

На правах рукописи

Волков Артём Николаевич

**ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ ПОСТРОЕНИЯ  
ИНФРАСТРУКТУРЫ И ПРЕДОСТАВЛЕНИЯ УСЛУГ СЕТЕЙ СВЯЗИ  
НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

2.2.15. Системы, сети и устройства телекоммуникаций

Автореферат  
диссертации на соискание ученой степени  
кандидата технических наук

Санкт-Петербург – 2021

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

Научный руководитель: доктор технических наук, профессор  
**Кучерявый Андрей Евгеньевич**

Официальные  
оппоненты:

**Татарникова Татьяна Михайловна**,  
доктор технических наук, доцент,  
Санкт-Петербургский государственный  
электротехнический университет "ЛЭТИ"  
им. В.И. Ульянова (Ленина), кафедра  
информационные системы, профессор кафедры

**Буранова Марина Анатольевна**,  
кандидат технических наук, доцент,  
Поволжский государственный университет  
телекоммуникаций и информатики, кафедра  
информационной безопасности, доцент кафедры

Ведущая организация:

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Российский университет дружбы народов»,  
г. Москва

Защита состоится 15 декабря 2021 года в 16.00 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](http://www.sut.ru).

Автореферат разослан 15 октября 2021 года.

Ученый секретарь  
диссертационного совета 55.2.004.01,  
д-р техн. наук, доцент

М.А. Маколкина

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** Тематика Искусственного Интеллекта (ИИ) в сетях связи в мире научных исследований появилась сравнительно недавно и вызывает все больший интерес со стороны как научных работ и проектов, так бизнеса и производства. Причиной большого интереса к применению технологий ИИ в сетях связи является надежда на решение множества трудноразрешимых задач сетей связи 5G и последующих поколений. Стоит отметить, что в Секторе Стандартизации МСЭ достаточно активно ведутся работы по разработке рекомендаций в области ИИ в сетях связи. При этом на базе 13 Исследовательской Комиссии (ИК), в декабре 2020 года была создана специальная Фокус-группа по исследованию и стандартизации автономных сетей, в открытии которой принимал участие также и автор данной диссертационной работы. Термин «Автономные сети связи» носит прикладной характер, с точки зрения стандартизации и технической реализации сетей связи с ИИ. Актуальность направления «Искусственный Интеллект в сетях связи» можно также подтвердить существующими международными событиями. Например, в 2020 году в рамках “AI for Good Global Summit” прошел международный конкурс “AI/ML in 5G Challenge”, организуемый МСЭ под эгидой ООН, который собрал множество стран, известных исследовательских университетов и центров, а также крупных компаний таких, как: Vodafone, China Telecom, Cisco и другие.

**Степень разработанности темы.** В области Интернета Вещей, сетей связи пятого поколения, Искусственного Интеллекта в сетях связи существует немало работ отечественных и зарубежных ученых: В.М. Вишневского, Б.С. Гольдштейна, В.Г. Карташевского, А.Е. Кучерявого, А.И. Парамонова, К.Е. Самуйлова, В.К. Сарьяна, С.Н. Степанова, Ю.В Гайдамаки, Р.В. Киричка, Е.А. Кучерявого, Д.А. Молчанова, В.О. Тихвинского, А.С.А. Мутханны, J. Andrews, J. Araniti, A.А.А. Ateya, M.Z. Faten и других. Стоит отметить, что на данный момент большую часть составляют работы по сетям связи пятого поколения, новым услугам, беспроводным и mesh-сетям, технологиям Интернета Вещей. Однако вопрос Искусственного Интеллекта в сетях связи является достаточно свежей темой исследований. В рамках данного направления выделяют целый ряд высокоуровневых задач, одними из которых являются задачи идентификации трафика в сетях связи, прогнозирования нагрузки на сеть и ее устройства, эффективного распределения вычислительных ресурсов и задач в сети связи. Данные задачи требуют разработки соответствующих методов на основе технологий ИИ, учитывая особенно требования качества обслуживания для перспективных услуг сетей связи, необходимого быстродействия служебных сервисов и знания прогнозируемых данных. Выше озвученные вопросы и определяют цель, задачи, объект и предмет настоящей диссертационной работы.

**Цель работы и задачи исследования.** Целью диссертационной работы является исследование и разработка методов построения инфраструктуры и предоставления услуг сетей связи с использованием технологий Искусственного Интеллекта.

Для достижения поставленной цели в диссертации последовательно решаются следующие задачи:

- анализ концепций современных и перспективных сетей связи, учитывая долгосрочные до 2030 года перспективы;
- анализ тематики «Искусственный Интеллект в сетях связи», учитывая прогресс стандартизации автономных сетей в МСЭ-Т;
- анализ методов машинного обучения и Больших данных для задач мониторинга и управления трафиком в сетях связи;
- разработка метода идентификации трафика услуг в сетях связи пятого и последующих поколений;
- разработка структуры и метода взаимодействия туманных и граничных вычислений с поддержкой микросервисной архитектуры услуг;
- разработка метода прогнозирования нагрузки на контроллеры программно-конфигурируемых сетей.

**Объект и предмет исследования.** Объектом исследования являются сети связи пятого поколения 5G/IMT-2020, а предметом исследования – методы построения этих сетей и услуг на основе технологий ИИ.

**Научная новизна** полученных результатов состоит в следующем:

- Разработан метод идентификации трафика услуг в сетях связи пятого и последующих поколений, позволяющий исключить внесение дополнительных задержек и изменение структуры потоков.
- Предложена структура взаимодействия туманных и граничных вычислений, отличающаяся тем, что услуги реализуются в виде микросервисной архитектуры программного обеспечения с учетом динаминости пользователей.
- Разработан метод взаимодействия туманных и граничных вычислений, обеспечивающий функционирование микросервисной архитектуры услуг с возможностью живой миграции, позволяющий уменьшить время выполнения функции микросервиса путем рационального выбора устройства Туманных вычислений.
- Разработан метод прогнозирования нагрузки на контроллеры Программно-конфигурируемых сетей, позволяющий исключить зависимость программного обеспечения мониторинга от особенностей АПК производителей.

**Теоретическая и практическая значимость исследования:** Теоретическая значимость диссертационной работы состоит, прежде всего, в обосновании предстоящей децентрализации сетей связи, вычислительных облачных систем

обработки данных и реализации услуг, а также модернизации архитектур программного обеспечения услуг в сторону микросервисного подхода, с появлением сетей связи с ультрамалыми задержками. Кроме того, разработан метод идентификации трафика в программно-конфигурируемых сетях, основанный на аналитике метаданных потоков и алгоритмах машинного обучения, который позволяет не вносить дополнительных задержек в трафик на уровне передачи данных. Разработан также метод прогнозирования нагрузки на контроллер программно-конфигурируемой сети, который позволяет реализовывать мониторинг нагрузки и ее прогноз на уровне приложений программно-конфигурируемой сети, что в свою очередь позволяет уйти от зависимости мониторинговых решений как от производителя АПК контроллера программно-конфигурируемой сети, так и от зависимости операционной системы, на которой развернуто программное обеспечение контроллера. В дополнение предложена структура и метод взаимодействия туманных и граничных вычислений, при этом обеспечивающие поддержку микросервисной архитектуры услуг.

*Практическая значимость* диссертационной работы подтверждается актом внедрения и состоит в создании научно-обоснованных рекомендаций по планированию сетей связи пятого поколения в условиях внедрения сетей с ультрамалыми задержками и Искусственного интеллекта в сетях связи.

Основные теоретические и практические результаты работы были использованы в ФГУП НИИР при выполнении государственных контрактов по научно-техническому и методическому обеспечению выполнения Министерством цифрового развития, связи и массовых коммуникаций функций Администрации связи Российской Федерации в Секторе стандартизации электросвязи Международного Союза Электросвязи (МСЭ-Т) в работах по разработке стандартов для сетей связи пятого и последующих поколений с учетом использования технологий искусственного интеллекта и в Санкт-Петербургском государственном университете телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ) при чтении лекций и проведении практических занятий по дисциплинам «Интернет Вещей и самоорганизующиеся сети», «Искусственный интеллект в сетях и системах связи», «Интернет вещей», «Системы, сети и устройства телекоммуникаций», а также при выполнении НИР «Исследование проблемных вопросов сетевой поддержки перспективных услуг сетей связи 2030, включая телеприсутствие, и путей их решения, в том числе на основе технологий искусственного интеллекта, при подготовке отраслевых кадров».

**Методологические и теоретические основы исследования.** Для решения поставленных в диссертации задач использовались методы машинного обучения, теории оптимизации, имитационного и эмуляционного моделирования. Эмуляционное моделирование проводилось на базе модельной программно-конфигурируемой сети в рамках кафедры сетей связи и передачи данных СПбГУТ.

Имитационное моделирование выполнялось с помощью языка программирования Python версий 2.7, 3.4 с необходимыми программными пакетами и библиотеками обработки данных.

**Основные положения, выносимые на защиту.**

1. Метод мониторинга, идентификации трафика услуг в сетях связи пятого и последующих поколений, основанный на аналитике метаданных потоков и алгоритмах Машинного обучения, позволяющий исключить внесение дополнительных задержек и изменение структуры потоков.

2. Структура и метод взаимодействия туманных и граничных вычислений, основанные на алгоритмах Больших данных, обеспечивающие функционирование микросервисной архитектуры с возможностью живой миграции, позволяющие уменьшить время выполнения функции микросервиса за счет рационального распределения ресурсов на величину до 70%.

3. Метод прогнозирования нагрузки на контроллеры Программно-конфигурируемых сетей на основе технологий Искусственного интеллекта, использующий для оценки нагрузки анализ метаданных служебных потоков, что позволяет исключить зависимость программного обеспечения мониторинга от особенностей АПК производителя.

**Степень достоверности и апробация результатов.** Достоверность результатов диссертации подтверждается корректным применением математического аппарата, результатами имитационного моделирования с использованием пакетов и библиотек языка программирования Python версий 2.7 и 3.4, результатами эмуляционного моделирования на базе модельной программно-конфигурируемой сети лаборатории кафедры ССиПД СПбГУТ, а также обсуждением результатов диссертационной работы на конференциях и семинарах, публикаций основных результатов диссертации в ведущих рецензируемых журналах.

*Апробация результатов.* Основные результаты диссертационной работы докладывались и обсуждались на следующих международных и российских конференциях и семинарах: международный глобальный саммит «Искусственный интеллект во Благо» 2020 года (виртуальный формат, Женева), проводимый МСЭ под эгидой ООН, а также на заседании 13 ИК в рамках 21 вопроса МСЭ-Т 1-12 марта 2021 года (виртуальный формат, Женева), конференциях DCCN (International conference on Distributed Computer and Communication Networks: Control, Computation, Communications) и NEW2WAN (International Conference on Next Generation Wired/Wireless Networking).

**Публикации.** По теме диссертации опубликовано 8 научных работ, из них: 4 статьи в рецензируемых научных журналах, рекомендованных ВАК при Минобрнауки России; 2 статьи в изданиях, индексируемых в международных базах данных; 2 – в других изданиях и материалах научных конференций.

**Соответствие специальности.** Содержание диссертации соответствует пунктам 3, 4, 11 и 14 паспорта специальности 05.12.13 – Системы, сети и устройства телекоммуникаций.

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

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав с выводами по каждой из них, заключения, списка литературы и 3 приложений. Общий объем работы – 194 страницы, из них основного текста 163 страницы. Работа содержит 38 рисунков и 3 таблицы. Список литературы включает 54 источника.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** обоснована актуальность диссертационной работы, описано состояние исследуемой проблемы, сформулированы цели и задачи работы, перечислены основные научные результаты диссертации, определена научная новизна и практическая ценность результатов, описана область их применения, представлены основные положения, выносимые на защиту, приведены сведения об аprobации работы, публикациях по теме работы, описана структура диссертации и ее объем.

**В первой главе** был проведен анализ концепций современных и перспективных сетей связи, в том числе учитывая долгосрочную до 2030 года перспективы. Было определено, что сети 5G/IMT-2020 и последующего поколения, сети связи 2030, обладают такими новыми характеристиками как «ультрамалые задержки», «сверхнадежная связь», «сверхплотные сети», что требует разработки новых методов их построения. Требуется рассмотреть новые технологии и методы построения сетей связи и систем предоставления услуг, такие как Программно-конфигурируемые сети, Виртуализация сетевых функций, граничные вычисления с множественным доступом (МЕС), туманные вычисления, микросервисный принцип проектирования ПО услуг. Требуется изменение принципов управления сетью и услугами через применение технологий ИИ. Были определены особенно актуальные вопросы и сформулированы в формате следующих задач: однозначная эффективная идентификация трафика на уровне передачи данных, прогнозирование нагрузки на контроллеры Программно-конфигурируемых сетей, эффективное распределение вычислений на уровне туманной инфраструктуры и граничных вычислений с множественным доступом.

**Во второй главе** диссертационной работы был проведен анализ и сформулирован вывод о том, что Сети связи с ультрамалыми задержками и требования к ним задают вектор развития сетей и услуг в сторону их децентрализации. Данный класс сетей призван обеспечить работу новых типов

услуг, таких как, Тактильный Интернет, Медицинские сети, Цифровые Аватары, беспилотный транспорт и так далее. Для достижения поставленных требований необходимы изменения не только в технологиях построения сетей связи и услуг, но и изменения в методах мониторинга и управления трафиком, в задачи которого включается в том числе идентификация типа трафика.

Учитывая возможность программируемости SDN сетей и функциональные возможности протокола управления OpenFlow, был разработан метод идентификации трафика в сетях связи на основе метаданных потоков и искусственной нейронной сети. Стоит отметить, что для заранее зарегистрированных потоков трафика не стоит задача его распознавания, например, трафика VoIP или IpTV. В данной работе рассматривается задача выявления трафика Интернета Вещей в сети SDN из общих (заранее не зарегистрированных) потоков.

Проведение исследований предлагаемого метода включает в себя разработку приложения контроллера SDN, его тестирование. Для данных работ необходимо разработать лабораторный стенд Программно-конфигурируемых сетей, с поддержкой северного программного интерфейса, а также развернутых на нем сервисов Интернета Вещей, который будет эмулировать работу одного из приложений Умного города, а также сервис потокового видео.

В данной работе предлагается реализация аналитической системы на сервисном уровне инфраструктуры SDN. Система может быть представлена как АПК, так и ПО контроллера (рисунок 1).

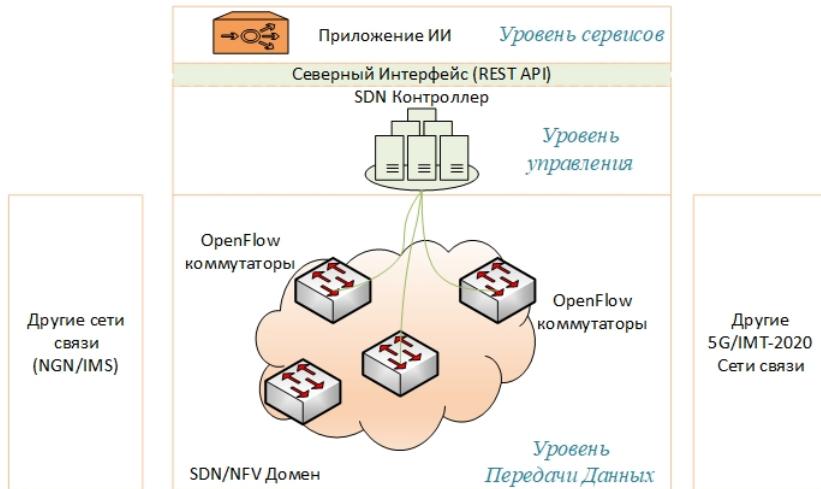


Рисунок 1 – Принципиальная архитектура

Таким образом, появляется возможность переносимости системы, независимости от среды передачи данных и тем более интеграции с уровнем передачи данных. Для аналитической системы все устройства и потоки являются «цифровым объектом», имеющим ряд параметров и функций, представленных набором методов. Такой уровень абстракции позволяет реализовать систему аналитики, которая будет работать с данными о потоках (метаданных) в режиме «На

лету». Таким образом, данная система позволит не вносить дополнительных задержек в трафик, а также каким-либо образом изменять его активность. Можно сказать, что система «наблюдает» за активностью потоков и составляет «общую картину» происходящего в подконтрольной сети связи. Если проанализировать данные, которые отображаются в двух глобальных частях таблицы OpenFlow: Match Field и Actions, то можно прийти к выводу, что на их основе можно составить метамодель потоков.

Одной из важных особенностей этих данных является то, что на основе счетчиков «Byte Count» и «Packet Count» нельзя точно определить точную длину пакета в потоке. Так как за один момент времени счетчики могут быть равны: «Byte Count» – 1500, «Packet Count» – 2. Соответственно, на основе этих данных нельзя точно определить длину каждого из пакетов, зарегистрированных в потоке за промежуток времени:  $\Delta T = 1$  [с].

Таким образом, за произвольный период времени  $\Delta T$ , имея отсчеты значений [Byte Count], [Packet Count], [TimeStamp] возможно составить набор данных с установленной структурой данных, где каждый отсчет отображает мгновенное значение [ByteCount\_delta] и [PacketCount\_delta]. Структура формируемого на основании запросов  $\text{DataSet}_{\text{RQ}}$  с «сырыми» данными (1.1) и формула (1.2) его преобразования в требуемый формат  $\text{DataSet}_{\text{ML}}$  с мгновенными значениями (1.3) приведены ниже.

Пусть,  $\text{PacketCount\_delta} = \text{PC}_{\text{delta}}$ ,

$\text{ByteCount\_delta} = \text{BC}_{\text{delta}}$ , а

$\text{TimeStamp\_deltas} - \text{TS} = 1$  [sec.] = const, тогда:

$$\text{DataSet}_{\text{RQ}} = \begin{array}{ccc} [\text{TimeStamp}] & [\text{ByteCount}] & [\text{PacketCount}] \\ \text{TimeStamp}_{11} & \text{ByteCount}_{12} & \text{PacketCount}_{13} \\ \dots & \dots & \dots \\ \text{TimeStamp}_{N1} & \text{ByteCount}_{N2} & \text{PacketCount}_{N3} \end{array} \quad (1.1)$$

$$\begin{cases} \text{BC\_delta}_{N2} = \text{ByteCount}_{N2} - \text{ByteCount}_{(N-1)2}, & \text{if } N \geq 1 \\ \text{PC\_delta}_{N2} = \text{PacketCount}_{N2} - \text{PacketCount}_{(N-1)2}, & \text{if } N \geq 1 \end{cases} \quad (1.2)$$

$$\text{DataSet}_{\text{ML}} = \begin{array}{ccc} [\text{TimeStamp}] & [\text{ByteCount}] & [\text{PacketCount}] \\ \text{TS} & \text{BC}_{\text{delta}12} & \text{PC}_{\text{delta}13} \\ \text{TS} & \text{BC}_{\text{delta}22} & \text{PC}_{\text{delta}23} \\ \dots & \dots & \dots \\ \text{TS} & \text{BC}_{\text{delta}N2} & \text{PC}_{\text{delta}N3} \end{array} \quad (1.3)$$

При этом расчет суммарных значений параметров за установленный промежуток времени производится по следующим формулам (1.4, 1.5):

$$ByteCount_{\Delta T} = \sum_{N=1}^{N=\Delta T/TS} BC_{delta_{N2}} \quad (1.4);$$

$$PacketCount_{\Delta T} = \sum_{N=1}^{N=\Delta T/TS} PC_{delta_{N2}} \quad (1.5);$$

В качестве искусственной нейронной сети была выбрана рекуррентная нейронная сеть с дополнительными слоями LSTM (с англ. – долговременной краткосрочной памятью). В данной работе глубинный модуль LSTM, как часть искусственной нейронной сети (ИНС) позволяет выявлять закономерности влияния данных предыдущих отсчетов на текущие с учетом большого разброса значений между ними.

Структура обучающего набора данных  $DataSet_{ML_{train}}$  выглядит следующим образом (формула 1.6):

	[TypeOfTraffic]	[TimeStamp]	[ByteCount]	[PacketCount]	
DataSet <sub>ML</sub> =	IoT	TS	BC <sub>delta<sub>12</sub></sub>	PC <sub>delta<sub>13</sub></sub>	
	...	...	BC <sub>delta<sub>22</sub></sub>	PC <sub>delta<sub>23</sub></sub>	...
	Video	TS	BC <sub>delta<sub>N2</sub></sub>	PC <sub>delta<sub>N3</sub></sub>	
	others	TS	BC <sub>delta<sub>(N+1)2</sub></sub>	PC <sub>delta<sub>(N+1)3</sub></sub>	

Архитектура сети содержит 2 полно связанных глубинных уровня LSTM и 2 полно связанных глубинных уровня RNN (Recurrent Neural Network, с англ. Рекуррентная нейронная сеть), каждый из которых содержит 7 скрытых нейронов.

Параметры тренировки ИНС:

- Оптимизатор: Adam;
- Количество эпох обучения: 40;
- Количество образцов на итерацию: 1024;
- Скорость обучения: 0,0025.

В течении процесса обучения нейронной сети приложением наблюдались следующие параметры: Train Accuracy (Точность обучения); Test Accuracy (Точность прохождения теста ИНС); Train loss (Ошибки во время обучения); Test loss (Ошибки во время прохождения теста ИНС).

График, отображающий данные параметры в процессе прохождения эпох обучения, приведен на рисунке 2. В дополнении, на рисунке 3 – приведено приближение (увеличен масштаб части графика на рисунке 2) для более детального отображения разницы между графиками.

Кроме выше приведенных графиков, приведенных на рисунках 2 и 3 – была рассчитана матрица противоречий ИНС (в терминологии ИНС с англ. – Confusion Matrix) обучения искусственной нейронной сети. Получившаяся матрица приведена на рисунке 4 (см. ниже).

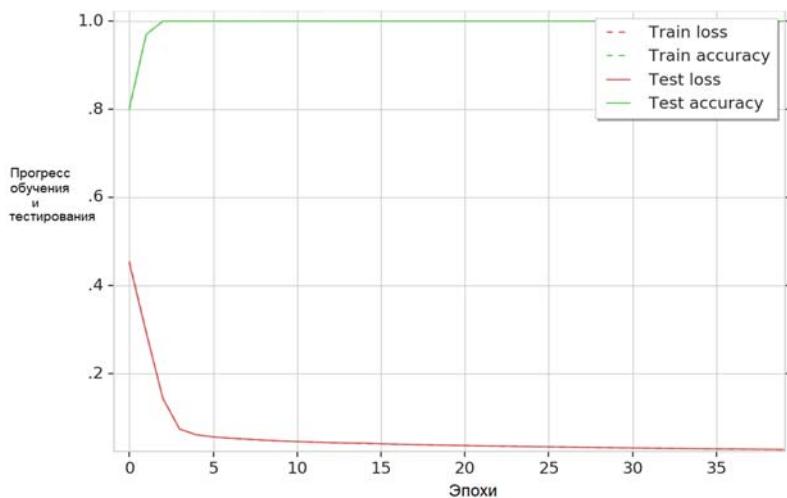


Рисунок 2 – График обучения и тестирования ИНС

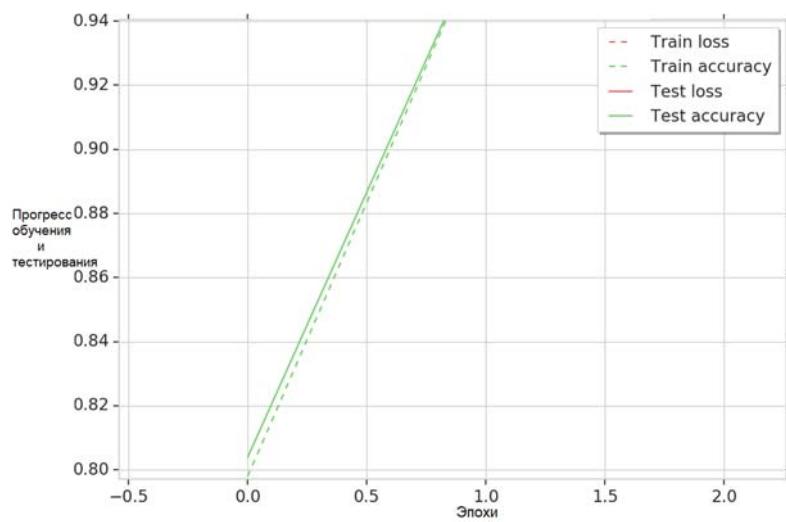


Рисунок 3 – График обучения и тестирования ИНС (масштаб)

На рисунках 2 и 3, где отображен процесс обучения сети, хорошо видно, что для поставленной задачи разработанная ИНС успешно прошла процесс обучения. В результате процесса обучения нейронной сети и проверки ее работы на тестовых наборах данных в обученном состоянии разработанная нейронная сеть может идентифицировать генерируемый поток Интернета Вещей с вероятностью 99,7 %. Результат тестирования показал работоспособность разработанного метода по факту идентификации трафика.

**В третьей главе** диссертационной работы приводится исследование причин и обоснование необходимости перехода к децентрализации облачных вычислений как части вычислительно-сетевой инфраструктуры сетей связи пятого и последующего поколения.

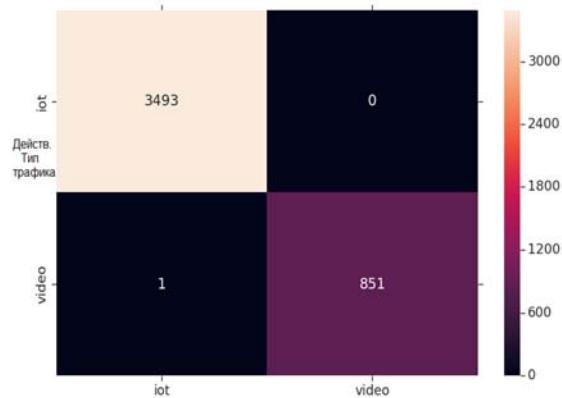


Рисунок 4 – Матрица противоречий (Confusion Matrix)

На рисунке 5 приведен предлагаемый фреймворк/структура взаимодействия распределенных вычислений с микросервисной поддержкой. Данный фреймворк призван объединить облачные структуры многоуровневых пограничных вычислений и туманных вычислений, а также сетевую инфраструктуру, построенную на основе технологий SDN/NFV, учитывая в том числе уровень оркестрации.

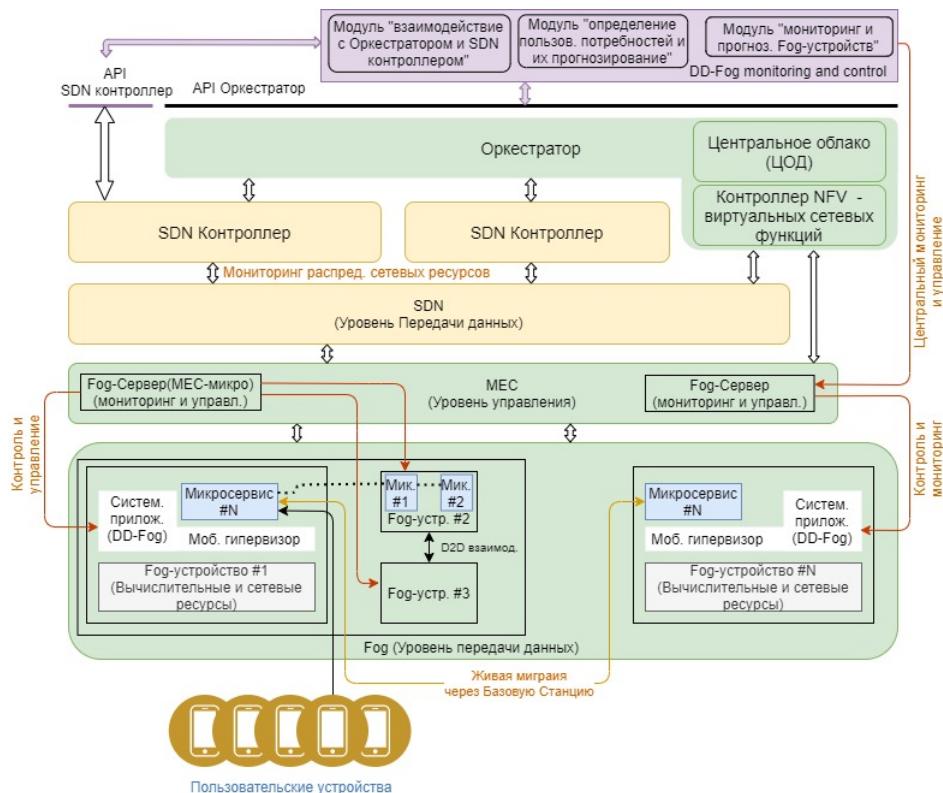


Рисунок 5 – Предлагаемый фреймворк/структура взаимодействия

В данном случае рассматривается подход к управлению через северный инфраструктурный интерфейс сети. Благодаря уровню абстрагирования от физических ресурсов, есть возможность создания служебных приложений-сервисов, которые реализуют системную и бизнес-логику оператора по управлению устройствами инфраструктуры. В данной главе также приводится типовая микросервисная архитектура ПО для высоконагруженных программных продуктов, и пример работы микросервисной услуги.

В данной диссертации рассматриваются: задача определения центра пользователей и задача определения устройства туманных вычислений для последующей миграции микросервиса.

Одним из исходных условий в рамках данной диссертационной работы также является статичность устройств ИВ, смартфонов и других устройств, образующих Fog-зоны. Радиусом кластера будет считаться расстояние между центром и самой удаленной точкой в кластере. В данной работе координаты x, y, z измеряются в метрах. Три плоскости позиционирования были выбраны для учета высоты

нахождения пользовательского устройства с целью корректного формирования соответствующих кластеров.

Работа выбранного алгоритма производится в следующие пять шагов:

1. Задаются входные данные: координаты x, y и z точек. Также задаются количество кластеров и их первоначальные центры.

Присвоение пользователей (пользовательских устройств) к ближайшим центрам кластеров. Для этого используется Евклидова метрика расстояния между точками в пространстве:  $\sqrt{(x - x_k)^2 + (y - y_k)^2 + (z - z_k)^2}$ , где x, y, z – координаты самой точки.

2. Определение центров масс:  $C_{mx} = \frac{\sum_{i=0}^l x_i}{l}$ ,  $C_{my} = \frac{\sum_{i=0}^l y_i}{l}$ ,  $C_{mz} = \frac{\sum_{i=0}^l z_i}{l}$ , где l – количество Fog-устройств, входящих в кластер.

3. Сравнение центров масс и предполагаемых центров кластеров.

4. Если центры масс и предполагаемые центры кластеров равны, то центры кластеров считаются окончательно определенными и все приписанные к ним пользователи помечаются как элементы данного кластера. Если они не равны, то шаги 2–4 повторяются заново, но с предполагаемыми центрами кластеров равными центрам масс, определенным на данной итерации.

Второй исследуемой задачей в рамках данной диссертационной работы является задача определения тех Fog-устройств, которые имеют свободные необходимые вычислительные возможности и удовлетворяют условиям миграции на них микросервиса(-ов) с последующим развертыванием и включением в общую архитектуру. Для задач нахождения глобального максимума или минимума используют так называемые алгоритмы глобальной оптимизации. К одному из таких алгоритмов относятся алгоритмы роевого интеллекта – PSO (на англ. Particle Swarm Optimization).

Таким образом, данный алгоритм определяется следующим образом:

Каждая отдельная частица  $i$  состоит из трех векторов: её положение в D-мерном пространстве поиска  $\bar{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ , лучшая найденная позиция,  $\bar{p}_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ , направленная скорость движения  $\bar{v}_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ . При запуске алгоритма, частицы равномерно, случайным образом инициализируются по всему пространству поиска, при этом скорость частиц также инициализируется случайным образом. Сформированные частицы перемещаются по пространству поиска с помощью довольно простого набора уравнений обновления векторов частицы. Алгоритм обновляет весь рой на каждом временном шаге, обновляя скорость и положение каждой частицы в каждом измерении по следующим правилам:

$$v_{id} = v_{id} + c\varepsilon_1(p_{id} - x_{id}) + c\varepsilon_2(p_{gd} - x_{id}); \quad (2.1)$$

$$x_{id} = x_{id} + v_{id}; \quad (2.2)$$

где  $c$  – постоянная ускорения.  $\varepsilon_1$  и  $\varepsilon_2$  – случайные числа в пределах  $[0; 1]$ ;  $p_{id}$  – лучшее положение из пройденных всеми частицами;  $p_{gd}$  – это положение, найденное любой соседней частицей. Процесс обновления кратко описан в алгоритме, представленном в виде таблицы 1 в формате псевдокода.

Таблица 1. Краткое описание алгоритма

Алгоритм обновления в PSO
<b>for</b> каждого шага $t$ <b>do</b> <b>    for</b> каждой частицы $i$ в рое <b>do</b> <b>        обновить</b> позицию $xt$ используя выражения $(3.3.6.1)$ и $(3.3.6.2)$ <b>        рассчитать</b> фитнес-функцию для $xt$ $f(xt)$ <b>        обновить</b> $pi$ , $pg$ <b>    end for</b> <b>end for</b>

Для решения текущей задачи в настоящей диссертационной работе были определены параметры, описывающие Fog-узел с точки зрения обеспечения качества обслуживания. Глобально задача сформулирована следующим образом: необходимо поддерживать время обслуживания за счет выбора Fog-узла, на который необходимо мигрировать микросервис. Таким образом, минимизированная фитнес-функция выглядит следующим образом:

$$T = \sum_{i=1}^n W_i \cdot TimeSlot_i,$$

где:  $T$  – рассчитываемый параметр в [мс],  $W_i$  – вес соответствующего параметра,  $TimeSlot_i$  – параметры, описывающие состояние Fog-устройства,  $n$  – количество таких параметров.

В данной диссертационной работе используются два параметра:

- $TimeSlot_1$  (задержка распространения),
- $TimeSlot_2$  (время обработки запроса микросервисом).

Так как оба параметра имеют одинаковое влияние на оцениваемый параметр, вес каждого из параметров ( $W_i$ ) равен 0,5, при этом сумма весов не должна превышать 1. Данные параметры рассчитываются в [мс], соответственно, нет необходимости их приводить в одну область значений, как это требуется при оценке параметров с различных областей определения.

В результате, в рамках данной работы фитнес-функция выглядит следующим образом:

$$T = \sum_{i=1}^2 W_i \cdot TimeSlot_i = 0,5 \cdot TimeSlot_1 + 0,5 \cdot TimeSlot_2.$$

Первый параметр определяется через тайм-трекер на уровне фреймворка. А второй параметр отправляется устройством – время обработки задачи.

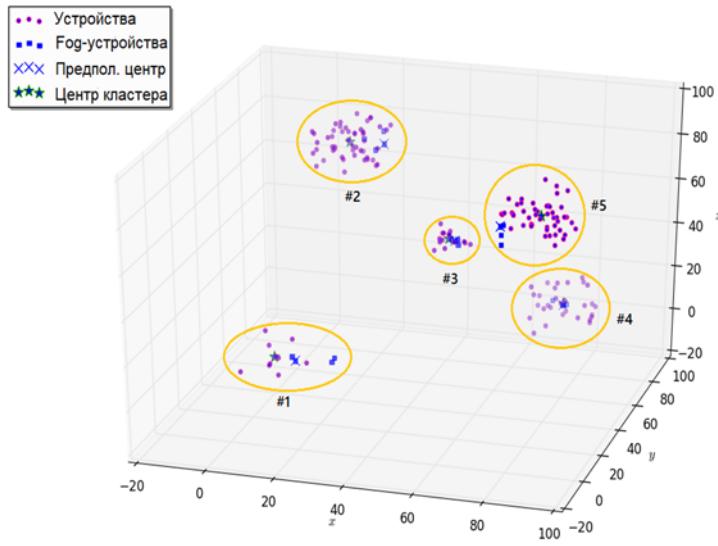


Рисунок 6 – Результат работы алгоритма К-средних

пределах:

$$TimeSlot_1 \in [0.5, 10] \text{ ms}, TimeSlot_2 \in [0.2, 2] \text{ ms}.$$

Результат работы алгоритма PSO, показал, что 13-е Fog-устройство (отмечен зеленым крестиком на рисунке 7) с показателями  $TimeSlot_1 = 1.18 \text{ [ms]}$ ,  $TimeSlot = 0.76 \text{ [ms]}$  и значением фитнесс-функции 0,97, имеет наименьший искомый показатель для дальнейшего размещения микросервиса на нем, а также удовлетворяет необходимым вычислительным и сетевым потенциалом.

Для моделирования предложенных алгоритмов были разработаны программные модели на основе языка программирования Python. Результат работы алгоритма К-средних отображен на рисунке 6.

Согласно заданной фитнесс-функции и описанному алгоритму PSO были получены результаты для 5-го кластера, которые представлены на рисунке 7. Данные были сгенерированы в следующих

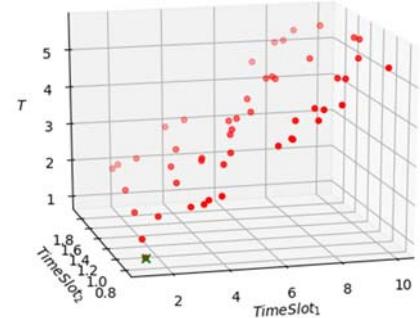


Рисунок 7 – Визуализация данных расчета фитнесс-функции и результата моделирования алгоритма роевого интеллекта – PSO

### Оценка эффекта применения алгоритма выбора Fog-устройства относительно времени выполнения функции микросервисом.

Таким образом, минимальное время выполнения функции на определенном Fog-устройстве:

$$T_{\text{функция,мин}} = 1.18 + 0.76 = 1.94 \simeq 2 \text{ [мс].}$$

Если же осуществлять выбор Fog-устройства равновероятно из всего набора устройств (в рассматриваемой Fog-зоне – 50 устройств), то время выполнения

функции может быть оценено средним значением  $T_{\text{функц.сред.}}$ . Рассчитаем данный параметр по нижеследующей формуле:

$$T_{\text{функц.сред.}} = \frac{1}{N} \sum_{i=1}^N TimeSlot_{1i} + TimeSlot_{2i} = \frac{1}{N} \sum_{i=1}^N T_{\text{функц.}i}.$$

Учитывая сгенерированные данные в модели выше, среднее время выполнения функции микросервисом при равновероятном распределении функции на устройства Fog-зоны составляет  $T_{\text{функц.сред.}} = 6.43[\text{мс}]$ .

Произведем точечную оценку эффекта применения алгоритма выбора Fog-устройства относительно времени выполнения функции микросервисом через отношение полученного минимального времени выполнения функции микросервисом  $T_{\text{функц.мин}}$  к среднему значению выполнения, в случае равновероятного выбора устройства  $T_{\text{функц.сред.}}$ .

$$\frac{(T_{\text{функц.мин}} \cdot 100\%)}{T_{\text{функц.сред.}}} = 30,90\%$$

Таким образом, в третьей главе было определено, что применение алгоритма роботого интеллекта (PSO) в предложенном фреймворке позволяет уменьшить время выполнения функции микросервиса за счет рационального распределения ресурсов на величину до 70% относительно равновероятного распределения ресурсов.

**В четвертой главе** был проведен анализ проблемы мониторинга контроллеров SDN. Приводится вывод о необходимости проведения мониторинга и прогнозирования нагрузки на контроллеры SDN с целью обеспечения устойчивости системы в условиях жестких рамок качества обслуживания предоставляемых услуг. В четвертой Главе был предложен метод мониторинга и прогнозирования нагрузки на контроллер SDN путем.

С целью подтверждения гипотезы о возможности прогнозирования нагрузки на вычислительные ресурсы контроллера SDN на основе метаданных суммы служебных потоков уровня управления был предложен метод проверки через применение многопараметрического корреляционного анализа. Для апробации был разработан стенд (рисунок 8) модельной программно-конфигурируемой сети, где на первой стадии был реализован метод многопараметрического корреляционного анализа с целью проверки выдвинутой гипотезы.

Метод многопараметрического корреляционного анализа достаточно подробно изложен в пункте диссертации 4.2.2 «Исследуемая модель и обоснование работоспособности метода через многопараметрический анализ».

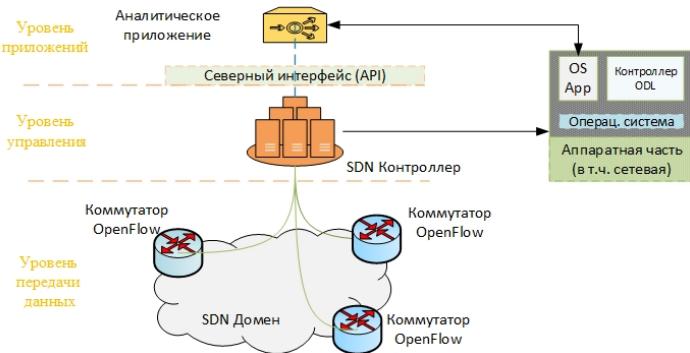


Рисунок 8 – Схема стенда

На рисунке 9 отображены наиболее интересующие результаты относительно взаимосвязи параметров ByteCount и нагрузки CPU. На графике 1 отображена естественная зависимость между двумя параметрами, которые напрямую связаны, для примера оценки параметров. На графике 2 рисунка 9 отображены распределения взвешенных оценок уже различных параметров, которые необходимо оценить с точки зрения зависимости изменений и можно сделать вывод, что согласно графику №2, существует зависимость между изменениями параметра ByteCount суммы служебных потоков и нагрузки на центральный процессор контроллера Программно-конфигурируемой сети.

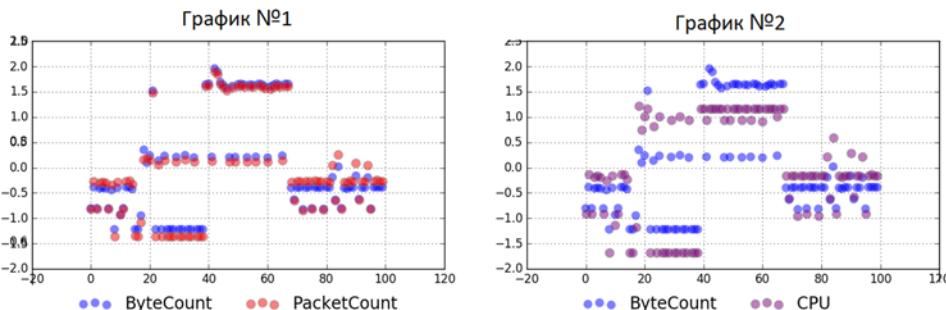


Рисунок 9 – Точечный график распределения взвешенных оценок

Для оценки уровня коррелированности исследуемых величин рассчитывается коэффициент корреляции. В таблице 2 приведены рассчитанные показатели корреляции  $\rho_U$  между величинами, отображенными на графиках рисунка 9.

Таблица 2 – Матрица коэффициентов корреляции

$\rho$	$BtCt PcCt$	$BtCp CpU$
значение	0.98	<b>0.89</b>

Все полученные значения коэффициентов корреляции удовлетворяют следующему свойству  $|\rho_{jk}| > 0$  и  $|\rho_{jk}| < 1$ . На основе данного свойства возможно сделать следующий вывод: существует связь между исследуемыми параметрами, и при этом чем больше величина коэффициента корреляции, тем в большей степени по значениям одного параметра возможно построить прогноз значения другого. Из полученных значений в данной работе интересует значение коэффициента

корреляции между  $BtCt|Cp$  равный 0,89 соответственно. Значение данного показателя позволяет сделать вывод о существовании зависимости между активностью служебных потоков OpenFlow и нагрузкой аппаратной части SDN-контроллера. При этом, стоит отметить, что явной линейной зависимости между ними нет. При этом полученный результат позволяет сделать вывод о том, что предложенный метод мониторинга и прогнозирования нагрузки на контроллер может эффективно работать.

В данной диссертационной работе рассматривается возможность реализации мониторинга нагрузки контроллера SDN при помощи интеллектуальной аналитики метаданных только группы служебных потоков OpenFlow. Схема сегмента с объектами исследования представлена на рисунке 10. Во второй главе была подмечена возможность составления метаданных потоков на основе двух глобальных частей таблицы потоков SDN-коммутатора, а именно: Match Field и Actions.

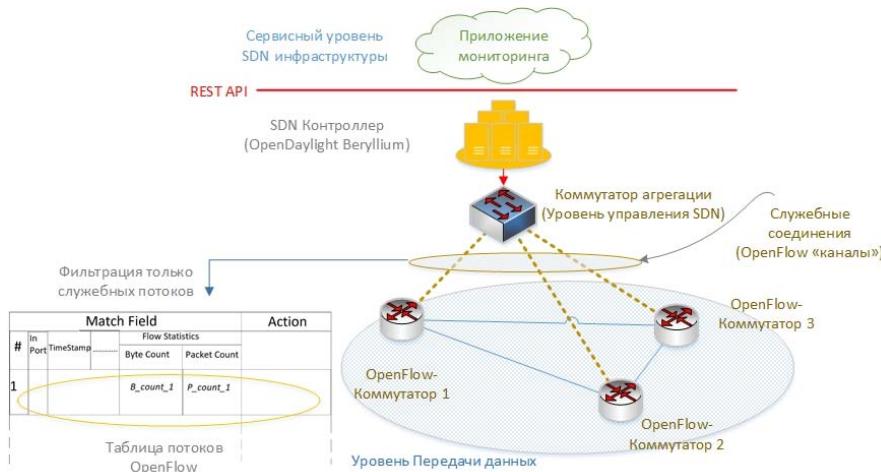


Рисунок 10 – Архитектура сегмента с объектами исследования

В качестве метода прогнозирования данных в четвертой главе было предложено использование ИНС. Архитектура предложенной ИНС отображена на рисунке 11. Нейронная сеть получает на вход данные фиксированной длины, для этого изначальный набор данных делиться на сегменты по 200 строк. Также исходный набор данных разделяется на два набора данных: обучающий и практический (тестовый), в соотношении 8:2. Архитектура нейронной сети является полностью рекуррентной и содержит 4 полносвязанных вложенных уровня нейронов, каждый из которых состоит из 10 нейронов.

Параметры обучения Искусственной Нейронной Сети:

- Оптимизатор: Adam.
- Количество эпох: 20.
- Количество образцов на итерацию: 1024.
- Скорость обучения: 0,0025.

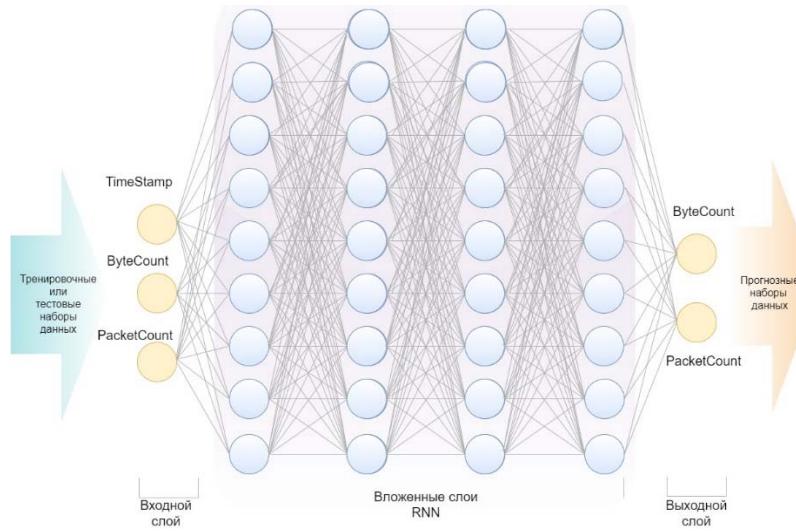


Рисунок 11 – Архитектура ИНС

В процессе обучения Искусственной Нейронной Сети (оценки точности прогнозирования нейронной сети), в качестве параметра оценки работы, наблюдался параметр MSE – Mean Square Error (с англ. среднеквадратическая ошибка). Изменение значения параметра MSE приведено на рисунке 12:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

где  $Y$  – вектор наблюдаемых значений прогнозируемой переменной,  $\hat{Y}_i$  – вектор спрогнозированных значений. Другими словами, MSE это среднее квадратов ошибок прогноза. В результате практического эксперимента было определено, что выбранная архитектура и параметры нейронной сети удовлетворяют для формирования прогнозирования нагрузки. В конечном итоге, получились следующие параметры:

$$MSE_{Train} = 4,54 * 10^{-6},$$

$$MSE_{TEST} = 1,5 * 10^{-5}.$$

Также, в процессе обучения нейронной сети, отслеживался процесс ее обучения на формирование прогнозов. Для наглядности процесса, строился график реальных и спрогнозированных значений нейронной сети на промежутке данных в 20 тыс. отсчетов. На рисунке 13 приведены несколько снимков экрана в процессе обучения нейронной сети, отображающие ее прогресс. В каждой из областей (графики № 1–4), приведенных на рисунке 13, есть два графика синего и зеленого цвета, отображающие реальные значения и спрогнозированные.

График №1 на первой линии рисунка 13 с подписью (Epoch 0, Batch 0) отображает реальный и первый прогноз. На графике видно, что спрогнозированные значения сильно отличаются от реальных. То есть график зеленого цвета сильно ниже основного – синего графика. Нижние два графика (график №3 и график №4),

с подписями (Epoch 19, Batch 250), (Epoch 19, Batch 300) соответственно отображают последние шаги обучения сети. По данным графикам видно, что Искусственная Нейронная Сеть обучилась и теперь может корректно прогнозировать активность суммы служебных потоков OpenFlow в реальной программно-конфигурируемой сети на уровне управления. Видно, что на последнем (нижнем правом графике №4), реальный график визуально совпадает с прогнозированным.

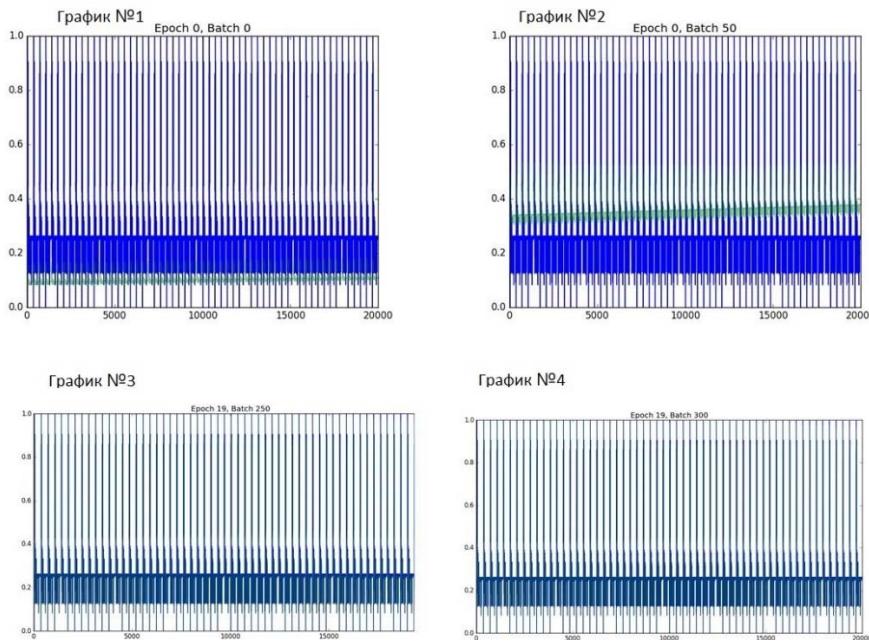


Рисунок 13 – Процесс обучения Искусственной Нейронной сети

## ЗАКЛЮЧЕНИЕ

В ходе выполнения настоящей диссертационной работы автором получены следующие основные результаты:

1. Был проведен анализ концепций современных и перспективных сетей связи, в том числе учитывая долгосрочные до 2030 года перспективы. Особенno был выделен класс сетей связи, имеющий такие характеристики, как «ультрамалые задержки», «сверхнадежная связь», «сверхплотные сети».

2. Так как определена необходимость обеспечения в характеристиках, указанных в п. 1, требуется рассмотреть новые технологии и методы построения сетей связи и предоставления услуг, таких как: Программно-конфигурируемые сети, Виртуализация сетевых функций, граничные вычисления с множественным доступом (МЕС), туманные вычисления (Fog), а также методы разработки и реализации программного обеспечения услуг, в частности – микросервисный архитектурный подход.

3. Было установлено, что сети связи с ультрамалыми задержками и требования к ним формируют вектор развития сетей и услуг в сторону их децентрализации.

4. Кроме означенных в п. 2 методов и технологий построения сетей связи пятого и последующих поколений, требуется изменение принципов управления сетью и услугами через применение технологий Искусственного Интеллекта. В данном направлении установлено, что особенно актуальными вопросами являются: однозначная идентификация трафика на уровне передачи данных, прогнозирования данного трафика, прогнозирование нагрузки на контроллеры Программно-конфигурируемых сетей, и эффективное распределение вычислений на уровне глубоко интегрированной туманной инфраструктуры и пограничных вычислений с множественным доступом. В диссертационной работе предложены методы решения данных задач.

5. В данной диссертации было проведено исследование современных архитектур относительно построения перспективных услуг в сетях связи. Было определено, что микросервисный архитектурный подход является одним из самых актуальных и реализующих необходимые требования к современному программному обеспечению, в том числе к программному обеспечению современных и перспективных услуг связи. Особенno обращается внимание на услуги сетей связи с ультрамалыми задержками и сверхнадежностью.

6. Была предложена структура/фреймворк взаимодействия распределенных туманных вычислений и пограничных вычислений с поддержкой микросервисных услуг.

7. В рамках предложенного фреймворка были рассмотрены две задачи, а именно: определение центра скопления пользователей, а также определение Fog-устройства, которое будет выбрано для живой миграции микросервиса услуги. На основе анализа были определены математические алгоритмы решения поставленных задач, такие как: К-средних и Роевого Интеллекта (PSO), которые в свою очередь входят в класс алгоритмов, относящихся к Искусственному Интеллекту. Полученные данные в результате программного моделирования показывают работоспособность предложенных алгоритмов с точки зрения фреймворка и перспективность развития данного решения.

8. Был предложен метод мониторинга и нагрузки контроллеров Программно-конфигурируемых сетей путем аналитики метаданных суммы служебных потоков OpenFlow уровня управления.

9. Для апробации предложенного метода прогнозирования был разработан стенд модельной программно-конфигурируемой сети, где на первой стадии был реализован метод многопараметрического корреляционного анализа с целью проверки гипотезы об использовании метаданных служебных потоков, результат тестирования которого подтвердил работоспособность предложенного метода. На

второй стадии был реализован метод прогнозирования потоков служебного трафика в программно-конфигурируемой сети, позволяющий оценить нагрузку на контроллеры SDN. В результате обучения и тестирования разработанной ИНС, значение MSE в тестовом варианте составил  $MSETEST = 1,5 \cdot 10^{-5}$ , что является достаточно высоким показателем. Стоит также отметить, что предложенный метод позволяет исключить зависимость мониторингового ПО от особенностей АПК-решений производителей.

## СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации в рецензируемых научных изданиях, рекомендованных ВАК

1. Волков, А.Н. Идентификация трафика сервисов в сетях связи IMT-2020 и последующего поколения на основе метаданных потоков и алгоритмов машинного обучения / А.Н. Волков, А.Е. Кучерявый // Электросвязь. – 2020. – № 11. – С. 21-28.
2. Волков, А.Н. Искусственный интеллект в сетях связи пятого и последующих поколений / А.Н. Волков, А.С. Мутханна, А.Е. Кучерявый // Электросвязь. – 2021. – № 1. – С. 17-22.
3. Волков, А.Н. Структура распределенной динамической вычислительной системы туманных вычислений для микросервисов (DD-Fog) // Электросвязь. – 2021. – № 7. – С. 34-43.
4. Волков, А.Н. Метод прогнозирования нагрузки на контроллеры SDN с помощью технологий Искусственного Интеллекта / А.Н. Волков, А.Е. Кучерявый // Электросвязь. – 2021. – № 2. – С. 31-38.

### Публикации в изданиях, входящих в Web of Science и Scopus

5. Artem, V., Ateya, A. A., Muthanna, A., & Koucheryavy, A. (2019). Novel AI-Based Scheme for Traffic Detection and Recognition in 5G Based Networks. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 11660 LNCS, pp. 243–255). Springer Verlag. [https://doi.org/10.1007/978-3-030-30859-9\\_21](https://doi.org/10.1007/978-3-030-30859-9_21)
6. Volkov A., Proshutinskiy K., Adam A.B.M., Ateya A.A., Muthanna A., Koucheryavy A. (2019) SDN Load Prediction Algorithm Based on Artificial Intelligence. In: Vishnevskiy V., Samouylov K., Kozyrev D. (eds) Distributed Computer and Communication Networks. DCCN 2019. Communications in Computer and Information Science, vol 1141. Springer, Cham. [https://doi.org/10.1007/978-3-030-36625-4\\_3](https://doi.org/10.1007/978-3-030-36625-4_3)

### Публикации в других изданиях

7. Волков А. Н., Мутханна А. С. А., Кучерявый А. Е. Сети связи пятого поколения: на пути к сетям 2030 // Информационные технологии и телекоммуникации. 2020. Том 8. № 2. С. 32–43. DOI 10.31854/2307-1303-2020-8-2-32-43.
8. Koucheryavy, A., Muthanna, A., Volkov, A.: AI/machine learning for ultra-reliable low-latency communication, ITU News Magazine, No.5, December 2020. pp. 62-65.

Подписано в печать 13.10.2021. Формат 60×84 1/16.

Печ. л. 1,0. Тираж 100 экз.

Отпечатано в СПбГУТ, 193232, Санкт-Петербург, пр. Большевиков, д. 22, корп. 1