитие предыдущих поколений с возросшими характеристиками и ключевыми особенностями, которые заключаются в интеграции космического, воздушного, наземного и морского сегментов (с англ. SAGSIN – Space Air Ground Sea Integrated Networks) и широкого применения искусственного интеллекта. К характеристикам относятся повышение надёжности передачи, достижение ультрамалых задержек, а также поддержка сверхвысокой плотности устройств. Ожидается, что сети шестого поколения создадут основу для широкого распространения новых сервисов, таких как иммерсивные услуги, смешанная реальность, автономный (беспилотный) транспорт, системы телеприсутствия, и будут способствовать развитию сетевых вселенных (метавселенных).

Сети 6G формируются как качественно новая архитектурная парадигма, в которой искусственный интеллект (ИИ) перестаёт быть инструментом оптимизации и становится неотъемлемым элементом сети. В основе данного подхода лежит концепция естественной интеграции искусственного интеллекта (AI-native), согласно которой механизмы ИИ закладываются в архитектуру сети на стадии проектирования, а не добавляются как надстройка над готовой инфраструктурой. В 2023 году этот принцип был закреплён в рекомендации Международного союза электросвязи (МСЭ) Р М.2160, где в качестве основополагающих принципов проектирования сетей связи шестого поколения выступают: устойчивость, безопасность и отказоустойчивость, подключение неохваченных пользователей и повсеместный искусственный интеллект для повышения общей производительности системы.

Параллельно на уровне международной стандартизации сформировалась концепция сети связи с искусственным интеллектом, представляющая собой телекоммуникационную сеть, в которой технологии ИИ интегрированы в качестве естественных компонентов на всех уровнях архитектуры с целью повышения эффективности, возможностей и надёжности систем и сетей связи. Резолюция 101 ВАСЭ-24 (Всемирная ассамблея стандартизации электросвязи в Нью-Дели, 2024) зафиксировала возрастающую актуальность ИИ и обусловленную этим потребность в надёжных технических стандартах для сетей связи с поддержкой ИИ, а также поручила МСЭ-Т продолжить работу по применению ИИ при разработке рекомендаций, связанных с эксплуатацией и управлением сетями, надёжностью, безопасностью и сетями с поддержкой ИИ. Тем самым на международном уровне была подтверждена стратегическая направленность развития телекоммуникационной отрасли в сторону глубокой интеграции ИИ в сетевую инфраструктуру.

Реализация указанных принципов в условиях реальной инфраструктуры представляет собой сложную задачу. Среди технологий, рассматриваемых как перспективная основа, всё чаще выделяют совместное использование туманных вычислений (Fog Computing, FC) и федеративного обучения (Federated Learning, FL) в сетях шестого поколения. FC позволяют перенести обработку данных ближе к их источнику: между облаком и пользовательскими устройствами формируется промежуточный вычислительный слой, который принимает на себя значительную часть нагрузки, снижает затраты на передачу данных и уменьшает задержку. FL

предполагает обучение моделей ИИ непосредственно на устройствах без передачи исходных данных в основной центр обработки данных (ЦОД), что устраняет узкие места, связанные с пропускной способностью сети, и обеспечивает соответствие требованиям защиты персональных данных. Сочетание этих технологий формирует ресурсную базу для реализации моделей и методов в сетях связи с ИИ.

Таким образом, разработка и исследование моделей и методов интеграции туманных вычислений и федеративного обучения в сетях связи шестого поколения представляет собой актуальную научную задачу, решение которой способствует практической реализации принципов «естественной интеграции ИИ» (AI-native) и концепции сети связи с ИИ в условиях распределённой гетерогенной граничной среды сетей связи шестого поколения.

Разработанность темы исследования. Существует множество работ, посвящённых развитию сетей IMT-2020/5G и последующих поколений, а также архитектур граничных и туманных вычислений. Значительный вклад в теоретические и экспериментальные исследования инфокоммуникационных сетей внесли такие российские учёные, как А.Е. Кучерявый, Е.А. Кучерявый, А.С.А. Мутханна, К.Е. Самуйлов, А.Н. Волков и О.А. Маркелов, А.С. Аджемов, В.М. Вишневский, Ю.В. Гайдамака, В.С. Елагин, А.И. Парамонов, С.Н. Степанов, В.Г. Карташевский, В.О. Тихвинский, Б.С. Гольдштейн, А.Б. Гольдштейн, а также ведущие зарубежные исследователи M. Z. Chowdhury, M. Latva-aho, H. Pennanen, C. Puliafito, M. Verma и A. A. Ateya.

Значительный вклад в развитие парадигмы распределённого искусственного интеллекта, в частности алгоритмов федеративного обучения и децентрализованной оптимизации сетей связи, внесли B. McMahan, P. Kairouz, G. B. Giannakis, W. Chen, T. Nishio, T. Wang, A.A. Ateya.

В области прогнозирования городского трафика с использованием моделей глубокого обучения и пространственно-временных графовых нейронных сетей важные результаты получены в трудах L. Zhao, B. Yu, L. Bai, T. Qi. Кроме того, вопросам развёртывания искусственного интеллекта на границе сети с применением современных сред выполнения, таких как WebAssembly, посвящены работы M. Garofalo, J. Bachmeier, S. Berard, P. Gackstatter.

Настоящая диссертация, в отличие от известных подходов к построению и функционированию сетей связи, направлена на решение научной задачи разработки моделей и методов интеграции туманных вычислений и федеративного обучения в условиях широкого распространения искусственного интеллекта на уровне инфраструктуры сетей связи шестого поколения.

Объект и предмет исследования. Объектом исследования являются сети связи на основе туманных вычислений и федеративного обучения.

Предметом исследования являются модели и методы интеграции туманных вычислений, федеративного обучения в сети связи шестого поколения.

Цель и задачи исследования. Целью диссертационной работы является исследование и разработка моделей и методов построения сетей шестого поколения на основе туманных вычислений и федеративного обучения.

Цель работы достигается путём последовательного решения следующих задач:

– анализ существующего положения в области исследования сетей шестого поколения, роли и места искусственного интеллекта, а также туманных вычислений и федеративного обучения.;

- анализ существующего положения дел в области международной стандартизации, в частности рекомендаций МСЭ-Т по искусственному интеллекту, сетям шестого поколения, туманным вычислениям и федеративному обучению;

- разработка метода выборки в раундах федеративного обучения на основе динамического порога скорости изменения градиента функции потерь с вычислением порога на уровне туманных вычислений для снижения коммуникационных издержек;

- разработка децентрализованной структуры федеративного обучения с туманными вычислениями на основе адаптивной графовой свёрточной рекуррентной нейронной сети для обеспечения автономного функционирования при нестабильном соединении с облачным уровнем на примере задачи прогнозирования городского транспортного потока в сервисах городских сетевых вселенных (CitiVerse);

- разработка структуры и метода кроссплатформенного развёртывания федеративного обучения на гетерогенных узлах туманных вычислений на основе технологии WebAssembly и среды выполнения Wasmtime;

Научная новизна. Научная новизна работы состоит в следующем:

1. Разработан метод выбора пользовательских устройств и модель федеративного обучения с туманными вычислениями для сетей шестого поколения, отличающийся от известных тем, что адаптивный порог выбора динамически пересчитывается на уровне туманных вычислений в каждом раунде на основе статистики скоростей изменения нормы градиента, что позволяет учитывать неоднородность локальных данных и динамику процесса сходимости, а также снизить объём сетевого взаимодействия между устройствами и центральным сервером на 39% при сохранении точности модели.

2. Разработана децентрализованная структура федеративного обучения с туманными вычислениями на основе адаптивной графовой свёрточной рекуррентной сети для прогнозирования транспортных потоков в сервисах городских сетевых вселенных (CitiVerse), отличающаяся от известных тем, что полностью исключает центральный сервер из цикла обучения и агрегации модели искусственного интеллекта, а механизм раздельной агрегации глобальных весов и локальных векторов встраивания обеспечивает баланс между общими знаниями и региональной адаптацией модели при решении проблемы Non-IID (неоднородных) данных, а также учитывает новую структуру туманных альянсов. Предложенный подход демонстрирует высокую эффективность: снижает потребность в сетевых ресурсах на ~48%, по сравнению с традиционными методами федеративного обучения, ускоряет сходимость модели при сохранении точности модели прогноза.

3. Разработаны структура и метод кроссплатформенного развёртывания федеративного обучения на гетерогенных узлах туманных вычислений в архитектуре «облако – туман – пользовательское устройство», отличающиеся от известных решений тем, что система байт-кода WebAssembly используется как инфраструктурный слой выполнения локального обучения моделей федеративного искусственного интеллекта на гетерогенных туманных узлах, что устраняет необходимость в перекомпиляции и переконфигурации под каждую целевую аппаратную архитектуру, в отличие от традиционного (т.н. нативного) подхода и контейнеризации, тем самым ускоряя процесс развёртывания модулей федеративного искусственного интеллекта в гетерогенной среде туманных вычислений, что существенно снижает сложность развёртывания в сетях шестого поколения.

Теоретическая значимость работы. Теоретическая значимость диссертационной работы определяется совокупностью разработанных моделей и методов интеграции туманных вычислений и федеративного машинного обучения в сетях связи шестого поколения. Полученные результаты расширяют научные представления о принципах построения распределённых систем искусственного интеллекта в условиях гетерогенной вычислительной инфраструктуры сетей связи шестого поколения и создают теоретическую основу для дальнейших исследований в данной области.

Каждый из трёх научных результатов обладает самостоятельной теоретической ценностью. Для задачи адаптивного выбора пользовательских устройств предложены формализованный метод и модель федеративного обучения с туманными вычислениями на основе скорости изменения нормы градиента функции потерь с аналитическими соотношениями для порога, пересчитываемого по статистике устройств предыдущего раунда непосредственно на уровне туманных узлов. Для решения задачи децентрализованной агрегации предложены иерархическая структура и модель федеративного обучения на основе туманных альянсов, предусматривающие кластеризацию узлов с использованием алгоритма HDBSCAN и механизм обеспечения отказоустойчивости. Теоретически обоснован подход к отдельной обработке глобальных весов модели и локальных векторов встраивания с учетом особенностей распределения данных в условиях Non-IID; для задачи кроссплатформенного развёртывания разработаны структура и метод развёртывания федеративного обучения с переносимым слоем выполнения на основе технологии WebAssembly на гетерогенных узлах туманных вычислений, устанавливающая количественные соотношения между накладными расходами среды выполнения и числом раундов обучения модели ИИ.

Практическая значимость работы. Практическая ценность работы состоит в разработке моделей и методов интеграции туманных вычислений и федеративного обучения, обеспечивающих практическую реализацию принципов естественного искусственного интеллекта в сетях связи (с англ., AI-native networks) в условиях распределённой гетерогенной сети шестого поколения.

Реализация результатов работы. Результаты диссертационной работы использованы в Санкт-Петербургском государственном университете телекоммуникаций им. проф. М.А. Бонч-Бруевича при чтении лекций и проведении практических занятий для бакалавров по дисциплине «Искусственный интеллект в сетях и системах связи» (Рабочая Программа регистрационный номер № 23.05/119-Д). При чтении лекций и проведении практических занятий для магистров по дисциплине «Современные проблемы науки в условиях перехода к сетям шестого поколения (6G)» (Рабочая Программа регистрационный номер № 24.05/537-Д). При чтении лекций и проведении практических занятий для аспирантов по дисциплине «Системы, сети и устройства телекоммуникаций» (Рабочая Программа регистрационный номер № 21.05/780-Д). Акты реализации диссертационных исследований представлены в Приложении к диссертации.

Методы исследования. При проведении исследований были использованы методы теоретического исследования, теории оптимизации, методы машинного обучения и глубокого обучения, включая графовые свёрточные сети и рекуррентные нейронные сети, методы пространственной кластеризации, методы имитационного и натурного моделирования.

Основные положения, выносимые на защиту:

1. Метод выбора пользовательских устройств и модель федеративного обучения с туманными вычислениями на основе динамического порога скорости изменения нормы градиента функции потерь, позволяющий одновременно учитывать неоднородность локальных данных и скорость изменения градиента на текущем этапе обучения модели искусственного интеллекта, а также снизить объём сетевого взаимодействия между пользовательскими устройствами и сервером на 39%.

2. Децентрализованная структура и модель федеративного обучения с туманными альянсами на основе адаптивной графовой свёрточной рекуррентной нейронной сети для прогнозирования городских транспортных потоков в сервисах городских сетевых вселенных (CitiVerse), обеспечивающая снижение сетевого взаимодействия до 48% по сравнению с традиционными методами федеративного обучения.

3. Структура и метод развёртывания федеративного обучения на основе WebAssembly (низкоуровневого байт-кода) в гетерогенных средах туманных вычислений, обеспечивающие кроссплатформенное выполнение единого модуля обучения нейронной сети на узлах с различной аппаратной архитектурой, снижение потребления оперативной памяти на 69,5% по сравнению со стандартной реализацией при сохранении точности и скорости сходимости модели федеративного искусственного интеллекта.

Достоверность результатов. Степень достоверности полученных результатов подтверждается корректным применением математического аппарата теории вероятностей, теории оптимизации и методов машинного обучения, результатами имитационного моделирования на открытых реальных наборах данных (MNIST, CIFAR-10, METR-LA, PeMSD7M) и результатами натурных экспериментов на физических устройствах различных аппаратных архитектур, а также публикациями и выступлениями на российских и международных научных конференциях.

Апробация работы. Основные положения диссертационной работы обсуждались и были одобрены на следующих конференциях и семинарах: Международная конференция по современным системам вычислений и связям следующего поколения (International Conference on Advanced Computing & Next-Generation Communication, ICACNGC 2024) – Санкт-Петербург, Российская Федерация – 2024; Международная конференция по технологиям голографического телеприсутствия и голосетевселенной (International Conference on Holographic Telepresence Technologies and Holnetverse) – Санкт-Петербург, Российская Федерация – 2024; 4-я Международная конференция по инженерии и технологиям (4th SLIIT International Conference on Engineering and Technology, SICET 2025) – Шри-Ланкийский институт информационных технологий (SLIIT), Малабе, Шри-Ланка – 9-12 сентября 2025; 9-я Международная конференция по сетям будущего и распределённым системам (The 9th International Conference on Future Networks and Distributed Systems, ICFNDS '25) – Дубай, ОАЭ – 8-9 декабря 2025; 80-я научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвящённая Дню радио – Санкт-Петербург, Российская Федерация – 2025; 81-я научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвящённая Дню радио – Санкт-Петербург, Российская Федерация – 2026; а также на научных семинарах кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича.

Публикации. Основные результаты диссертации изложены в 10 публикациях, в том числе в 3 работах, опубликованных в журналах из перечня ВАК Министерства образования и науки Российской Федерации; 4 работах в изданиях, включённых в международные базы цитирования Scopus и Web of Science; 3 работах в других научных изданиях и материалах конференций.

Соответствие паспорту специальности. Диссертационная работа выполнена по специальности 2.2.15. Системы, сети и устройства телекоммуникаций и соответствует следующим пунктам паспорта специальности: 1, 2, 4, 7, 8, 12.

Личный вклад автора. Все основные результаты диссертационной работы получены автором самостоятельно. В экспериментальных исследованиях автор принимал непосредственное участие на всех этапах.

Структура и объём диссертации. Диссертация состоит из введения, четырёх глав, заключения, списка использованных источников и двух приложений. Полный объём диссертации составляет 169 страниц машинописного текста. Работа содержит 36 рисунков, 24 таблицы и список из 108 наименований литературных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы исследования и определены её ключевые особенности, сформулированы цели и задачи работы, перечислены основные положения, выносимые на защиту, определена научная новизна результатов, указаны теоретическая и практическая ценность результатов, приведены сведения об апробации работы, перечислены публикации по теме выполненного исследования.

В первой главе проведён анализ направлений развития сетей шестого поколения, роли и места искусственного интеллекта, а также туманных вычислений и федеративного обучения как ключевых технологий практической реализации концепции сети связи с искусственным интеллектом. Показано, что сети IMT-2030/6G представляют собой качественно новую архитектурную парадигму, в рамках которой искусственный интеллект перестаёт быть инструментом оптимизации и становится неотъемлемым элементом сети в соответствии с принципом “AI-native” (с англ. естественная интеграция искусственного интеллекта), закреплённым в Рекомендации МСЭ-Р М.2160. Резолюция 101 ВАСЭ-24 закрепила концепцию сети связи с искусственным интеллектом, в которой технологии ИИ интегрированы в качестве естественных компонентов на всех уровнях архитектуры с целью повышения эффективности, возможностей и надёжности систем связи.

Далее в главе систематизирован комплекс рекомендаций МСЭ-Т серии Y.317x – Y.3172, Y.3173, Y.3224, а также пара Y.3400/Y.3401, — формирующих нормативную основу для встраивания машинного обучения и федеративного обучения в телекоммуникационную инфраструктуру. Сопоставление данных рекомендаций с условиями практического развёртывания сетей IMT-2030 выявило два фундаментальных технологических пробела. Во-первых, архитектура ML-конвейера согласно Рекомендации Y.3172 неявно ориентирована на централизованную модель сбора и обработки данных, что перестаёт соответствовать условиям масштабирования до миллиардов конечных узлов и ужесточающимся требованиям к защите персональных данных, — это обуславливает необходимость федеративного обучения, реализуемого непосредственно на уровне туманных узлов. Во-вторых, Рекомендация Y.3224 вводит роль внутрисетевого агрегатора федеративного обучения, однако оставляет открытым вопрос о его физическом размещении и координации в гетерогенной инфраструктуре, что делает задачу архитектурной интеграции туманных узлов в цикл FL самостоятельной и нерешённой научной задачей.

Для устранения выявленных пробелов предложена трёхуровневая иерархическая архитектура интеграции туманных вычислений и федеративного обучения «облако – туманные узлы – пользовательские устройства», реализующая принцип “AI-native” и концепцию сети связи с искусственным интеллектом. Узлы туманных вычислений выполняют функцию промежуточных агрегаторов федеративного обучения, что позволяет снизить коммуникационные издержки, уменьшить задержку агрегации и устранить единую точку уязвимости за счёт распределения функций агрегации между множеством туманных узлов. Разработанная архитектура служит единой основой для исследовательских задач, решаемых в последующих главах (рисунок 1).

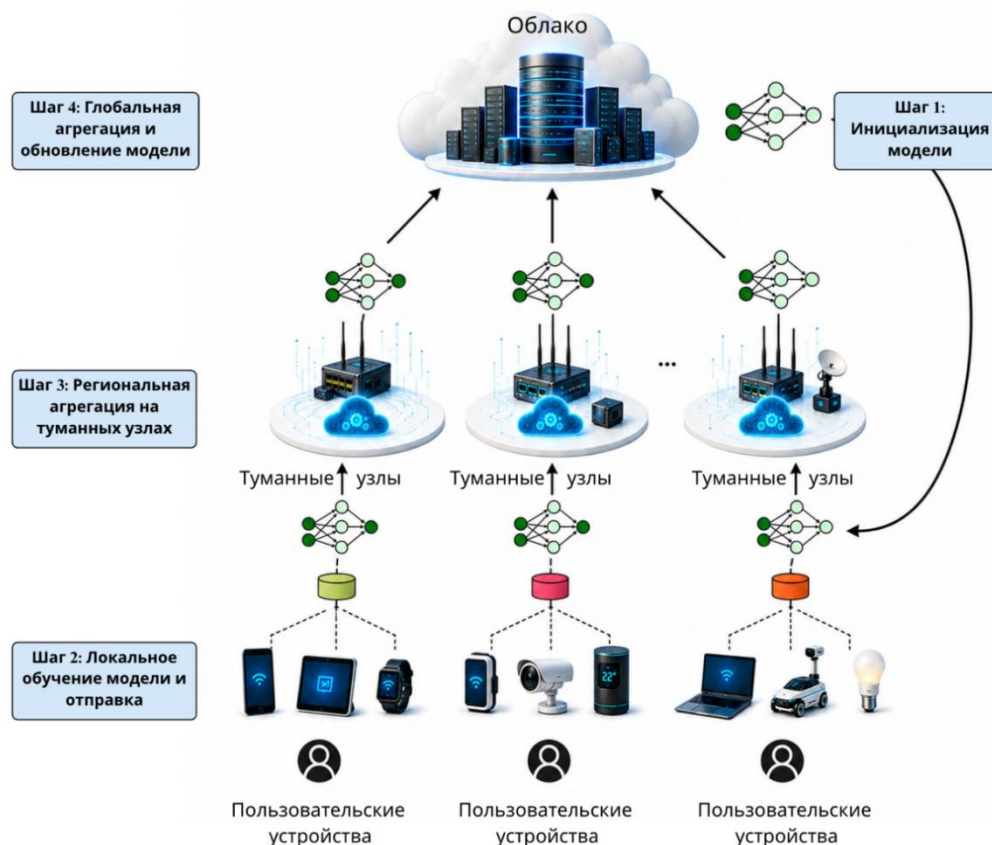


Рисунок 1. Трёхуровневая иерархическая архитектура интеграции туманных вычислений и федеративного обучения в сетях IMT-2030/6G

Во второй главе разработан метод выборки пользовательских устройств в интегрированной структуре федеративного обучения с туманными вычислениями на основе динамического порога скорости изменения нормы градиента функции потерь. В предложенной архитектуре уровень туманных вычислений выполняет функцию регионального агрегатора федеративного обучения: FC собирает обновления градиентов от граничных устройств, вычисляет взвешенное среднее и передаёт агрегированный результат на облачный уровень для формирования глобальной модели. Однако в условиях FL обновления от устройств, локальные данные которых статистически близки к глобальной модели, несут ограниченную информационную ценность для процесса сходимости и расходуют ресурсы канала впустую.

Структура интеграции FL и FC состоит из трёх уровней (рисунок 2). На уровне граничных устройств выполняется локальное обновление модели и вычисление скорости изменения нормы градиента; на уровне FC – фильтрация и промежуточная агрегация обновлений с динамическим пересчётом порога; на облачном уровне – глобальная агрегация модели и её распространение.

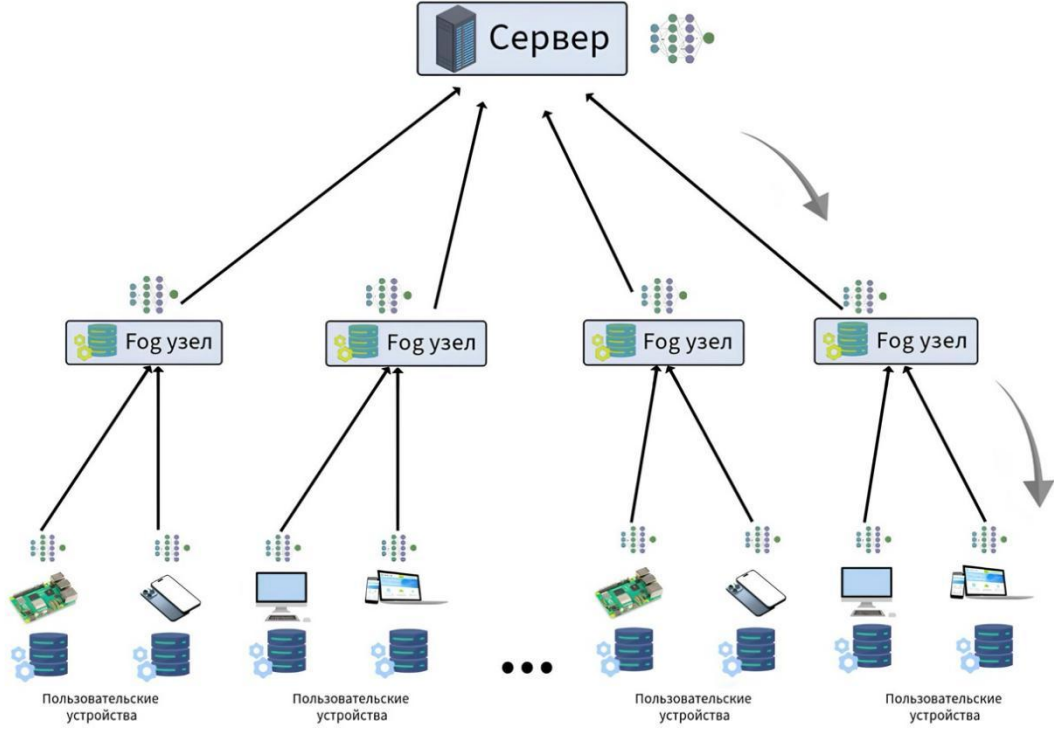


Рисунок 2. Структура интеграции FL и FC с тремя уровнями обработки

Скорость изменения нормы градиента функции потерь устройства k в раунде t определяется как:

$$r_t^{(k)} = \frac{\|\Delta\theta_t^{(k)}\|}{\|\Delta\theta_{t-1}^{(k)}\|} \quad (1)$$

где $\|\Delta\theta_t^{(k)}\|$ – норма градиента функции потерь устройства k в раунде t ; $\|\Delta\theta_{t-1}^{(k)}\|$ норма градиента функции потерь устройства k в раунде $(t-1)$.

Для каждого устройства k на итерации t обновление градиента задаётся как:

$$\Delta\theta = -\eta \nabla L^k(\theta_{t-1}) \quad (2)$$

где η – скорость обучения; $\nabla L(\theta_{t-1})$ – градиент функции потерь L относительно параметров модели θ на итерации t .

Узел туманных вычислений в каждом раунде t вычисляет среднее значение μ_t и стандартное отклонение σ_t скорости изменения функции потерь от устройств, отправивших обновления:

$$\mu_t = \frac{1}{K} \sum_{k=1}^K r_t^{(k)} \quad (3)$$

$$\sigma_t = \sqrt{\frac{1}{K} \sum_{k=1}^K (r_t^{(k)} - \mu_t)^2} \quad (4)$$

Динамический порог τ_t в раунде t определяется на основе среднего и стандартного отклонения скорости изменения градиента:

$$\tau_t = \mu_t + \alpha \sigma_t \quad (5)$$

где α – коэффициент корректировки, что позволяет настраивать чувствительность порога к колебаниям скорости изменения градиента.

Устройство k передаёт своё обновление модели в раунде t только при условии $r_t^{(k)} \geq \tau_t$; в противном случае передаётся лишь скаляр $r_t^{(k)}$ до ближайшего узла туманных вычислений. Такой подход позволяет адаптивно изменять количество передаваемых обновлений в соответствии с фазой процесса сходимости: на начальных этапах, когда градиенты демонстрируют значительную изменчивость, порог автоматически повышается, отсеивая малоинформативные обновления; по мере приближения модели к сходимости порог снижается, что предотвращает потерю значимых, но малых обновлений.

Экспериментальная оценка предложенного метода выполнена на наборах данных MNIST (70 тыс. изображений рукописных цифр) и CIFAR-10 (60 тыс. цветных изображений в 10 классах) с использованием простой свёрточной нейронной сети с двумя свёрточными слоями. Предложенный метод сравнивался с базовой моделью FL-FC и с моделью со случайной выборкой устройств. Кумулятивная связь с центральным сервером по раундам обучения на наборе MNIST представлены на рисунках 3, 4, 5, 6.

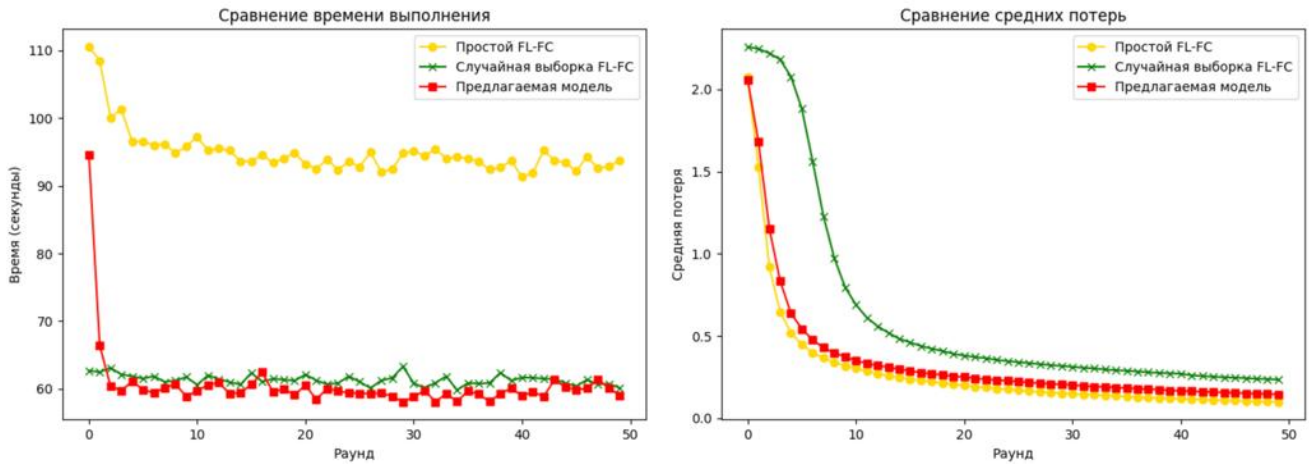


Рисунок 3. Время обучения и средние потери (MNIST)

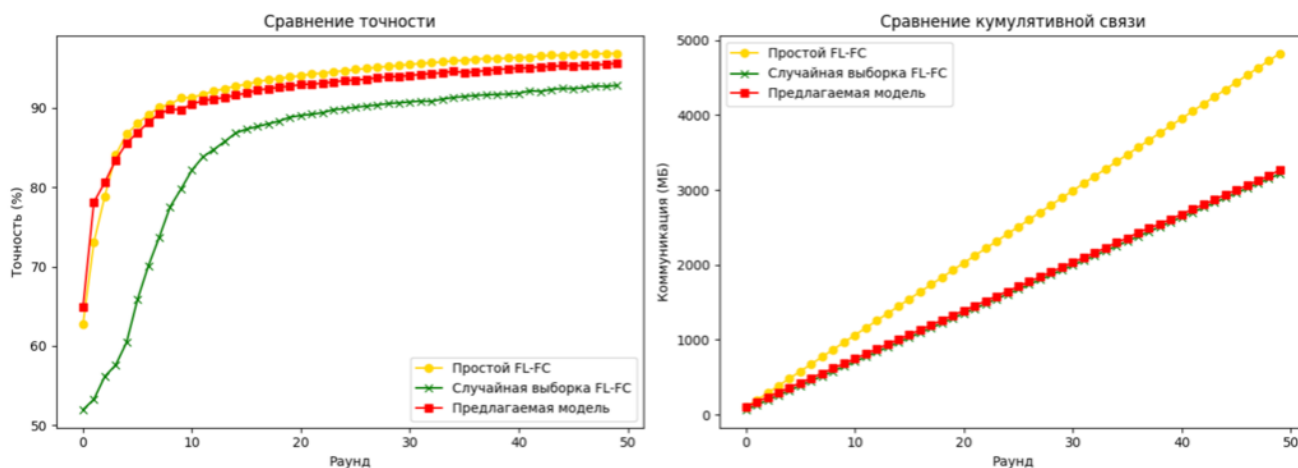


Рисунок 4. Значение средних потерь и кумулятивной связи (MNIST)

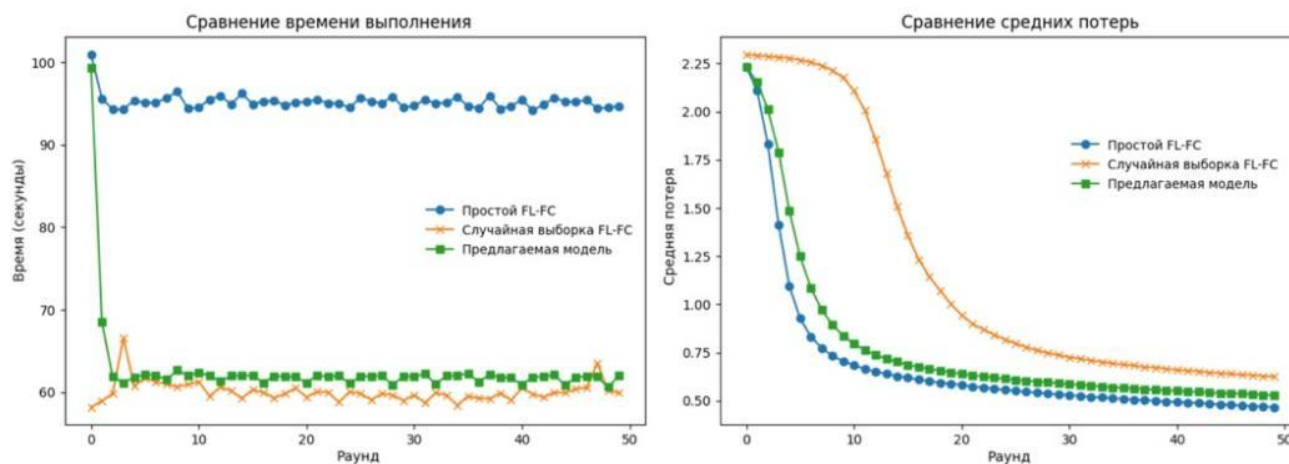


Рисунок 5. Время обучения и средние потери (CIFAR-10)

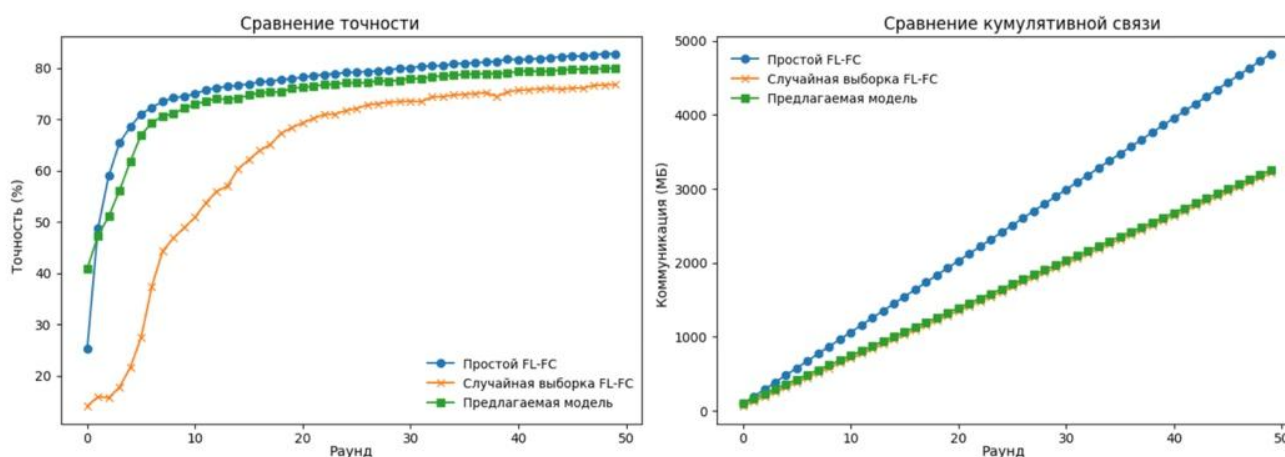


Рисунок 6. Средние потери и кумулятивная связь (CIFAR-10)

Результаты моделирования показывают, что предложенный метод позволяет снизить объём сетевого взаимодействия с центральным сервером на 39% на наборе MNIST и на 31% на наборе CIFAR-10 при сохранении точности модели на уровне базовой конфигурации FL-FC. Кроме того, предложенный метод обеспечивает сходимость за меньшее количество итераций по сравнению с моделью со случайной

выборкой устройств, поскольку обеспечивает участие в обучении тех устройств, чьи локальные данные несут наибольшую информационную ценность для глобальной модели.

В третьей главе Разработана децентрализованная структура FL с туманными альянсами, обеспечивающей автономное функционирование без непрерывного обращения к облачному агрегатору. Проведён анализ существующих методов прогнозирования городских транспортных потоков (FedAGCN, FedSTN, T-GCN, FASTGNN, FCGCN) и установлено, что все они сохраняют зависимость от центрального сервера агрегации, что создаёт риск единой точки отказа, не учитывает региональную неоднородность данных (Non-IID) и ограничивает устойчивость к нестабильности сетевого соединения с облаком.

Для устранения указанных недостатков предложена архитектура DFL-FC-AGCRN, объединяющая три взаимодополняющих компонента. Механизм одноранговой агрегации между узлами туманных вычислений полностью исключает центральный сервер из операционного цикла. Уровень туманных вычислений берёт на себя региональную координацию без постоянного обращения к облаку. Модель AGCRN выбрана в качестве основы локального обучения: она автоматически извлекает матрицу пространственных связей из данных и формирует индивидуальные параметры для каждого узла-датчика, что соответствует неоднородному характеру транспортных данных (Non-IID) в распределённой среде.

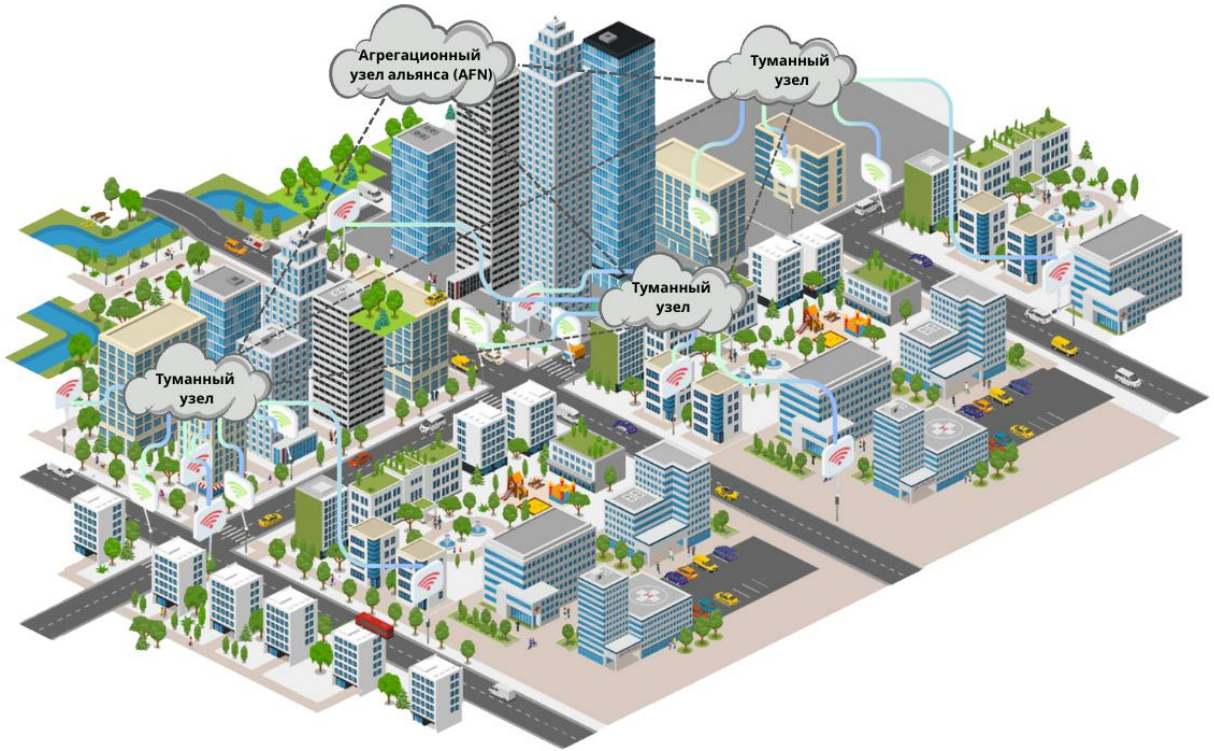


Рисунок 7. Концептуальная схема структуры DFL-FC-AGCRN

Локальное обучение на каждом туманном узле F_n управляется функцией потерь:

$$L(\theta_n) = \frac{1}{|B_n|} \sum_{(X,y) \in B_n} || f_{\theta_n}(X) - y ||_2^2 \quad (6)$$

где θ – вектор параметров модели; B_n – приватный набор данных узла F_n ; X – входная последовательность признаков; y – соответствующее целевое значение; $f_\theta(X)$ – прогнозная функция модели с параметрами θ ; $\|\cdot\|_2^2$ – квадрат евклидовой нормы.

Обновление параметров выполняется полностью на локальном уровне методом градиентного спуска, без передачи исходных данных на какой-либо сервер. Это одновременно обеспечивает конфиденциальность данных у источника и исключает роль центрального сервера из цикла обучения.

Ключевым методологическим вкладом главы служит механизм раздельной агрегации параметров. Параметры модели разделяются на глобальные (матрицы весов GCN–GRU), агрегируемые внутри альянса через узел AFN по формуле:

$$\theta_{agg,i}^{global} = \frac{1}{|B_i|} \sum_{F_n \in B_i} \theta_n^{global} \quad (7)$$

где $\theta_{agg,i}^{global}$ – агрегированные глобальные параметры альянса B_i ; $|B_i|$ – число узлов в альянсе; θ_n^{global} – глобальные параметры (матрицы весов GCN–GRU) туманного узла F_n после локального обучения

Локальные параметры при этом остаются на каждом узле и в процессе агрегации не участвуют. Такое разграничение позволяет модели одновременно обобщать закономерности, общие для всей сети, и сохранять региональную специфику каждой географической зоны – что принципиально недостижимо в централизованных схемах с традиционным усреднением всех параметров.

Построение топологии сети осуществляется в два этапа: сначала датчики кластеризуются алгоритмом HDBSCAN по матрице геодезических расстояний (формула гаверсинуса), затем туманные узлы объединяются в альянсы методом K-means по географическим координатам. Для каждого альянса основной узел AFN определяется по критерию минимакс-медоида:

$$Primary\ AFN(A) = \operatorname{argmin}_{j \in A} \max_{i \in A} D_{fog}(i, j) \quad (8)$$

где A – множество туманных узлов альянса; $D_{fog}(i, j)$ – геодезическое расстояние между узлами i и j , вычисленное по матрице попарных расстояний D_{fog} ; $\max_{i \in A} D_{fog}(i, j)$ – максимальное расстояние от кандидата j до любого узла альянса (внутриальянсовый радиус); argmin_j – узел, минимизирующий этот радиус.

Данный критерий гарантирует, что узел AFN является реально существующим элементом инфраструктуры и при этом минимизирует максимальное расстояние до любого участника альянса, снижая тем самым задержку внутриальянского взаимодействия. Для повышения отказоустойчивости дополнительно определяется резервный узел AFN по аналогичному критерию среди оставшихся узлов; механизм переключения на резервный узел функционирует автономно на узлах-участниках альянса без обращения к облаку.

Моделирование и оценка. Экспериментальная проверка проведена на двух открытых наборах данных системы измерения производительности Caltrans: METR-LA (207 датчиков) и ReMSD7(M) (228 датчиков), для горизонтов прогнозирования 5, 15 и

30 минут. Сравнивались три модели: централизованный AGCRN (верхняя граница точности), стандартный FL-AGCRN и предложенный DFL-FC-AGCRN. Для оценки точности прогноза используются три стандартные метрики: средняя абсолютная ошибка (MAE, Mean Absolute Error), корень из среднеквадратичной ошибки (RMSE, Root Mean Square Error), симметричная средняя абсолютная процентная ошибка (sMAPE, Symmetric Mean Absolute Percentage Error).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$\text{sMAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \quad (11)$$

где y_i – фактическое значение, $\hat{y}_i$ – прогнозное значение, n – число прогнозируемых примеров.

Точность прогноза DFL-FC-AGCRN сохраняется на конкурентном уровне по отношению к централизованному эталону на всех трёх горизонтах, а темп деградации с увеличением горизонта сопоставим с FL-AGCRN. По скорости сходимости предложенная модель достигает целевого значения функции потерь за 13 раундов против 18 у FL-AGCRN, что соответствует ускорению на 27,8%. Аналогичные результаты получены на наборе PeMSD7(M), подтверждая масштабируемость архитектуры при увеличении числа датчиков. Этот результат принципиально важен: модель использует географическую структуру сети для ускорения сходимости, что напрямую снижает коммуникационные издержки в условиях сетей шестого поколения.

Таблица 1. Влияние скорости сходимости на затраты на связь

Метрика	FL-AGCRN	DFL-FC-AGCRN	Экономия
Комм. затраты за раунд	34.33 МБ	17.71 МБ	48.4%
Затраты до достижения Loss 0,325 (FL) / 0,290 (DFL)	617,94 МБ (18 раундов)	232,18 МБ (13 раундов)	62.4%
Затраты до достижения Loss 0,320 (FL) / 0,285 (DFL)	1476,19 МБ (43 раунда)	589,38 МБ (33 раунда)	60.1%
Суммарные затраты за 100 раундов	3433.00 МБ	1771.43 МБ	48.4%

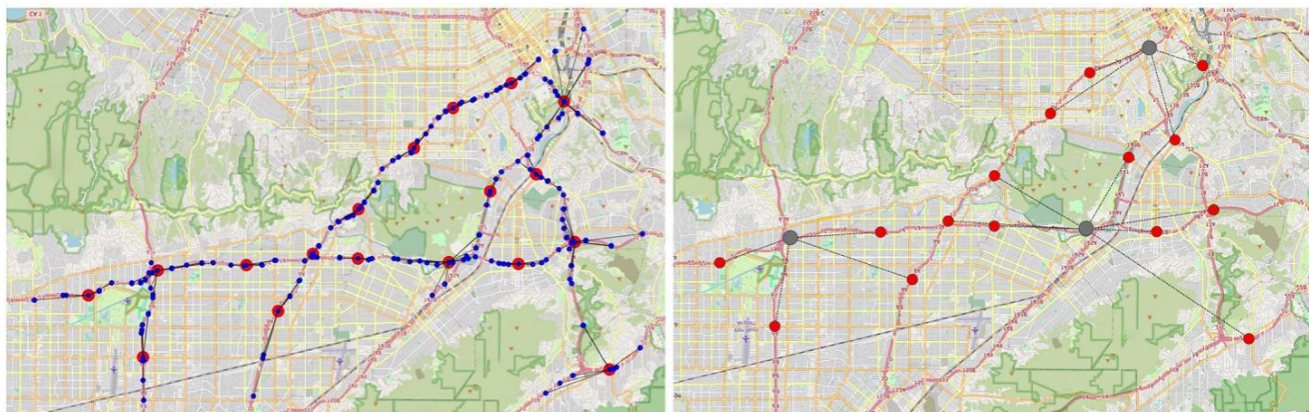


Рисунок 8. Развёртывание системы на наборе данных METR-LA

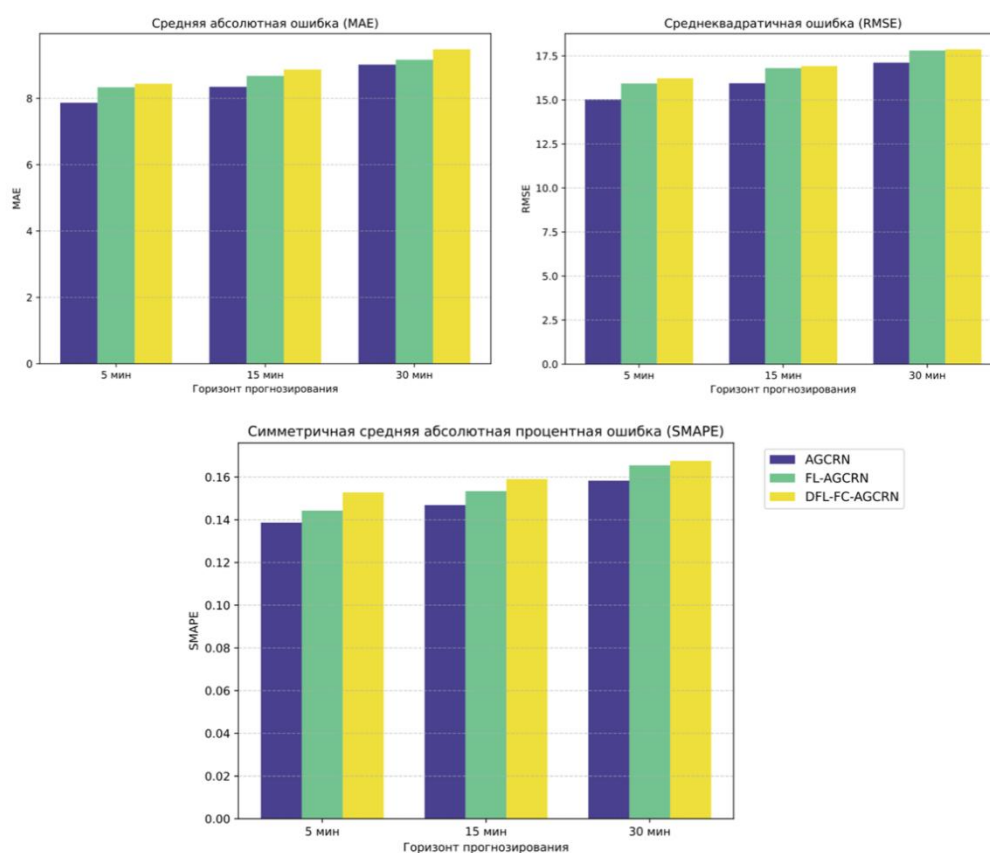


Рисунок 9. Точность моделей на наборе данных METR-LA

Результаты третьей главы подтверждают, что архитектура DFL-FC-AGCRN устраняет единую точку отказа, сохраняет автономную работоспособность при потере облачного соединения и обеспечивает точность прогноза в условиях неоднородных данных. Полученные результаты создают основу для четвёртой главы, в которой решается задача переносимости модуля локального обучения FL на гетерогенную инфраструктуру туманных вычислений.

В четвёртой главе разработана структура и метод кроссплатформенного развёртывания федеративного обучения на гетерогенных узлах туманных вычислений на основе технологии WebAssembly (Wasm) и среды выполнения Wasmtime. Инфраструктура туманных вычислений в реальных сетях IMT-2030 представляет

собой физически неоднородную совокупность устройств. Туманные узлы функционируют на различных архитектурах (ARM64, ARM, x86, RISC-V) под управлением разнородных операционных систем (Linux, Android и RTOS), при этом их вычислительные ресурсы существенно различаются. Код, скомпилированный под конкретную аппаратную архитектуру, не может быть исполнен на другой без повторной компиляции. Это обстоятельство превращает масштабное развёртывание единого программного стека федеративного обучения на гетерогенном парке туманных узлов в самостоятельную инженерную задачу, требующую отдельного решения. Предложенная структура реализует на практике принцип «скомпилировано однажды — выполняется везде» применительно к задаче локального обучения FL на гетерогенной инфраструктуре FC, показана на рисунке 10.

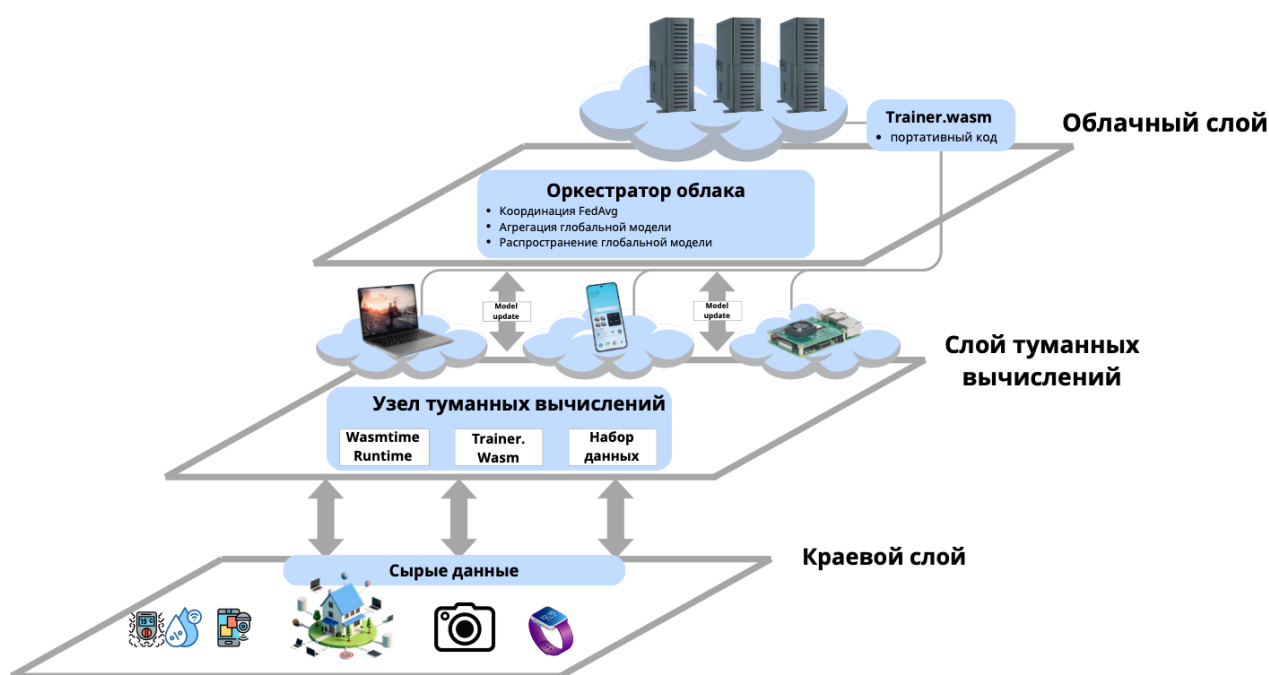


Рисунок 10. Интегрированная структура с WebAssembly

Для устранения данной проблемы предложена структура кроссплатформенного развёртывания FL на основе технологии WebAssembly и среды выполнения Wasmtime. Модуль локального обучения реализован на языке Rust и компилируется однократно в портативный бинарный файл формата wasm32-wasi. Полученный модуль `trainer.wasm` распространяется из облака и исполняется на любом туманном узле через среду Wasmtime без перекомпиляции — вне зависимости от аппаратной архитектуры и операционной системы. Выбор языка Rust обоснован совместимостью его архитектурных принципов с требованиями среды Wasm. Управление памятью без сборщика мусора в сочетании с абстракциями с нулевыми накладными расходами позволяет формировать компактные исполняемые модули с производительностью, близкой к нативной.

Системная модель построена по трёхуровневой схеме «облако – туман – пользовательское устройство». Центральная задача FL состоит в минимизации глобальной функции потерь на данных, распределённых по туманным узлам, без передачи исходных данных:

$$\min_{\theta} F(\theta) = \sum_{f=1}^F \frac{D_f}{N} F_f(\theta), \quad N = \sum_{f=1}^F D_f \quad (12)$$

где $F_f(\theta) = \frac{1}{D_f} \sum_{(x,y) \in B_f} l(\theta; x, y)$ – локальная функция потерь на наборе данных B_f узла f , $l(\theta; x, y)$ – функция потерь для одного образца, F – общее число туманных узлов.

Обновление параметров на каждом узле осуществляется методом стохастического градиентного спуска по мини-пакетам (x_b, y_b) размером B :

$$\theta \leftarrow \theta - \eta \cdot \nabla l(\theta; x_b, y_b)$$

где η – скорость обучения. По завершении E локальных эпох облачный оркестратор агрегирует обновления по алгоритму FedAvg:

$$\theta^{(r+1)} \leftarrow \sum_{f \in S_r} \frac{D_f}{\sum_{j \in S_r} D_j} \theta_f^{(r+1)} \quad (13)$$

где S_r – подмножество участвующих узлов в раунде r , $\theta_f^{(r+1)}$ – локальные параметры узла f после обучения.

Принципиально важно, что слой WebAssembly не изменяет математическую логику алгоритма. Траектория сходимости в среде Wasm-FL совпадает с нативной реализацией, что гарантирует алгоритмическую корректность результатов обучения. Модель безопасности WASI реализует принцип наименьших привилегий. По умолчанию модуль `trainer.wasm` лишён доступа к файловой системе, сети и системным часам. Хостовая среда явно предоставляет лишь два разрешения: чтение данных из выделенного каталога и запись обновлённых весов в выходной каталог. Этот механизм защищает туманные узлы от потенциально вредоносного кода обучения вне зависимости от источника его происхождения.

Моделирование и оценка. Экспериментальная проверка проведена в научно-исследовательской лаборатории Meganetlab 6G СПбГУТ им. проф. М. А. Бонч-Бруевича на физических устройствах: двух Raspberry Pi 4 (ARM64, Linux) и смартфоне Xiaomi 13 (ARM, Android), показана на рисунке 11. Набор данных MNIST разделён по узлам в режиме Non-IID с применением распределения Дирихле при $\beta = 0,5$ для эмуляции реального дисбаланса меток. Обучаемой моделью служит многоклассовая логистическая регрессия с 7850 параметрами. Сравнивались два варианта исполнения одного и того же кода на Rust: нативная компиляция под конкретную архитектуру и компиляция в `wasm32-wasi` с исполнением через Wasmtime.



Рисунок 11. Экспериментальное развёртывание системы на физических устройствах различных аппаратных архитектур

Сравнительная оценка нативного выполнения и выполнения через Wasm проведена для разного числа раундов обучения. Результаты экспериментов представлены на рисунках 12, 13, 14, 15, 16.

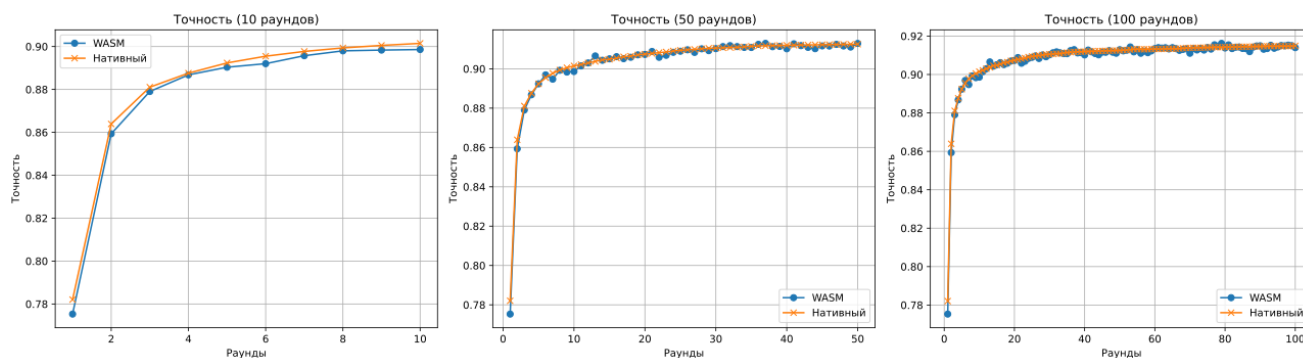


Рисунок 12. Точность по раундам обучения

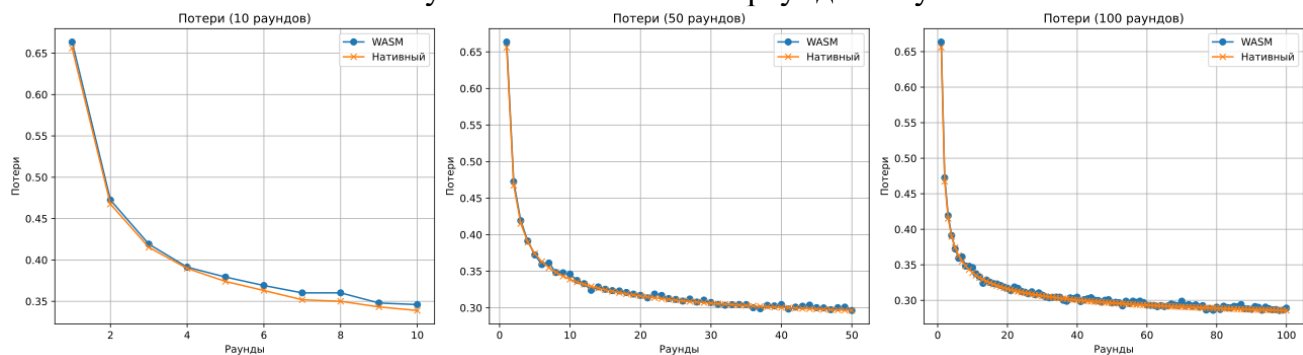


Рисунок 13. Потери по раундам обучения

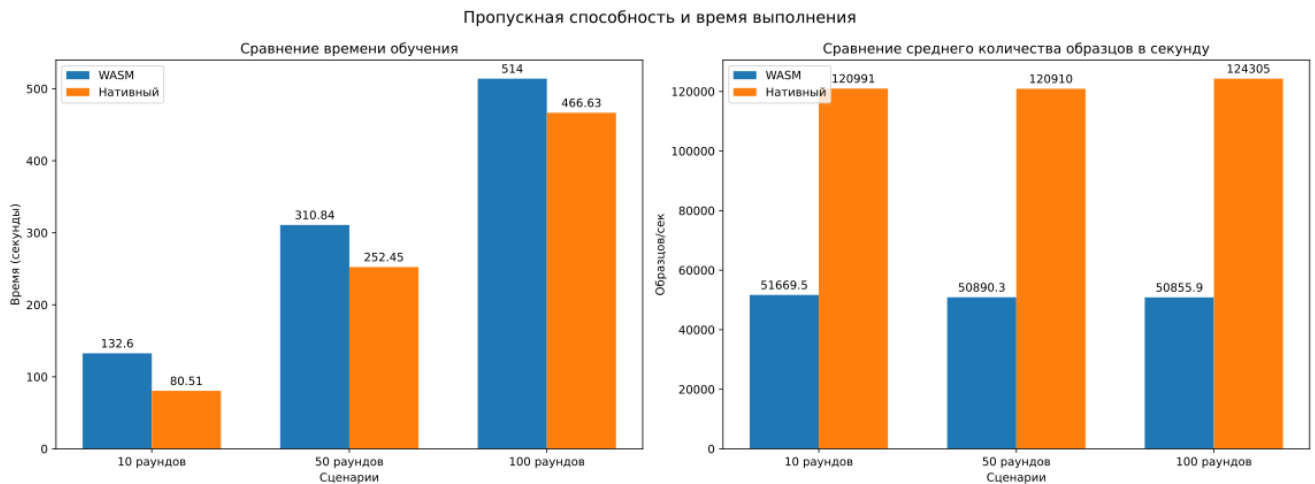


Рисунок 14. Сравнение времени обучения и средней пропускной способности

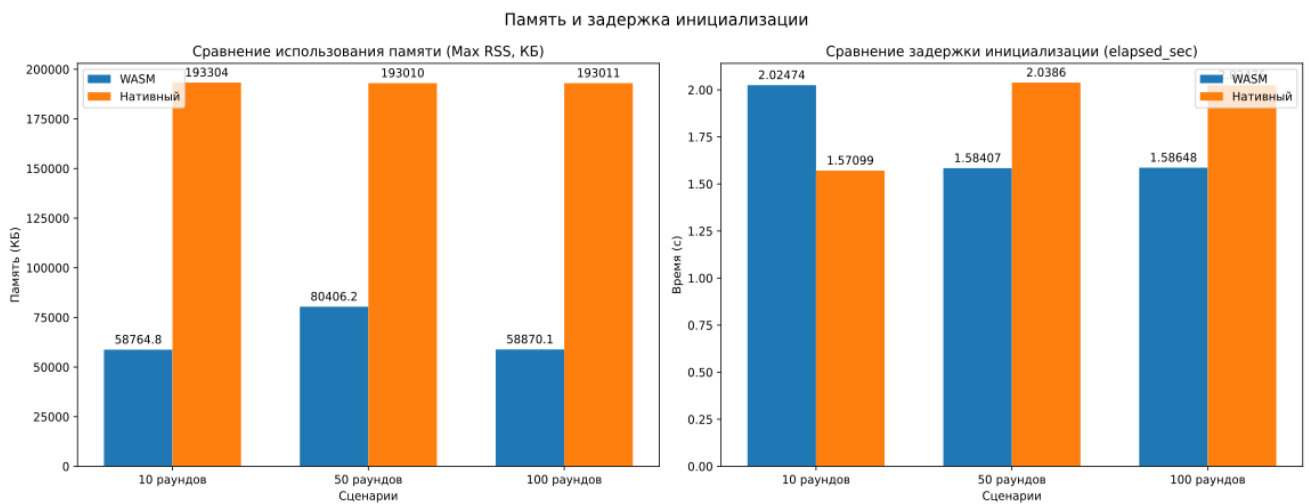


Рисунок 15. Потребление памяти и среднее время инициализации по раундам

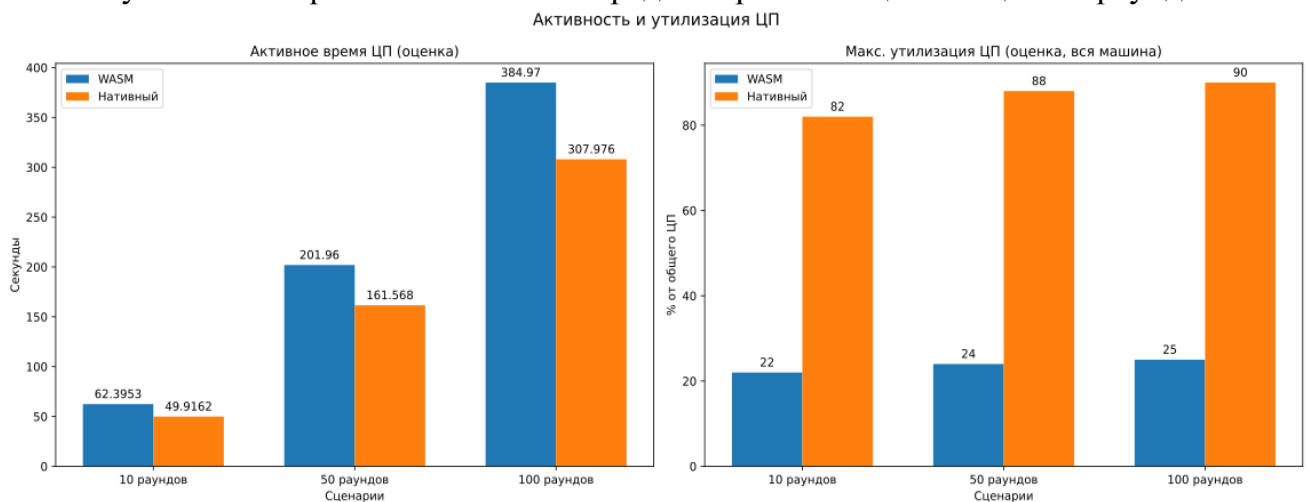


Рисунок 16. Активное время CPU и максимальная загрузка %CPU в процессе обучения

По алгоритмической корректности оба варианта демонстрируют практически идентичные траектории сходимости. После 100 раундов обучения расхождение по точности не превышает 0,3% (Wasm: 91,49% против Native: 91,49%), а разница по

функции потерь составляет менее 0,003. По ресурсным показателям преимущества Wasm весьма существенны: потребление оперативной памяти снижается на 69,5% – с 193,0 МБ до 58,8 МБ, загрузка процессора сокращается с 82-90% до 22-25%. Накладные расходы по времени выполнения уменьшаются монотонно с ростом числа раундов: с 64,6% при 10 раундах до 10,2% при 100 раундах. Это свидетельствует об успешной амортизации затрат на инициализацию среды Wasm в ходе длительной эксплуатации.

Результаты четвёртой главы подтверждают, что предложенная структура на основе WebAssembly реализует принцип «скомпилировано однажды – выполняется везде» применительно к задаче локального обучения FL на гетерогенной инфраструктуре туманных вычислений. При этом сохраняется алгоритмическая эквивалентность нативному коду и обеспечивается значительная экономия ресурсов памяти и процессора. Совокупность результатов четырёх глав формирует целостную архитектурную концепцию интеграции туманных вычислений и федеративного обучения в сетях связи шестого поколения, обеспечивающую практическую реализацию принципа “AI-native” и концепции сети связи с искусственным интеллектом и отвечающую на исследовательские задачи, сформулированные в первой главе диссертации.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

В диссертационной работе получены следующие основные результаты:

1. Решена научная задача разработки и исследования интегрированной архитектуры туманных вычислений и федеративного обучения в сетях шестого поколения, а также моделей и методов её практической реализации на гетерогенной инфраструктуре, обеспечивающей реализацию принципа “AI-native” и концепции сети связи с искусственным интеллектом согласно Рекомендации МСЭ-Р М.2160 и Резолюции 101 ВАСЭ-24.

2. Проведён анализ направлений развития сетей шестого поколения на основе Рекомендации МСЭ-Р М.2160, Резолюции 101 ВАСЭ-24 и Рекомендаций МСЭ-Т серии Y.317х. Установлено, что сети связи шестого поколения представляют собой качественный переход к интегрированной распределённой платформе, объединяющей связь, вычисления и искусственный интеллект в соответствии с принципом «естественного ИИ» (с англ. AI-native). Выявлены три фундаментальных пробела, включающие ограничения пропускной способности канала связи, недостаточные возможности алгоритмов для распределённого принятия решений и аппаратную неоднородность граничных узлов. Предложена трёхуровневая интегрированная архитектура туманных вычислений и федеративного обучения, реализующая концепцию сети связи с искусственным интеллектом путём отображения функциональных компонентов на физическую инфраструктуру «облако – туманные узлы – пользовательские устройства».

3. Разработан метод выбора пользовательских устройств и модель федеративного обучения с туманными вычислениями для сетей шестого поколения, отличающийся от известных тем, что адаптивный порог выбора динамически пересчитывается на уровне туманных вычислений в каждом раунде на основе статистики скоростей изменения нормы градиента, что позволяет учитывать неоднородность локальных данных и динамику процесса сходимости, а также снизить объём сетевого взаимодействия между устройствами и центральным сервером на 39% при сохранении точности модели.

4. Разработана децентрализованная структура федеративного обучения с туманными вычислениями на основе адаптивной графовой свёрточной рекуррентной сети для прогнозирования городских транспортных потоков в сервисах городских сетевых вселенных (CitiVerse), отличающаяся от известных тем, что полностью исключает центральный сервер из цикла обучения и агрегации модели искусственного интеллекта, а механизм раздельной агрегации глобальных весов и локальных векторов встраивания обеспечивает баланс между общими знаниями и региональной адаптацией модели при решении проблемы Non-IID (неоднородных) данных, а также учитывает новую структуру туманных альянсов. Предложенный подход демонстрирует высокую эффективность: снижает потребность в сетевых ресурсах на ~48 %, по сравнению с традиционными методами федеративного обучения, ускоряет сходимость модели при сохранении точности модели прогноза.

5. Разработаны структура и метод кроссплатформенного развёртывания федеративного обучения на гетерогенных узлах туманных вычислений в архитектуре «облако – туман – пользовательское устройство», отличающиеся от известных решений тем, что система байт-кода WebAssembly используется как инфраструктурный слой выполнения локального обучения моделей федеративного искусственного интеллекта на гетерогенных туманных узлах, что устраняет необходимость в перекомпиляции и переконфигурации под каждую целевую аппаратную архитектуру, в отличие от традиционного (т.н. нативного) подхода и контейнеризации, тем самым ускоряя процесс развёртывания модулей федеративного искусственного интеллекта в гетерогенной среде туманных вычислений, что существенно снижает сложность развёртывания в сетях шестого поколения.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в рецензируемых научных изданиях, рекомендованных ВАК:

1. Данг, В.Т. Структура федеративного машинного обучения с туманными вычислениями и метод выборки пользовательских устройств на основе динамического порога / В.Т. Данг, А.Н. Волков // Электросвязь. – 2026. – № 5. – С. 13-25.
2. Данг, В.Т. Туманные альянсы: децентрализованная облачная структура с федеративным машинным обучением для городских сетевых вселенных / В.Т. Данг, А.Н. Волков // Труды учебных заведений связи. – 2026. – Т. 12. – № 3. – С. 81-97.
3. Dang, V.T. A WebAssembly-Enabled Federated Learning Deployment Framework for Fog Computing Environments / V.T. Dang, A.Y. Koucheryavy // T-Comm: Телекоммуникации и транспорт. – 2026. – Т. 20. – № 6. – С. 52-62.

Публикации в изданиях, входящих в международные базы цитирования (Scopus, Web of Science):

4. Van Thang, D. A Framework Integrating Federated Learning and Fog Computing Based on Client Sampling and Dynamic Thresholding Techniques / D. Van Thang, A. Volkov, A. Muthanna, I. A. Elgendy, R. Alkanhel, D. N. K. Jayakody, A. Koucheryavy // IEEE Access. – январь 2025. – DOI: 10.1109/ACCESS.2025.3571979.
5. Van Thang, D. Future of Telepresence Services in the Evolving Fog Computing Environment: A Survey on Research and Use Cases / D. Van Thang, A. Volkov, A. Muthanna, A. Koucheryavy // Sensors (MDPI). – май 2025. – DOI: 10.3390/s25113488.

6. Van Thang, D. A New Challenges of Mirror Worlds: Network, Cloud Resources and Energy Issue / Mai Cong Huong, A. Volkov, D. Van Thang, A. Muthanna, A. Koucheryavy, M. Ibrahim // ICFNDS '25: Proceedings of the 9th International Conference on Future Networks and Distributed Systems. – 05 May 2026. – P. 1076-1086. – DOI: 10.1145/3789692.3789830.

7. Van Thang, D. Task Scheduling Problem in Fog Computing Environment with Improved Memetic Algorithm / D. Van Thang, A. Volkov, A. Muthanna, O. Vorozheikina, A. Koucheryavy // SLIIT Journal, Proceedings of the 4th SLIIT International Conference on Engineering and Technology (SICET 2025). – март 2026. – DOI: 10.54389/OVFU5465.

Публикации в других научных изданиях и материалах конференций:

8. Данг, В.Т. Анализ влияния Green IT на распределённые туманные вычисления: возможности и вызовы / В.Т. Данг, А.Н. Волков // Научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвящённая Дню радио. – 2025. – № 1(80). – С. 171-174.

9. Данг, В.Т. Услуги телеприсутствия и их будущее в развивающейся среде туманных вычислений: обзор исследований и примеров использования / В.Т. Данг, А.Н. Волков // Информационные технологии и телекоммуникации. – 2024. – Т. 12. – № 3. – С. 29-51.

10. Данг, В.Т. Проблема конвергенции федеративного машинного обучения, туманных вычислений и автономных сетей в архитектуре сетей 6G / В.Т. Данг, А.Н. Волков // 81-я научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвящённая Дню радио. – Санкт-Петербург, 20-24 апреля 2026 г.