

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
КАФЕДРА АНАЛИЗА ДАННЫХ

Направление подготовки/специальность: 01.04.02 Прикладная математика
и информатика (магистратура)

Направленность (профиль) подготовки: Анализ данных

Управление параметрами передачи multi-link Wi-Fi устройства при помощи RL модели

(магистерская диссертация)

Студент:

Котюшев Михаил Юрьевич

Научный руководитель:

Климаков Андрей Владимирович,
канд. физ.-мат. наук

Москва, 2021

Содержание

1	Введение	4
1.1	Описание предметной области	4
1.2	Организация работы	6
2	Окружение, постановка задачи и используемая модель	8
2.1	Верхнеуровневая постановка задачи	8
2.2	Реализация мултилинк архитектуры	8
2.3	Параметры сценария	11
2.4	RL-окружение	13
2.5	RL-агент	15
3	Обучение модели	18
3.1	Алгоритм и параметры обучения	18
3.2	Обучение RL-агента	20
4	Тестирование и обсуждение результатов	22
4.1	Бейзлайн	22
4.2	Метрики	22
4.3	Тестовые сценарии	23
4.4	Результаты	25
5	Заключение	27
	Приложение	30
A	Exponentially weighted moving average	30

В	Параметры симуляций из тренировочной выборки	31
	Список литературы	34

Аннотация

В данной работе реализуется RL-модель для оптимальной балансировки нагрузки между линками multi-link Wi-Fi устройства, исследуется качество полученного решения в сравнении с бейзлайном с точки зрения метрик качества Quality of Experience.

В сетевом симуляторе ns-3 реализована архитектура Multi-link operation в соответствии с опубликованным документом спецификации находящегося в разработке стандарта Wi-Fi 7.

Полученная модель показывает лучшее качество по сравнению с бейзлайном в широком диапазоне тестовых сценариев с точки зрения 99% квантилей задержек сети Wi-Fi и задержек показа IPTV кадра, а также с точки зрения максимального поддерживаемого битрейта IPTV передачи без потерь пакетов. В отдельных сценариях достигается уменьшение 99% квантилей задержек показа кадра IPTV до 10 раз, увеличение максимального передаваемого битрейта без потери пакетов до 80 МБпс по сравнению с бейзлайном.

Данная работа проведена при поддержке компании Huawei, предоставившей симуляционную платформу и вычислительные мощности.

1 Введение

1.1 Описание предметной области

Multi-link Wi-Fi

Стандарт IEEE 802.11 [1], описывающий базовые протоколы физического и MAC уровня модели OSI для Wi-Fi, оставляет широкие возможности для улучшения пропускной способности сети и сетевых задержек путем настройки параметров передачи. Теоретические и прикладные исследования различных подходов к оптимизации проводятся для таких задач, как resource allocation [2, 3], rate control [4].

В ранних версиях стандарта до Wi-Fi 6 включительно для одновременной передачи между двумя станциями использовался только один *линк* — логическая ассоциация передающего и принимающего устройства [5]. Стандарт Wi-Fi 7 (так же известный как IEEE 802.11be), разработка которого началась в январе 2019 года [6], вводит поддержку Multi-Link Operation (MLO) — одновременного использования для передачи нескольких линков. Отсюда естественным образом возникает задача балансировки нагрузки между несколькими доступными для передачи линками для достижения оптимальных характеристик передачи, таких как задержка и пропускная способность.

Приложения reinforcement learning

Reinforcement Learning (RL) используется для решения прикладных задач оптимального управления в различных областях, таких как управление беспилотными автомобилями [7], спортивные прило-

жения [8, 9], индустриальная оптимизация ресурсов [10] и т. д.

В сетевых областях, таких как LTE и Wi-Fi, RL также используется для решения задач оптимального управления. Так, в [11] авторы решают задачу resource allocation для базовой станции LTE при помощи LSTM Deep Q-network (DQN) [12], выбирая в качестве награды суммарную пропускную способность по всем доступным частотам. Эксперименты были проведены в симуляторе с использованием реальных данных о трафике.

В [13] рассматривается задача выбора оптимального канала и мощности сигнала для Vehicle-to-vehicle LTE передачи в окружении с многими агентами. Наградой выступает суммарная пропускная способность всех задействованных агентов с ограничением на максимальную задержку передачи, в качестве состояния — Channel state information (CSI) каналов, решения всех агентов с предыдущего шага, а также оставшееся время до нарушения ограничения на максимальную задержку. Эксперименты были проведены в симуляторе.

Работа [14] рассматривает задачу выбора оптимального data rate для Wi-Fi передачи между перемещающимися устройствами. Для решения задачи используется Q-learning, в качестве награды выступает функция от пропускной способности и соотношения потерянных пакетов к успешно доставленным. Эксперименты были проведены в сетевом симуляторе OMNeT++.

Симуляционная платформа

Реализация лабораторного окружения для сбора данных, тестирования алгоритмов требует постоянных финансовых и временных затрат на поддержку лабораторного оборудования в актуальном состоянии. При этом программное обеспечение для сетевых устройств,

таких как роутеры, ретрансляторы, мобильные станции, IPTV приемники и т. д. почти всегда является проприетарным и доступа к исходному коду сетевого стека нет, что дополнительно затрудняет реализацию и тестирование сетевых алгоритмов в реальных устройствах. Поэтому для тестирования алгоритмов транспортного уровня, уровней IP и MAC используются сетевые симуляторы. В данной работе используется симулятор Network Simulator 3 (ns-3) [15] с открытым исходным кодом, поддерживающий стандарт Wi-Fi 6 и подходящий для реализации отдельных технологий стандарта Wi-Fi 7 (в частности, MLO).

Данная работа посвящена разработке RL-модели для оптимальной балансировки нагрузки между линками multi-link Wi-Fi устройства и исследованию качества модели в различных сценариях с точки зрения сетевых задержек и пропускной способности.

1.2 Организация работы

Работа организована следующим образом. В главе 2 дается верхнеуровневая постановка задачи с точки зрения поддержки передачи с низкими задержками сети, описывается симуляционная платформа и RL-окружение. Также в этой главе дается постановка задачи RL и описывается архитектура RL-агента. В главе 3 описывается процесс обучения RL-агента. Глава 4 посвящена обсуждению методологии тестирования, а так же обсуждению достигнутых результатов. В заключении дается краткий обзор проделанной работы, приводятся основные результаты и список планируемых улучшений. В приложении находится таблица параметров сценариев из обучающего набора,

а так же определение схемы усреднения статистик для наблюдений.

Ниже приведена таблица расшифровок аббревиатур, используемых в данной работе.

Таблица 1: Таблица расшифровок аббревиатур

Термин	Значение
AP	Access point
CSI	Channel state information
DQN	Deep Q-network
EDCA	Enhanced distributed channel access
ER	Experience replay
EWMA	Exponentially weighted moving average
EWMC	Exponentially weighted moving counter
FIFO	First in first out
IP	Internet protocol
IPTV	Internet protocol television
LSTM	Long short-term memory
LTE	Long Term Evolution
MAC	Medium access control
MLD	Multi-link device
MLO	Multi-link operation
OSI	Open systems interconnection (сетевая модель)
PTT	Packet transmission time
QoE	Quality of Experience
RL	Reinforcement learning
SN	Sequence number
STA	Station

2 Окружение, постановка задачи и используемая модель

2.1 Верхнеуровневая постановка задачи

В отличие от предыдущего стандарта Wi-Fi 6, ставившего главной целью увеличение пропускной способности [16], стандарт Wi-Fi 7 в дополнение к этому ставит целью достижение низких задержек для поддержки приложений реального времени [6]. Таким образом, целью данной работы является разработка решения для балансировки нагрузки устройства Wi-Fi с несколькими линками, улучшающего как характеристики сетевых задержек, так и пропускную способность устройства.

Более детальное описание используемого типа трафика и сценария дано в параграфе 2.3 ниже в этой главе. Описание бейзлайна и метрик, используемых для оценки качества решения, дано в главе 4.

2.2 Реализация мултилинк архитектуры

Для реализации модели мулти-линк Wi-Fi передачи и соответствующего RL окружения бы выбран дискретно-событийный сетевой симулятор ns-3 [15], поддерживающий сетевой стек от уровня приложения до физического уровня модели OSI для таких сетевых технологий, как Ethernet, cellular networks (LTE), WiMAX и, в частности, Wi-Fi. ns3 доступен для использования в исследовательских и обучающих целях под лицензией GNU GPLv2 как проект с открытым исходным кодом [17].

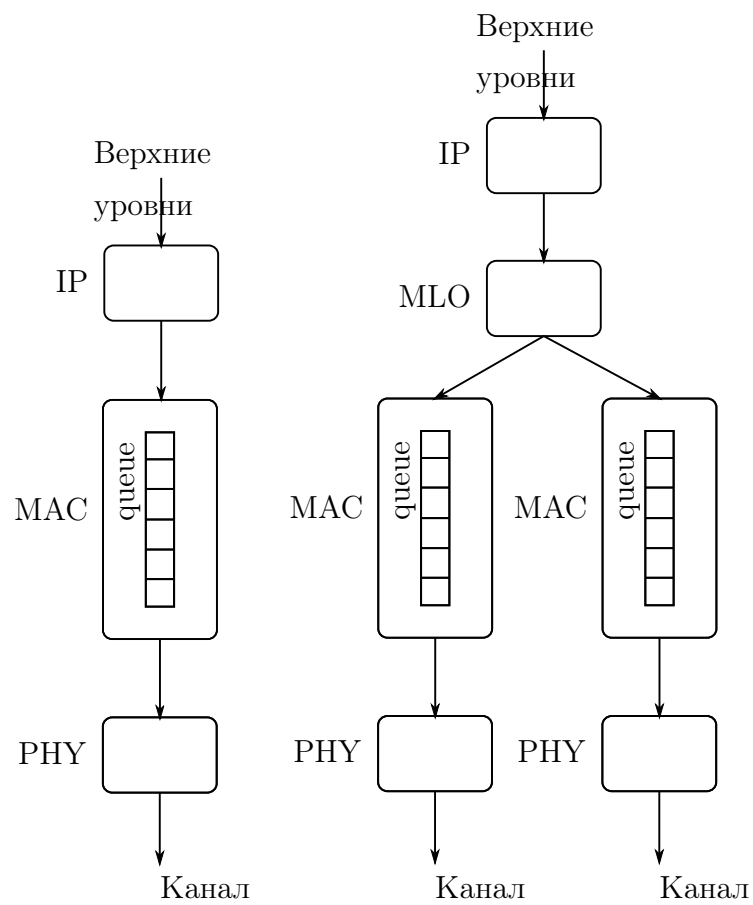


Рис. 1: Сравнение архитектуры передающей стороны 802.11ah станции (слева) и MLD с реализованной архитектурой MLO (справа).

Ввиду того, что MLO — часть стандарта Wi-Fi 7 [18], находящегося во время выполнения данной работы в разработке, полноценной реализации MLO в ns-3 нет. Следуя опубликованному документу спецификации IEEE 802.11be [19], была реализована следующая архитектура MLO (см. блок схему архитектуры на рис. 1) с отдельными очередями пакетов MAC-уровня:

- каждый MLD представляет собой агрегацию нескольких (от 1 до 3) 802.11ax станций, имеющих общий сетевой стек с IP уровня и выше, называемых линками (см. [19] глава 2.1)
- по аналогии с MAC-адресом в 802.11ax, каждому MLD присваивается уникальный MLD-адрес (см. [19] глава 2.3)
- станции одного MLD имеют одинаковый MAC-адрес, равный MAC-адресу MLD
- вводится новый промежуточный уровень модели OSI — MLO — располагающийся между IP и MAC и поддерживающий следующий функционал
 - на MLO уровне передающей стороны каждый получаемый с IP уровня пакет направляется в один из MAC-уровней агрегированных станций
 - по аналогии со сквозной нумерацией MAC-пакетов 802.11ax (MAC sequence number) [5], вводится общая сквозная нумерация пакетов MLD (MLO SN) и каждому приходящему с IP уровня пакету присваивается следующий по порядку MLO SN (см. [19] параграф 2.5.2)

- на приемной стороне реализуется Reordering Buffer (см. [19] глава 2.5), в котором производится восстановление порядка пакетов согласно MLO SN

Таким образом, на уровне MLO возникает задача выбора оптимального линка для передачи пакета, подход к решению которой подробно описан в нижеследующих параграфах.

2.3 Параметры сценария

В сценариях для обучения и тестирования используется топология, изображенная на рис 2. Выбирая различные значения битрейта интерферирующих сетей, можно получить различные загруженности для разных линков, а следовательно — простор для оптимизации над бейзлайном (см. 4.1).

Параметры сети

Следующие параметры были использованы для каждого из линков во всех тренировочных и тестовых симуляциях:

- 5 GHz канал
- ширина Wi-Fi канала 20 МГц
- число передающих антенн на AP: 2, принимающих на STA: 2

Модель трафика

В качестве трафика была выбрана модель IPTV, генерирующая поток пакетов, схожий по характеристикам с потоком пакетов от реального H.264 кодека [20]: редкие и тяжеловесные I-кадры, частые

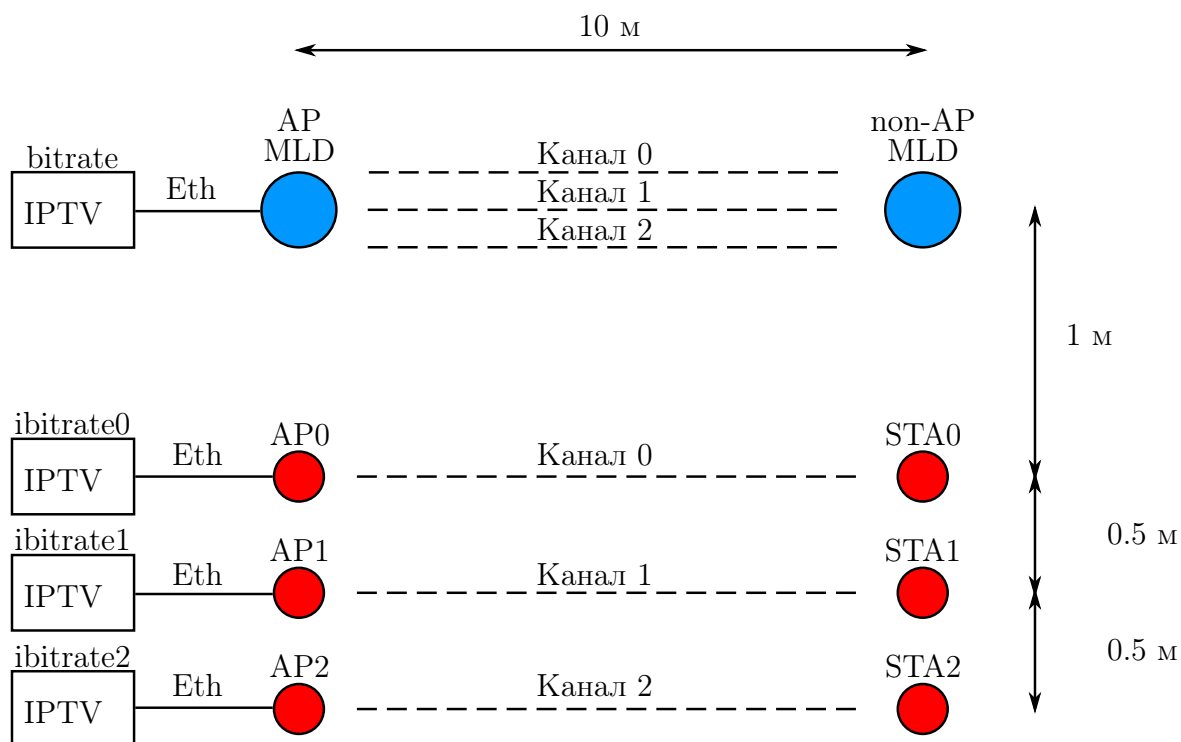


Рис. 2: Топология сценария симуляции.

Основная сеть (выделена синим): AP MLD с тремя линками передает IPTV трафик с битрейтом bitrate на non-AP MLD, расстояние между AP MLD и non-AP MLD равно 10 м.

Интерферирующие сети (выделены красным): на каждом из линков AP_i с одним линком выполняет передачу IPTV-трафика с битрейтом ibitrate_i на STA_i , расстояние между AP и STA также равно 10 м.

и легкие Р- и В-кадры. Параметры модели: частота кадров 25 FPS, соотношение I/P/V кадров — 1/4/12.

2.4 RL-окружение

В этом параграфе приведены описания пространства наблюдений и соответствующих признаков, а также пространства действий и награды.

Пространство наблюдений

Запись наблюдения и пересылка его агенту вызывается приходом с IP-уровня нового пакета для передачи. Пространство наблюдений описывается следующими признаками (здесь все временные признаки даны в наносекундах):

- EWMA (см. приложение А) задержки передачи пакета с MLO NS j на линке i : время между вставкой пакета в MAC очередь i -го линка (момент времени $t_{enqueued}^{ij}$) и успешной передачей пакета (момент времени t_{acked}^{ij})

$$L_{ij} = t_{acked}^{ij} - t_{enqueued}^{ij}$$

- EWMA packet transmission time (PTT) с MLO NS j на линке i — удельной задержки на число пакетов в очереди:

$$PTT_{ij} = \frac{t_{acked}^{ij} - t_{enqueued}^{ij}}{N_{ij} + 1}$$

где N_{ij} — число пакетов в очереди в момент $t_{enqueued}^{ij}$

- EWMA значения битрейта (байт / с), вычисленного при выборе схемы кодирования на физическом уровне для линка i

- EWMA времени, проведенного физическим уровнем линка i в состоянии BUSY (передача другого устройства на канале, не предназначенная данному устройству)
- EWMA времени, проведенного физическим уровнем линка i в состоянии IDLE (не находится в состоянии BUSY, не осуществляет передачу и не осуществляет прием)
- число пакетов в очереди i -го линка
- EWMC числа пакетов, переданного i -го линком

Признаки нормализуются коэффициентами, указанными в таблице 2.

Таблица 2: Коэффициенты нормализации признаков

№	Признак	Коэффициент
1	Задержка Wi-Fi	5e9 нс
2	РТТ	5e9 нс
3	Битрейт	1e8 bps
4	Длина BUSY состояния	1e7 нс
5	Длина IDLE состояния	1e7 нс
6	Число пакетов в очереди	1e3
7	Число переданных пакетов	1e3

Таким образом, размерность пространства наблюдений равна $7 * N$, где N — число линков MLD, доступное для передачи пакета.

Пространство действий

В качестве пространства действий рассматривается дискретное

пространство возможных индексов линка, выбранного для передачи данного пакета.

Награда

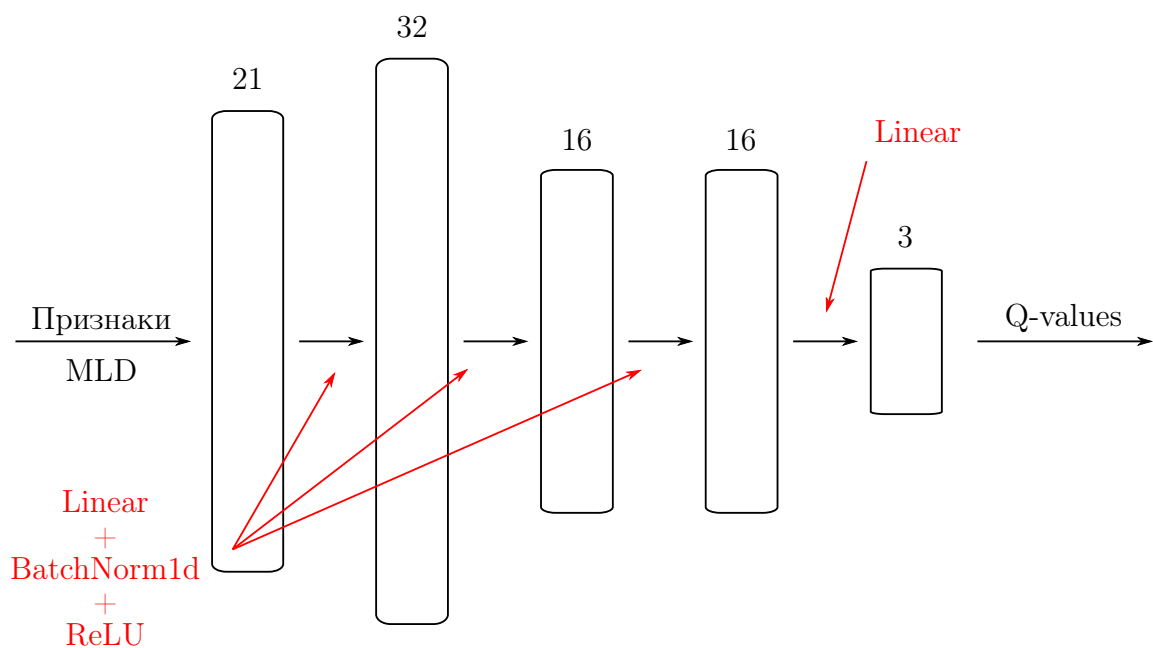
С учетом верхнеуровневой постановки задачи минимизации задержек сети (см. параграф 2.1), различные функции наград могут быть рассмотрены: линейные [21], агрегированные значения задержек, логарифмические, экспоненциальные. Применение монотонной функции не несет существенных изменений, поэтому принципиальных отличий между такими функциями наград нет.

В качестве награды при выборе для передачи пакета j линка i была выбрана логарифмическая награда $R_{ij} = -\log_{10} \frac{L_{ij}}{10^6}$. Здесь задержка переведена в миллисекунды перед логарифмированием для удобства восприятия.

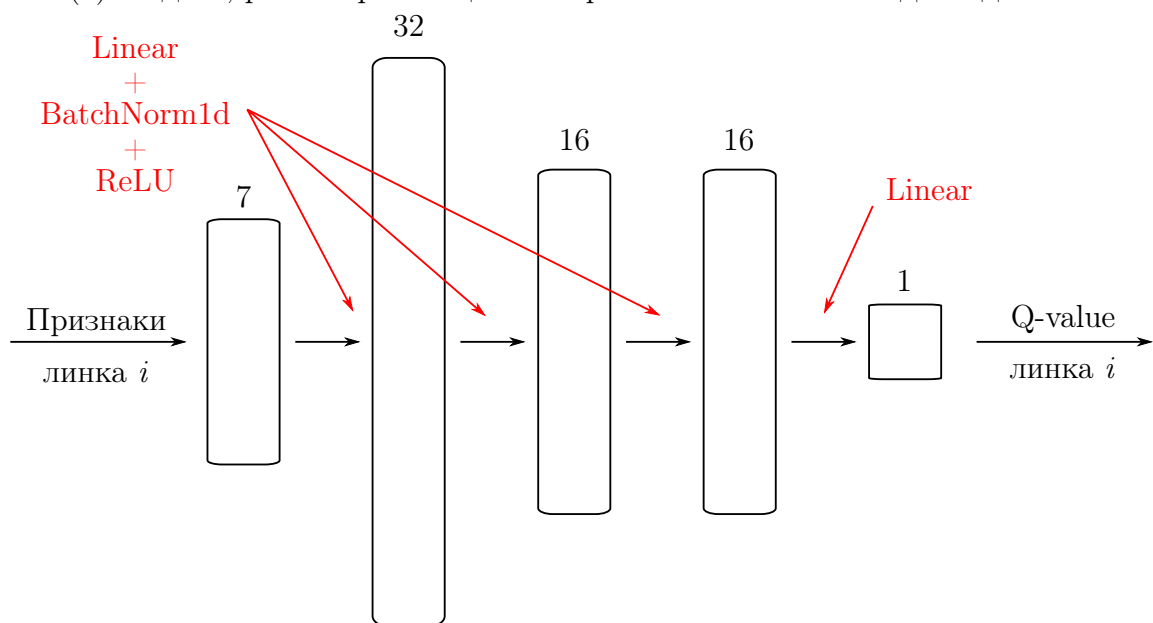
Преимуществом такого выбора награды, в отличие от агрегированных статистик задержки, таких, как средняя задержка по окну или статистик уровня приложения, таких, как задержка показа кадра IPTV, является однозначное соответствие между наградой и совершенным действием, что исключает проблему Credit Assignment.

2.5 RL-агент

В качестве базовой нейросетевой архитектуры была выбрана полносвязная DQN-сеть. Было рассмотрено две схожие архитектуры RL-агента: первая, берущая в рассмотрение полное описание наблюдения для всех линков MLD (см. рис. 3а), и вторая — "сиамская" с общими весами, оценивающая наблюдения каждого линка по отдельности и выдающая Q-value для текущего линка (см. рис. 3б).



(a) Модель, рассматривающая все признаки MLD как входные данные.



(b) Модель, рассматривающая каждый линк отдельно от остальных. Применяется последовательно к каждому доступному для передачи линку.

Минусом архитектуры, принимающей полное описание MLD, является зависимость размера модели от числа линков, а следовательно — ограничение области применения: модель может быть использована только в MLD устройствах с таким же количеством линков, которое использовалось при обучении агента. Принимая данный фактор во внимание, в качестве финальной архитектуры агента была выбрана "сиамская" архитектура с общими весами.

В главе 3 описан процесс обучения выбранной архитектуры, а также полученные результаты.

3 Обучение модели

3.1 Алгоритм и параметры обучения

Описанная в параграфе 2.5 RL-модель была реализована на языке Python 3 с использованием фреймворка глубинного обучения PyTorch [22]. Для интеграции RL-модели и окружения симулятора ns-3 (выполнение сбора наблюдений, наград и отправка выбранных агентом действий) был использован пакет ns3-gym [23]. Диаграмма обмена данными изображена на рис. 4.

В качестве награды использовалась функция от величины задержки пакета, которая становится известной только после успешной передачи пакета или потери пакета по критическому порогу времени нахождения в очереди в 500 мс. Из этого следует необходимость использовать механизм отложенной награды. Сразу после выполнения действия запись с наблюдением помещаются в Experience Replay (ER) буфер без приложенного значения награды. В интервале между событиями принятия решения агентом окружение собирает значения наград для пакетов, которые были доставлены или потеряны в течение этого интервала, и в момент следующего принятия решения отправляет эти данные агенту (на рис. 4 собранные отложенные награды помечены как $\{R\}$), который обновляет соответствующие записи в ER буфере.

Для эффективного выполнения ER буфером операций удаления старейшей записи (при переполнении буфера) и поиска по ключу (выставление отложенной награды ранее полученным записям) используется две структуры данных: FIFO очередь из ключей с мак-

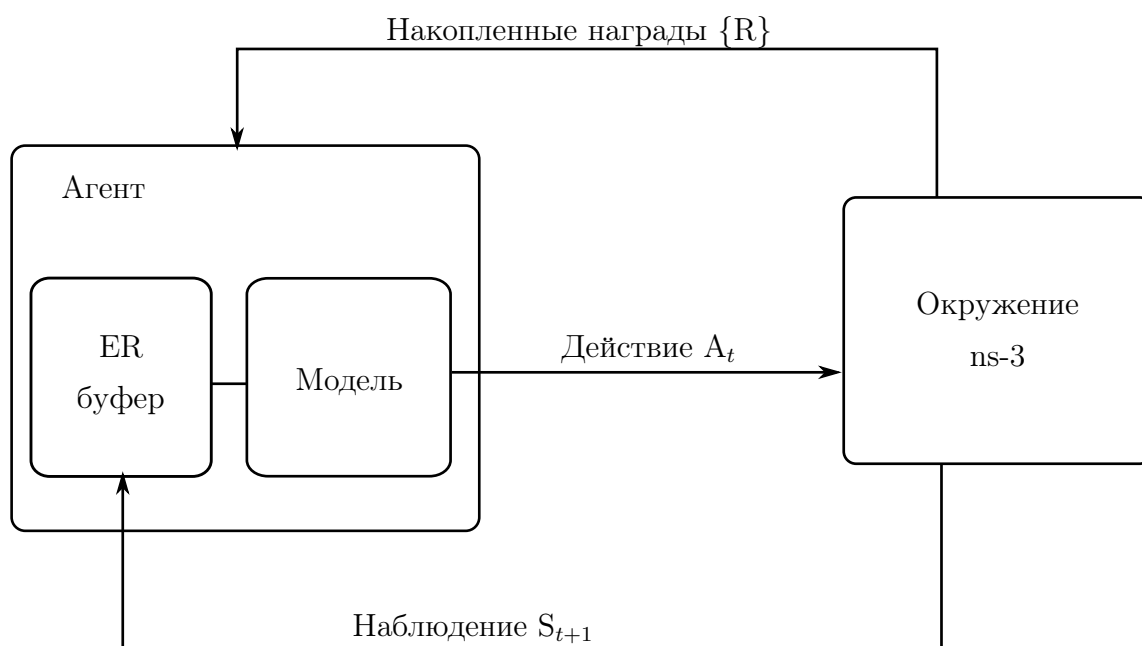


Рис. 4: Схема взаимодействия окружения ns-3 и RL-агента. Окружение сообщает агенту свое состояние, а так же накопленные отложенные награды, агент помещает наблюдение и награды в ER буфер, выбирает следующее действие согласно политике модели и сообщает его окружению.

симальным размером, равным размеру ER буфера, а также хеш-таблица, адресуемая по ключу, в которой хранятся записи — наблюдение, награда (если уже получена), выбранное действие. В качестве ключа используется кортеж (номер симуляции, уникальный идентификатор пакета). При переполнении буфера старейший ключ вынимается из очереди и удаляется из хеш-таблицы. При получении отложенной награды происходит быстрый поиск по ключу в хеш-таблице и вставка награды в соответствующую запись.

3.2 Обучение RL-агента

Всего для обучения были использованы 35 симуляций с различными значениями битрейтов интерференции на линиях, а так же значениями битрейта MLD (см. таблицу 4 в приложении В).

Для отсека краевых эффектов (таких как начало и конец вещания интерферирующих станций, разгон параметров EDCA и т. д.) обучение проводилось только на данных из середины симуляции: каждая симуляция длилась 18.5 секунд, из которых первые 15.5 секунд приходились на "разогрев" сети, а последняя секунда отводилась на "отключение" сети. Таким образом, эффективное время обучения алгоритма на каждой из симуляций составило 2 секунды. При этом из-за различных значений битрейта MLD число записей, получаемых из разных симуляций, было различно — см. также таблицу 4.

При обучении использовался ER buffer с максимальной вместимостью 500000 записей, что превышает общее количество записей, полученных при обучении агента. Таким образом, все записи, полу-

ченные при обучении, использовались при тренировке модели.

Обучение проводилось в течение двух эпох: в течение первой эпохи агент обучался в режиме онлайн, в течение второй эпохи — в режиме оффлайн (на содержащихся в ER буфере значениях).

Размер батча был выбран равным 64, записи семплировались из ER буфера равномерно. Итерация тренировки проводилась каждые 64 шага.

Кривая зависимости лосса от числа шагов обучения приведена на рис. 5. Из графика видно, что наблюдается рост значений лосса от времени. Это связано со структурой и порядком симуляций, выполняющихся в процессе обучения: при больших значениях битрейта MLD и битрейтов интерференции MLD испытывает большие задержки и амплитуда значения наград растет, а в тренировочном наборе сценариев самые большие значения битрейта MLD появляются в ER буфере как раз в конце первой эпохи обучения. Причем, согласно таблице 4, записей с такими высокими значениями задержек преобладающее количество. Все эти факторы обуславливают рост значений лосса от времени.

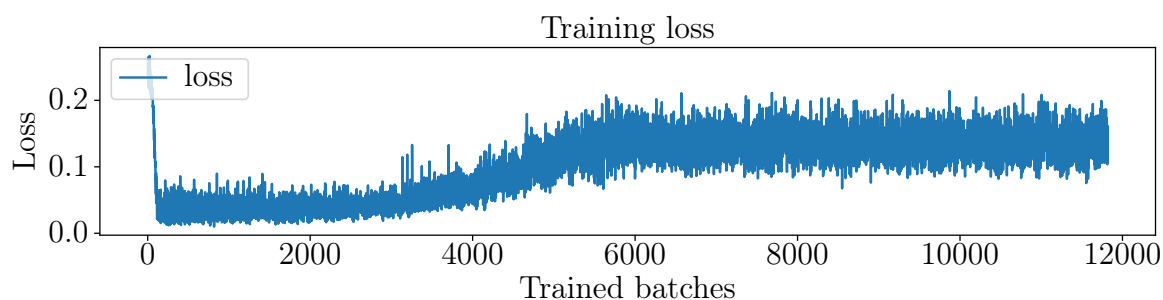


Рис. 5: Зависимость батч-лосса от числа шагов обучения.

4 Тестирование и обсуждение результатов

4.1 Бейзлайн

В качестве бейзлайна предлагается рассматривать решение, отправляющее пакеты на доступные для передачи линки по принципу Round-robin [24]: каждый следующий пакет передается следующим линком, при достижении конца списка линков отсчет начинается заново. Такой бейзлайн обеспечивает равномерную загрузку всех доступных линков, разделяя приходящий трафик на равные доли. Минусом такого подхода является то, что такое решение не учитывает текущее состояние линка: заполненность очередей, уровень интерференции, среднюю задержку и т. п., и, следовательно, в случаях кардинально различного состояния линков является не оптимальным.

4.2 Метрики

В рассматриваемых сценариях с IPTV-трафиком первоочередную роль играет Quality of Experience (QoE) конечного пользователя: передача должна не допускать потерь пакетов, задержки сети худшего случая должны быть минимизированы. Следующий набор метрик отражает описанные требования:

- 99% квантиль задержки L_{ij}
- 99% квантиль задержки показа IPTV кадра
- Процент потерянных кадров в течение симуляции.

Выбор 99% квантиля позволяет увидеть минимальную задержку худших 1% пакетов, что отражает сетевые задержки худшего случая. Фильтруя симуляции, в которых было потеряно больше 1% пакетов, получим максимальный битрейт MLD, который может быть обслужен в рамках данного сценария не выходя за пределы хорошего QoE.

4.3 Тестовые сценарии

В качестве тестовых сценариев были выбраны сценарии со значениями битрейтов интерференции, представленные в таблице 3. Для каждого из сценариев были запущены симуляции длиной 50 секунд со значениями битрейта MLD из интервала $[10, 300]$ с шагом 10 МБпс. Каждая симуляция была запущена 10 раз с различными значениями random seed, значения метрик усреднены по random seed.

Такой выбор тестовых сценариев призван покрыть характерные случаи сетевой интерференции: низкая интерференция (например, "10,10,10"), средняя ("30,30,30") и высокая ("60,60,60"), а также различные случаи с асимметричной интерференцией на трех линках.

Таблица 3: Параметры тестовых симуляций

№	Битрейты интерференции, МБпс	Представлены в обучающих сценариях
1	10,60,60	\pm только для $[10, 200]$
2	10,30,60	\pm только для $[10, 200]$
3	60,10,10	\pm только для $[10, 200]$

Таблица 3: Параметры тестовых симуляций (продолжение)

№	Битрейты интерферен- ции, МБпс	Представлены в обучающих сценариях
4	60,60,60	\pm только для {10, 30, 60}
5	30,30,30	\pm только для {10, 30, 60}
6	10,10,10	\pm только для {10, 30, 60}
7	60,30,30	-
8	30,60,60	-
9	30,10,10	-
10	10,30,30	-

Из таблицы 3 видно, что часть комбинаций параметров битрейтов интерференции для тестовых сценариев не представлена в обучающей выборке целиком (7-10), часть представлена только для выбранных значений битрейтов MLD (4-6), а часть представлена почти целиком, за исключением интервала битрейтов MLD (200, 300] с шагом 10 МБпс (1-3).

Еще одним отличием тестовых сценариев от сценариев, использованных в процессе обучения, является динамичность интерференции: с периодом 3 секунды для случайной интерферирующей сети происходит изменение битрейта интерференции на битрейт из интервала

$$\text{ibitrate}_i \pm 0.3 * \text{ibitrate}_i$$

где ibitrate_i — битрейт AP_i , указанный в таблице. Выбор нового

значения битрейта происходит равномерно из указанного интервала. Введение такого изменения позволяет проверить устойчивость модели к динамическим изменениям параметров симуляции, а также позволяет протестировать устойчивость RL-агента на сценариях со значениями битрейтов интерференции, не входивших в тестовую выборку.

4.4 Результаты

Сравнение метрик, описанных в параграфе 4.2, для бейзлайна и обученного RL-агента на тестовых сценариях представлено на рис. 6. Результаты позволяют сделать следующие выводы:

- RL-агент обыгрывает бейзлайн во всех сценариях со средней и высокой интерференцией с точки зрения максимального значения поддерживаемого битрейта MLD по QoE и не проигрывает с этой точки зрения в сценариях с низкой и средней интерференцией
- для всех сценариев наблюдается существенный (до одного порядка) выигрыш в метриках 99% квантилей задержек на высоких значениях битрейта MLD, граничащих с максимально поддерживаемым по QoE
- наблюдается аномальная область проигрыша по значениям 99% квантилей обеих метрик задержек для сценариев с низким и средним значением битрейтов интерференции (битрейт MLD 130-190 МБпс)

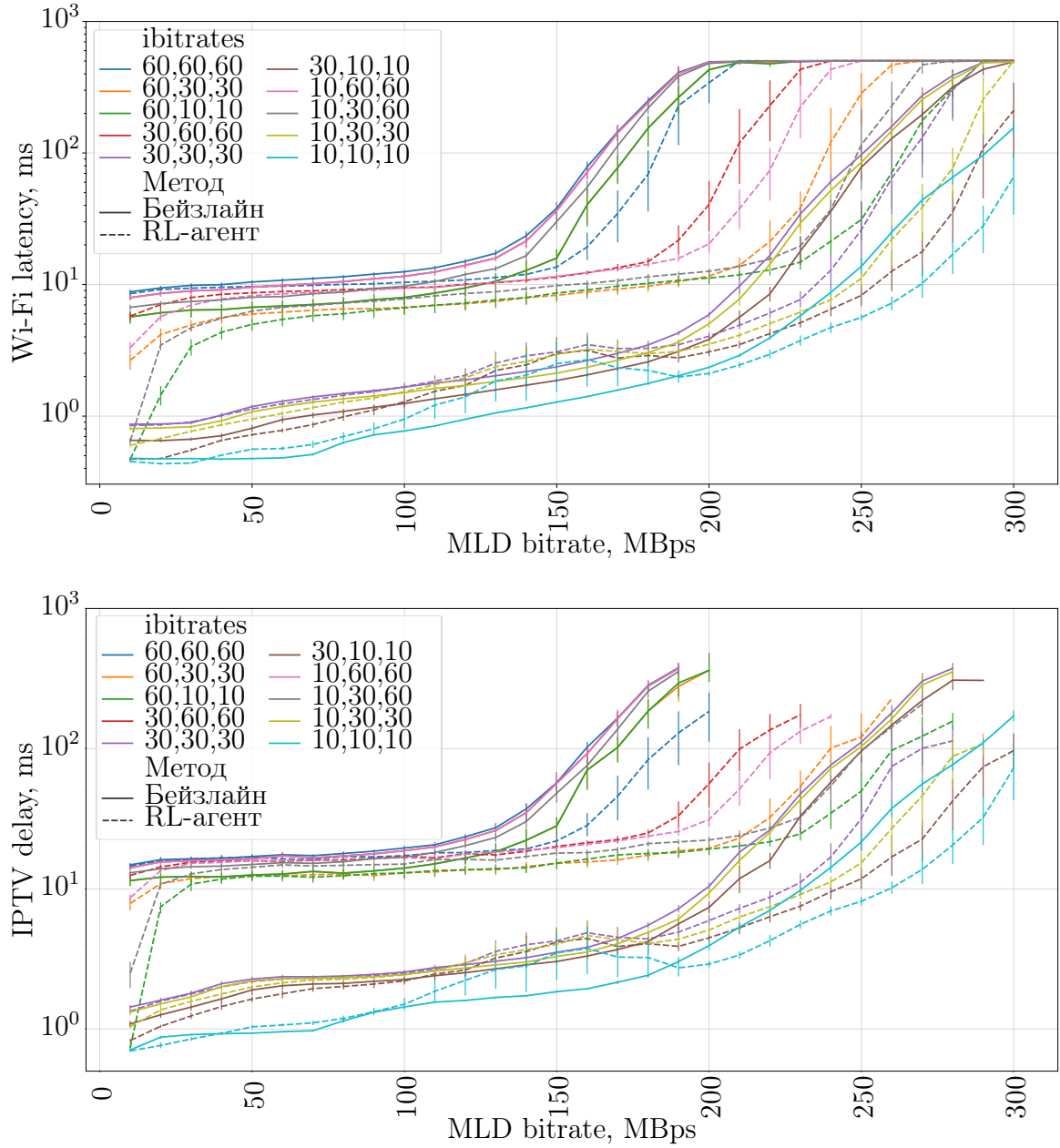


Рис. 6: Значения 99% квантилей задержек Wi-Fi (сверху) и задержек показа IPTV кадра (снизу) в логарифмическом масштабе для бейзлайна и RL-агента на тестовых сценариях. Каждая точка — среднее по 10 запускам с различными random seed, вертикальными линиями отмечены значения максимума и минимума по random seed.

- для низких значений битрейта MLD на сценариях с низкой и средней интерференцией не наблюдается существенного различия между бейзлайном и RL-агентом
- не наблюдается существенных отличий между результатами на сценариях, частично использованных в процессе обучения и на сценариях, не входивших в обучающую выборку

Также можно заметить, что кривая бейзлайна для сценария "60,10,10" (выделена зеленым сплошным цветом) и сценария "60,30,30" (выделена оранжевым сплошным цветом) совпадают почти во всех точках. Это можно объяснить несовершенством работы бейзлайна в случае, когда один из трех линков подвержен высокой интерференции по сравнению с другими. Действительно: бейзлайн определит ровно треть всех пакетов в симуляции на худший линк, занимая этими пакетами всю верхнюю треть распределения задержек, и, таким образом, определяя 99% квантиль независимо от того, какие значения будут на двух оставшихся линках. Похожую картину можно наблюдать и в поведении кривой для сценария "10,30,60".

5 Заключение

В данной работе исследовались возможности балансировки нагрузки Wi-Fi устройства с несколькими линками при помощи RL-агента в сценариях на большом диапазоне интерферирующих нагрузок на линках. На основе спецификации находящегося в разработке стандарта Wi-Fi 7 была разработана архитектура Multi-link operation

в сетевом симуляторе ns-3, обучена простая DQN RL-модель и измерены QoE характеристики полученного решения.

Обученная модель уверенно обходит предложенный бейзлайн с точки зрения максимального передаваемого битрейта без потери пакетов на всех рассмотренных тестовых сценариях. Кроме того, для больших значений транслируемых битрейтов (≈ 200 МБпс) для всех рассматриваемых сценариев было получено уменьшение 99% квантилей задержки сети Wi-Fi и задержки показа кадра IPTV. В отдельных сценариях достигается уменьшение 99% квантилей задержек показа кадра IPTV до 10 раз, увеличение максимального передаваемого битрейта без потери пакетов до 80 МБпс.

Также тестирование модели показало, что для низких значений передаваемого битрейта (до 100 МБпс) решение выдает значения метрик не хуже предложенного бейзлайна для всех рассмотренных сценариев. Для сценариев с низкой и средней интерференцией решение показывает ухудшение метрик задержек для диапазона средних передаваемых битрейтов, тем не менее демонстрируя улучшение с точки зрения максимального передаваемого битрейта.

Среди планируемых направлений дальнейших исследований и планов работы над платформой:

- исследование причин аномального ухудшения метрик в диапазоне средних битрейтов MLD в сценариях с низкой и средней интерференцией и улучшение работы алгоритма в таких случаях
- измерение и оптимизация сложности вычислений для внедрения в реальные Wi-Fi устройства с поддержкой грядущего стан-

дарт Wi-Fi 7

- расширение тестовой выборки симуляциями с более высокими значениями битрейта MLD для более точной оценки максимальной пропускной способности, достигаемой решением

Приложение

A Exponentially weighted moving average

В качестве усреднения для некоторых признаков используется exponentially-weighted moving average (EWMA) или exponentially-weighted moving counter (EWMC) с характеристикой затухания $T = 10$ ms, задающееся следующими формулами:

$$V_{\text{EWMA}}^k = \frac{S_{\text{EWMA}}^k}{W_{\text{EWMA}}^k}$$

S_{EWMA}^k и W_{EWMA}^k инициализируются 0 и 1 соответственно и обновляются при каждом новом наблюдении нового признака V^k следующим образом:

$$\begin{aligned} S_{\text{EWMA}}^k &= V^k + \exp\left(\frac{-(t-t_{\text{last}})}{T}\right) * S_{\text{EWMA}}^i \\ W_{\text{EWMA}}^k &= 1 + \exp\left(\frac{-(t-t_{\text{last}})}{T}\right) \end{aligned}$$

где t — время наблюдения, t_{last} — время предыдущего наблюдения.

EWMC рассматривается в качестве усредненного числа событий (например, приход пакета с IP-уровня), для вычисления EWMC берется значение S_{EWMA}^k для признака, V^k которого всегда равен 1 (событие случилось).

В Параметры симуляций из тренировочной выборки

Таблица 4: Параметры тренировочных симуляций

№ симуляции	Битрейт MLD, МБпс	Битрейты интерференции, МБпс	Количество записей в ER буфер
1	10,10,10	10	2014
2	30,30,30	10	2014
3	60,60,60	10	2014
4	10,10,10	30	5485
5	30,30,30	30	5485
6	60,60,60	30	5479
7	10,10,10	60	10856
8	30,30,30	60	10855
9	60,60,60	60	10828
10	10,60,60	10	2010
11	10,60,60	30	5476
12	10,60,60	60	10852
13	60,10,10	10	2011
14	60,10,10	30	5481
15	60,10,10	60	10849
16	10,30,60	10	2014
17	10,30,60	20	3798
18	10,30,60	30	5482

Таблица 4: Параметры тренировочных симуляций (продолжение)

№ симуляции	Битрейт MLD, МБпс	Битрейты интерферен- ции, МБпс	Количество записей в ER буфер
19	10,30,60	40	7332
20	10,30,60	50	9047
21	10,30,60	60	10854
22	10,30,60	70	12762
23	10,30,60	80	14732
24	10,30,60	90	16632
25	10,30,60	100	17743
26	10,30,60	110	19574
27	10,30,60	120	21602
28	10,30,60	130	23205
29	10,30,60	140	24855
30	10,30,60	150	27325
31	10,30,60	160	28761
32	10,30,60	170	30828
33	10,30,60	180	31786
34	10,30,60	190	34024
35	10,30,60	200	29979

Из соображений экономии времени обучения рассматривались только дискретные значения битрейтов интерференции: 10 (низкая нагрузка), 30 (средняя нагрузка) и 60 (высокая нагрузка). Также таб-

лица поделена на следующие группы симуляций: 1-3, 4-6, 7-9 — одинаковая интерференция различного уровня, разные битрейты MLD, 10-12 — один линк с низкой интерференцией и два с высокой интерференцией, 13-15 — один линк с высокой интерференцией и два с низкой интерференцией, 16-35 — фиксированное "асимметричное" значение битрейтов интерференции "10,30,60" и битрейты MLD в интервале $[10, 200]$ с шагом 10 МБпс.

Список литературы

- [1] *802.11*. URL: https://en.wikipedia.org/wiki/IEEE_802.11 (дата обр. 01.06.2021).
- [2] Dmitry Bankov и др. “IEEE 802.11ax uplink scheduler to minimize, delay: A classic problem with new constraints”. в: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 2017, с. 1—5. DOI: 10.1109/PIMRC.2017.8292382.
- [3] Alessandro Checco и Douglas Leith. “Proportional Fairness in 802.11 Wireless LANs”. в: *Communications Letters, IEEE* 15 (сент. 2011), с. 807—809. DOI: 10.1109/LCOMM.2011.060811.110502.
- [4] Alexander Krotov, Anton Kiryanov и Evgeny Khorov. “Rate Control With Spatial Reuse for Wi-Fi 6 Dense Deployments”. в: *IEEE Access* 8 (2020), с. 168898—168909. DOI: 10.1109/ACCESS.2020.3023552.
- [5] *802.11ax Standard*. 2021. URL: https://standards.ieee.org/standard/802_11ax-2021.html (дата обр. 01.06.2021).
- [6] *802.11be Project Authorization Request*. 2019. URL: https://www.ieee802.org/11/PARs/P802_11be_PAR_Detail.pdf (дата обр. 01.06.2021).
- [7] B Ravi Kiran и др. *Deep Reinforcement Learning for Autonomous Driving: A Survey*. 2021. arXiv: 2002.00444 [cs.LG].
- [8] David Silver и др. “Mastering the game of Go with deep neural networks and tree search”. в: *Nature* 529 (2016), с. 484—503. URL:

<http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html>.

- [9] OpenAI и др. *Dota 2 with Large Scale Deep Reinforcement Learning*. 2019. arXiv: 1912.06680 [cs.LG].
- [10] *Deepmind AI reduces Google data centre cooling bill by 40%*. 2016. URL: <https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40> (дата обр. 01.06.2021).
- [11] Ursula Challita, Li Dong и Walid Saad. “Proactive resource management for LTE in unlicensed spectrum: A deep learning perspective”. в: *IEEE Transactions on Wireless Communications* 17.7 (2018), с. 4674—4689. ISSN: 15361276. DOI: 10.1109/TWC.2018.2829773. arXiv: 1702.07031.
- [12] Yanhua Huang. “Deep Q-networks”. в: *Deep Reinforcement Learning: Fundamentals, Research and Applications* (2020), с. 135—160. DOI: 10.1007/978-981-15-4095-0_4.
- [13] Hao Ye и Geoffrey Ye Li. “Deep Reinforcement Learning for Resource Allocation in V2V Communications”. в: *IEEE International Conference on Communications* 2018-May (2018). ISSN: 15503607. DOI: 10.1109/ICC.2018.8422586. arXiv: 1711.00968.
- [14] Giovanni Peserico и др. “Rate Adaptation by Reinforcement Learning for Wi-Fi Industrial Networks”. в: *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2020-September* (2020), с. 1139—1142. ISSN: 19460759. DOI: 10.1109/ETFA46521.2020.9212060.

- [15] *ns-3 network simulator*. URL: <https://www.nsnam.org/about/> (дата обр. 01.06.2021).
- [16] *802.11ax Project Authorization Request*. 2014. URL: <https://mentor.ieee.org/802.11/dcn/14/11-14-0165-01-0hew-802-11-hew-sg-proposed-par.docx> (дата обр. 01.06.2021).
- [17] *ns-3 GitLab*. URL: <https://gitlab.com/nsnam/ns-3-dev> (дата обр. 01.06.2021).
- [18] *802.11be*. URL: https://en.wikipedia.org/wiki/IEEE_802.11be (дата обр. 01.06.2021).
- [19] IEEE Task Group BE. *802.11be Specification Framework Document*. 2021. URL: <https://mentor.ieee.org/802.11/dcn/19/11-19-1262-23-00be-specification-framework-for-tgbe.docx> (дата обр. 01.06.2021).
- [20] *MPEG-4*. URL: https://en.wikipedia.org/wiki/Advanced_Video_Coding (дата обр. 01.06.2021).
- [21] Nathan Jay и др. *Internet Congestion Control via Deep Reinforcement Learning*. 2019. arXiv: 1810.03259 [cs.NI].
- [22] Adam Paszke и др. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. в: *Advances in Neural Information Processing Systems 32*. под ред. H. Wallach и др. Curran Associates, Inc., 2019, с. 8024—8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.

- [23] Piotr Gawłowicz и Anatolij Zubow. “ns3-gym: Extending OpenAI gym for networking research”. В: *arXiv* (2018). ISSN: 23318422. arXiv: 1810.03943.
- [24] *Round-robin scheduling*. URL: https://en.wikipedia.org/wiki/Round-robin_scheduling (дата обр. 01.06.2021).