

Эффективная виртуализация сетевых ресурсов в системах MU-MIMO¹

Д.Е. Зудин^{*,**}, М.В. Любогощев^{*,**}, Е.М. Хоров^{*}

^{*}Институт проблем передачи информации им. А.А. Харкевича РАН, Москва

^{**}Московский физико-технический институт (национальный исследовательский университет), Москва

Поступила в редколлегию 03.03.2023 г., Принята 12.05.2023 г.

Аннотация—В работе рассматривается задача виртуализации ресурсов в беспроводных сетях 5G, использующих технологию множественного доступа с использованием MU-MIMO (англ.: Multiple User-Multiple Input Multiple Output). Виртуализация сетевых ресурсов является ключевым подходом к распределению сетевых ресурсов в сетях пятого поколения, однако MU-MIMO усложняет задачу распределения ресурсов беспроводного канала между виртуальными подсетями, а также между отдельными клиентами. Для решения данной задачи в работе используется архитектура виртуализации DeSlice, и разрабатывается метод планирования радиоресурсов, учитывающий особенности как задачи виртуализации ресурсов, так и MU-MIMO и различных типов трафика, включая трафик приложений виртуальной реальности и веб-трафик. С помощью имитационного моделирования демонстрируется, что разработанный метод планирования радиоресурсов значительно повышает качество обслуживания для обоих типов трафика по сравнению со стандартной системой. Результаты оценки производительности демонстрируют эффективность разработанного метода для распределения ресурсов в MU-MIMO системе с поддержкой технологии виртуализации сетевых ресурсов.

КЛЮЧЕВЫЕ СЛОВА: 5G, виртуализация сетевых радиоресурсов, VR, MU-MIMO, приложения виртуальной реальности, веб-трафик.

DOI: 10.53921/18195822_2023_23_2_241

1. ВВЕДЕНИЕ

Развертывание сетей пятого поколения (5G) требует значительных капиталовложений от сотовых операторов. Поэтому для ускорения перехода на системы 5G важно разработать эффективные решения для совместного использования сетевой инфраструктуры. Одним из таких решений является виртуализация сетевых ресурсов, которое позволяет создавать несколько гибко настраиваемых и изолированных виртуальных сетей в рамках общей физической сети. Данная технология имеет потенциал для эффективного предоставления дифференцированных услуг абонентам и позволяет реализовывать все основные сценарии использования 5G в рамках единой инфраструктуры [1].

Однако виртуализация ресурсов в беспроводных сетях является сложной проблемой по нескольким причинам. Во-первых, мгновенное состояние беспроводного канала меняется со временем, в т.ч. из-за мобильности клиентов. Во-вторых, трафик разных виртуальных подсетей имеет различные и зачастую несовместимые требования к качеству обслуживания (англ.: Quality of Service, QoS). В-третьих, хотя использование технологии MU-MIMO и повышает

¹ Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда № 21-79-10431, <https://rscf.ru/project/21-79-10431/>

емкость и пропускную способность системы, но существенно усложняет задачу распределения радио ресурсов, а также оценку объема радиоресурсов, необходимого для удовлетворения требований к качеству обслуживания.

Проблема виртуализации сетевых ресурсов является многокритериальной стохастической задачей оптимизации, которую можно решать с помощью различных методов. Существующую литературу по данной теме можно разделить на три основные группы. Первая группа фокусируется на планировании MU-MIMO трафика с различными требованиями к качеству обслуживания, оптимизируя различные критерии [2–5]. Однако предлагаемые решения не нацелены на использование в системах 5G с виртуализацией сетевых ресурсов и поэтому не рассматривают проблему изоляции виртуальных сетей, которая является критически важной для технологии виртуализации [8].

Во второй группе работ исследуется виртуализация радиоресурсов в системах без MU-MIMO и используются различные подходы, такие как стохастическая оптимизация [8], теория игр [9, 10], машинное обучение [11, 12] или эвристические методы [13]. Однако в системе MU-MIMO проблема виртуализации и планирования радиоресурсов значительно отличается из-за дополнительной возможности обслуживать несколько клиентов в одном ресурсном блоке, и поэтому приведенные решения неприменимы.

Третья группа исследований рассматривает виртуализацию сетевых ресурсов с MU-MIMO, но сводит эту проблему к распределению подмножеств антенн между виртуальными сетями для обеспечения требований к качеству обслуживания [14, 15]. Однако этот подход может быть неэффективным, поскольку каждой виртуальной подсети выделяется только небольшое количество антенн, что ограничивает преимущества систем MIMO. Поэтому необходимы более эффективные решения, которые могут полностью использовать преимущества технологии MU-MIMO.

В данной статье метод решения описанной проблемы основывается на архитектуре DeSlice [16] с целью упростить проблему виртуализации сетевых ресурсов и превратить ее в несколько более простых однокритериальных и взаимосвязанных задач оптимизации. Формулируется оптимизационная задача для сценария совместного обслуживания веб- и VR-трафика в системе MU-MIMO с поддержкой виртуализации сетевых ресурсов. Для указанной задачи предлагается решение на основе взвешенного пропорционально-честного (англ.: Weighted Proportional Fair, WPF) алгоритма планирования [17].

Основным вкладом данной статьи является набор новых алгоритмов планирования ресурсов, которые позволяют эффективно использовать архитектуру DeSlice в системе MU-MIMO, чтобы улучшить качество обслуживания для веб- и VR-трафика. Полученные результаты показывают значительный прирост производительности по сравнению с системой без поддержки технологии виртуализации сетевых ресурсов, демонстрируя эффективность предложенного метода распределения ресурсов в MU-MIMO системе с поддержкой технологии виртуализации сетевых ресурсов.

Дальнейшее изложение построено следующим образом. В разделе 2 представлены обзор архитектуры DeSlice и постановка задачи. Затем в разделе 3 описывается предлагаемый метод и набор разработанных алгоритмов планирования радиоресурсов в системах MU-MIMO. В последующем разделе 4 с помощью имитационного моделирования оценивается производительность предлагаемого решения и проводится его сравнение с известными методами. Наконец, в разделе 5 подводятся основные результаты работы и обсуждаются потенциальные направления для будущих исследований.

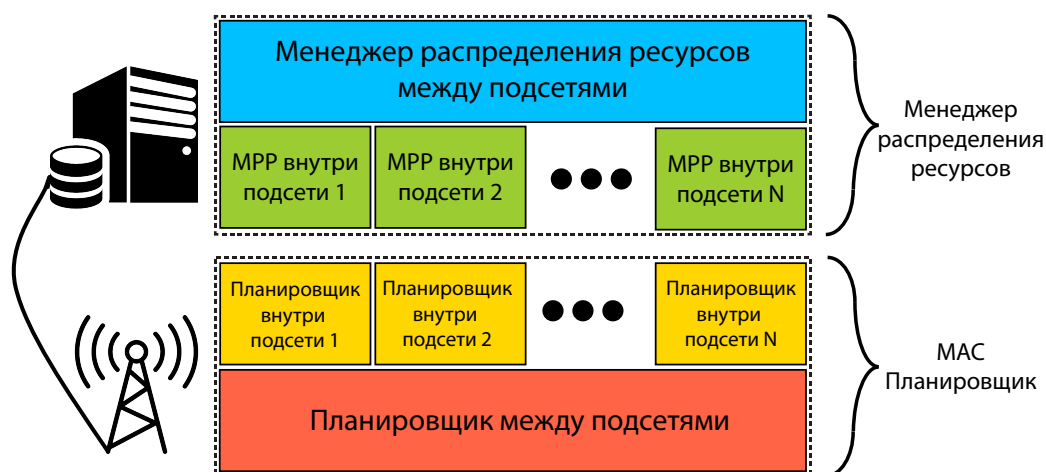


Рис. 1. Архитектура DeSlice [16].

2. АРХИТЕКТУРА DESLICE И ПОСТАНОВКА ЗАДАЧИ

Проблема виртуализации сетевых ресурсов заключается в том, что различные виртуальные сети и типы трафика имеют различные требования к качеству обслуживания. Качество обслуживания оценивается с помощью различных метрик, которые с точки зрения оптимизации могут быть не выпуклыми и вычисляться на различных временных интервалах: от миллисекунд до часов. Кроме того, базовая станция 5G (англ.: 5G base station, gNB) должна решать проблему распределения ресурсов в режиме реального времени. Поэтому идея архитектуры DeSlice, рис. 1, заключается в упрощении комплексной задачи путем разбиения ее на более простые взаимозависимые однокритериальные задачи оптимизации долгосрочного и краткосрочного распределения ресурсов внутри и между подсетями.

Основной идеей предложенного в работе [16] решения является формулировка требований к качеству обслуживания различных подсетей и их клиентов в виде доли сетевых ресурсов и временного окна, в котором должна гарантироваться необходимая доля ресурсов. Однако предлагаемое авторами решение не учитывает особенности использования технологии MU-MIMO в современных сетях 5G. Основное отличие проблемы планирования радиоресурсов в системах MU-MIMO заключается в том, что несколько клиентов могут обслуживаться одновременно в одном ресурсном блоке, и пропускная способность беспроводного канала связи с клиентом зависит также и от того, с какими клиентами он разделяет данный ресурсный блок. В результате чего взаимосвязь доли ресурсов и требований к качеству обслуживания становится более сложной.

В данной работе исследуется система, состоящая из одной многоантенной базовой станции 5G (gNB) и нескольких многоантенных пользовательских устройств (англ.: User Equipment, UE), подключенных к сети Интернет через gNB. Пользовательские устройства являются либо веб-клиентами, скачивающими веб-страницы, либо VR-клиентами, загружающими интерактивные видеопотоки VR с облачных серверов. Все веб-клиенты принадлежат одной виртуальной подсети, а все VR-клиенты — другой, и виртуализация ресурсов беспроводного канала осуществляется в рамках архитектуры DeSlice.

MU-MIMO планирование осуществляется по следующему алгоритму. В каждом временном слоте (англ.: Time Transmission Interval, TTI) gNB объединяет пользовательские устройства на основе корреляций их беспроводных каналов в низкокоррелированные между собой группы

(далее, MU-группы). Затем мощность равномерно распределяется между MU-группами, и вся доступная gNB полоса пропускания распределяется между клиентами внутри каждой MU-группы.

Целью данной работы является повышение качества совместного обслуживания веб- и VR-трафика в рамках описанной системы по сравнению с существующими решениями. Для этого ставится задача разработать метод планирования радиоресурсов и набор алгоритмов краткосрочного и долгосрочного планирования радиоресурсов, учитывающие особенности как задачи виртуализации ресурсов, так и MU-MIMO и трафика приложений виртуальной реальности и веб-трафика.

3. РАЗРАБОТАННЫЕ АЛГОРИТМЫ

Для адаптации архитектуры DeSlice к работе в системе 5G с поддержкой технологии MU-MIMO необходимо оценить общее количество доступных ресурсов беспроводного канала и долю ресурсов достаточных для удовлетворения требований к качеству обслуживания различных потоков. Для этого разрабатываются следующие алгоритмы, работающие на верхних уровнях архитектуры DeSlice, рис. 1. Во-первых, подсчитывается среднее число MU-групп K . Во-вторых, оценивается идеальная пропускная способность для каждого клиента, предполагая, что число MU-групп соответствует среднему, а все доступные ресурсы в MU-группе клиента i выделяются только ему:

$$C_i^{ideal} = \mathcal{F} \left(\text{SINR}^{avg} \cdot \frac{\rho_i}{K} \right), \quad (1)$$

где \mathcal{F} — отображение SINR в пропускную способность, K — среднее число MU-групп, ρ_i — средняя фактическая корреляция между клиентом i и его соседями по MU-группе во время передачи, и SINR^{avg} — средний SINR клиента.

Основываясь на идеальной пропускной способности, оценивается доля ресурсов, выделяемая клиенту i для удовлетворения его требований к качеству обслуживания, соответствует следующему выражению:

$$\xi_i = \frac{R_i^{tgt} \cdot S_{MU_j}}{C_i^{ideal}}, \quad (2)$$

где R_i^{tgt} — целевая скорость передачи данных для UE i , S_{MU_j} — сумма весов клиентов в группе MU j .

При выделении долгосрочных долей ресурсов виртуальным подсетям учитываются требования к качеству обслуживания различных типов трафика. Например, VR-сессии требуют стабильной пропускной способности, обеспечивающей необходимую пропускную способность для VR потока. Поэтому долгосрочная доля ресурсов выделяется VR-клиенту в соответствии с следующим выражением:

$$w_i^{VR} = \frac{b_i}{C_i(t)}, \quad (3)$$

где $C_i(t)$ — идеальная пропускная способность беспроводного канала, определенная в выражении 1, а b_i — целевой битрейт VR-потока клиента i . Оставшаяся часть ресурсов выделяется виртуальной подсети, которая обслуживает веб-трафик. Таким образом,

$$S_{web} = K - \sum_{i \in U_{VR}} w_i^{VR}.$$

Все веб-потоки имеют одинаковый приоритет и должны получать одинаковые долгосрочные доли ресурсов. Для этого выделенная веб-подсети доля ресурсов равномерно распределяется между всеми активными веб-потоками:

$$w_i^{web} = \frac{S_{web}}{N_{web}(t)}, \quad (4)$$

где $N_{web}(t)$ — количество активных веб-клиентов.

На нижних уровнях архитектуры DeSlice MAC-планировщик решает задачу максимизации взвешенной суммы средних пропускных способностей, где веса клиентов w_i соответствуют их целевым долгосрочным долям ресурсов, а τ_i соответствуют целевым временным интервалам, на которых необходимо выделять целевую долю ресурсов клиентам. Пусть $s_i(t)$ — объем данных, переданных клиенту i в ТТИ t . Тогда $\overline{r_i(t)} = (1 - \frac{1}{\tau_i})\overline{r_i(t-1)} + (\frac{1}{\tau_i})\frac{s_i(t)}{TTI}$ является экспоненциально взвешенным скользящим средним пропускной способности в окне τ_i . Требуется максимизировать следующую функцию полезности:

$$U = \sum_i w_i \cdot \log(\overline{r_i(t)}). \quad (5)$$

К данному выражению применяется алгоритм градиентного спуска и вычисляется метрика клиента. Таким образом, метрика клиента i для обслуживания в RB k в ТТИ t задается выражением:

$$\mu_{ik}^{WPF}(t) = \frac{w_i}{\tau_i} \cdot \frac{r_{ik}(t)}{r_i(t)}. \quad (6)$$

Представленное уравнение используется самым нижним блоком архитектуры DeSlice. На завершающем шаге планирования данный блок выбирает из лидеров каждой виртуальной подсети клиента с наивысшей метрикой.

4. ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ

Для оценки производительности системы с различными алгоритмами планирования используется платформа имитационного моделирования NS-3 [18].

Рассматривается система с переменным суммарным числом веб- и VR-пользователей. Веб-клиенты загружают страницы с размерами, описываемыми случайной величиной, имеющей логарифмическое распределение со средним значением 25 МБ, минимальным 1 МБ и максимальным 100 МБ. Время между запросами моделируется случайной величиной с усеченным экспоненциальным распределением со средним значением 5 секунд, минимумом 1 секундой и максимумом 10 секундами. VR-потоки моделируются с помощью записей трафика, генерируемого приложением облачной виртуальной реальности Pico Neo Streaming Assistant [19]. Их средний битрейт равен 15 Мбит/с, а частота кадров — 60 кадров в секунду. Каждая VR-сессия начинается в начале эксперимента и продолжается до конца. Моменты подключения различных VR-клиентов к серверу облачной виртуальной реальности смещены друг относительно друга в пределах нескольких ТТИ, поэтому VR-сессии не начинаются одновременно. В таблице 1 приведены другие параметры эксперимента.

Рассматриваются три варианта настройки 5G MU-MIMO системы:

- *Legacy No Slicing*: стандартная система с алгоритмом планирования PF [20] без виртуализации.
- *Weighted Delay*: система с планировщиком maximum-largest weighted delay first (MLWDF) [21].

Таблица 1. Параметры эксперимента.

Несущая частота	3.6 GHz
Ширина канала	100 MHz
Модель канала	3GPP TR 38.901 EPA
Число антенн gNB/UE	64/4
Высота gNB/UE	30/1.5 м
Длительность TTI	0,5 мс
Модельное время	40 с
Число экспериментов на одну точку	20

– *DeSlice MIMO*: архитектура DeSlice с алгоритмами, описанными в разделе 3.

VR-клиенты используют облачные интерактивные приложения, такие как игры, образовательные или инженерные платформы. В приведенных сценариях требуется плавное воспроизведение потока VR, поэтому для моделирования VR качества восприятия (англ.: Quality of Experience, QoE) используется следующий критерий удовлетворенности пользователя VR-сессией. VR-кадр считается доставленным успешно, если он доставлен целиком и с задержкой менее 50мс. Тогда VR-пользователь считается удовлетворенным, если в течение VR-сессии было доставлено успешно более 95% кадров. Для веб-трафика важно загружать страницы как можно скорее, поэтому в качестве метрики QoE для веб-трафика используется среднее геометрическое пропускной способности уровня приложений (англ.: Application Layer Perceived Throughput, ALPT). Выбор метрики обусловлен тем, что она отражает как скорость загрузки веб-страниц, так и справедливость распределения ресурсов между клиентами.

Для оценки эффективности разработанного метода обратимся к рис. 2 и 3, на которых приведены результаты экспериментов в рассматриваемом сценарии. Во-первых, отметим, что решение *Weighted Delay* не справляется с совместным обслуживанием двух типов трафика даже в низкой нагрузке. Излишне приоритизируя VR-клиентов, оно выделяет недостаточно ресурсов веб-клиентам, и тем самым перегружает сеть и удовлетворяет менее 80% VR-сессий даже при малом числе VR-клиентов в сети.

По сравнению с *Legacy* решением, разработанный в данной статье метод распределения ресурсов удовлетворяет более 90% VR-сессий вплоть до 36 VR-клиентов в системе, т.е. увеличивает эффективную емкость сети на 50%. Одновременно с этим, разработанный метод также позволяет повысить качество обслуживания веб-клиентов. Прирост среднего геометрического пропускной способности веб-клиентов составляет минимум 25% на всем рассмотренном диапазоне нагрузок на сеть.

5. ЗАКЛЮЧЕНИЕ

В работе архитектура виртуализации ресурсов беспроводного канала DeSlice адаптирована для работы в системах 5G MU-MIMO, обслуживающих веб- и VR-трафик. Разработан метод планирования радиоресурсов и набор алгоритмов краткосрочного и долгосрочного планирования радиоресурсов, совместимые с архитектурой DeSlice и учитывающие особенности как задачи виртуализации ресурсов, так и MU-MIMO и трафика приложений виртуальной реальности и веб-трафика. С помощью имитационного моделирования продемонстрировано, что разработанные алгоритмы значительно повышают качество обслуживания обоих типов трафика по сравнению с существующими решениями. В частности, прирост средней геометрической скорости загрузки веб-страниц превышает 25%, а прирост емкости сети для VR-трафика превышает 50%.

В дальнейшей работе планируется усовершенствовать разработанные алгоритмы для поддержки URLLC и видеотрафика в системах MU-MIMO, поскольку данные типы трафика име-

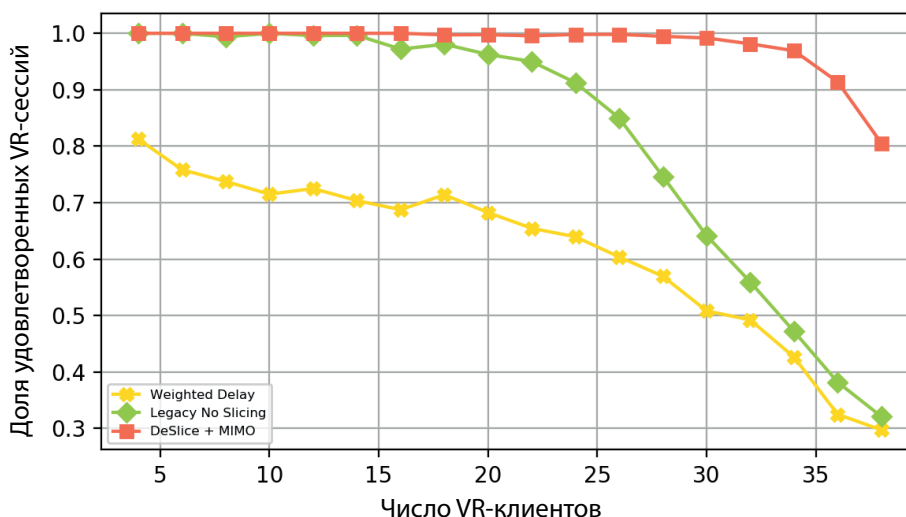


Рис. 2. Зависимость доли удовлетворенных VR-сессий от числа VR-клиентов при 48 веб-клиентах в системе.

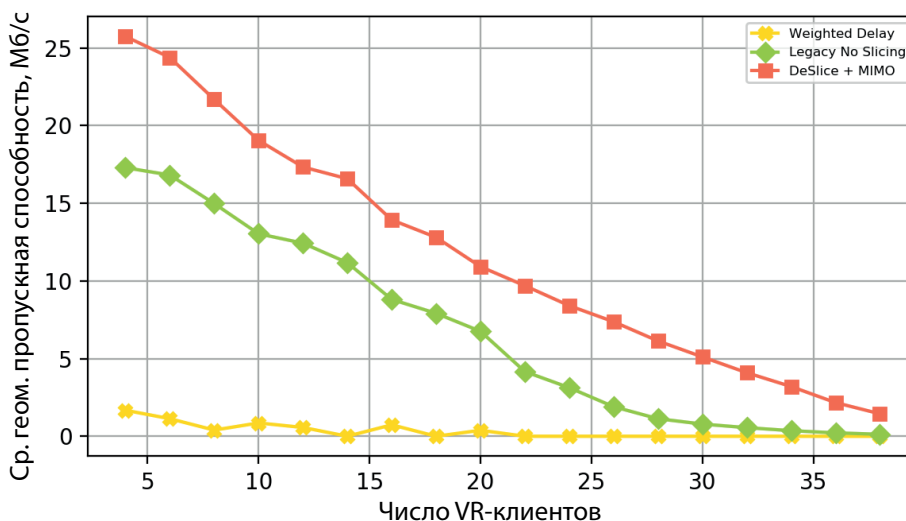


Рис. 3. Зависимость средней геометрической пропускной способности веб-клиентов от числа VR-клиентов при 48 веб-клиентах в системе.

ют отличные от рассмотренных в данной работе требования к качеству обслуживания и составляют важную часть трафика в современной сети Интернет.

СПИСОК ЛИТЕРАТУРЫ

1. I.F. Akyildiz, A. Kak, E. Khorov, A. Krasilov, A. Kureev, ARBAT: A flexible network architecture for QoE-aware communications in 5G systems, *Computer Networks*, 2018, vol. 147, pp. 262–279, ISSN 1389-1286, doi: 10.1016/j.comnet.2018.10.016.
2. Huang C. -W, Althamary I., Chou Y. -C., Chen H. -Y., Chou C. -F. A DRL-Based Automated Algorithm Selection Framework for Cross-Layer QoS-Aware Scheduling and Antenna Allocation in Massive MIMO Systems. *IEEE Access*, 2023, vol. 11, pp. 13243–13256.
3. Akhtar J., Rajawat K., Gupta V. and Chaturvedi A. K. Joint User and Antenna Selection in Massive-MIMO Systems With QoS-Constraints. *IEEE Systems Journal*, 2021, vol. 15, no. 1, pp. 497–508.

4. Ajib W., Haccoun D., Frigon J. -F. An efficient QoS-based scheduling algorithm for MIMO wireless systems. *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference*, 2005, pp. 1579-1583.
5. Lebedeva I., Yusupov R., Krasilov A. Multiplexing of URLLC and eMBB Traffic in a Downlink Channel with MU-MIMO. *J. Commun. Technol. Electron*, 2022, vol. 67, pp. 1506-1512.
6. Dangi R., Jadhav A., Choudhary G., Dragoni N., Mishra M.K., Lalwani P. ML-Based 5G Network Slicing Security: A Comprehensive Survey. *Future Internet*, 2022, vol. 14, no. 14, p. 116.
7. Hurtado Sanchez J.A., Casilimas K, Caicedo Rendon O.M. Deep Reinforcement Learning for Resource Management on Network Slicing: A Survey. *Sensors*, 2022, vol. 22, no. 8, p. 3031.
8. Papa A., Jano A., Ayvasik S., Ayan O., Gursu H.M., Kellerer W. User-Based Quality of Service Aware Multi-Cell Radio Access Network Slicing. *IEEE Transactions on Network and Service Management*, 2022, vol. 19, no. 1, pp. 756-768.
9. Li Y., Wang Y., Jin Y., Cheng X., Xu L., Liu G. Research on Wireless Resource Management and Scheduling for 5G Network Slice. *International Wireless Communications and Mobile Computing (IWCMC)*, 2021, pp. 508-513.
10. Adamuz-Hinojosa O., Munoz P., Ameigeiras P., Lopez-Soler J. M. Potential-Game-Based 5G RAN Slice Planning for GBR Services. *IEEE Access*, 2023, vol. 11, pp. 4763-4780.
11. Dangi, R., Lalwani, P. Harris Hawks optimization based hybrid deep learning model for efficient network slicing in 5G network. *Cluster Comput*, 2023.
12. Yan M., Feng G., Zhou J., Sun Y., Liang Y.-C. Intelligent Resource Scheduling for 5G Radio Access Network Slicing. *IEEE Transactions on Vehicular Technology*, 2019 vol. 68, no. 8, pp. 7691-7703.
13. I. F. Akyildiz, E. Khorov, A. Kiryanov, D. Kovkov, A. Krasilov, M. Liubogoshchev, D. Shmelkin, and S. Tang, XStream: A new platform enabling communication between applications and the 5G network, in *2018 IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, IEEE, 2018, pp. 1-6, doi: 10.1109/GLOCOMW.2018.8644183.
14. Marquez C., Gramaglia M., Fiore M., Banchs A., Costa-Perez X. Resource Sharing Efficiency in Network Slicing. *IEEE Transactions on Network and Service Management*, 2019, vol. 16, no. 3, pp. 909-923.
15. Sun Y., Peng M., Mao S., Yan S. Hierarchical Radio Resource Allocation for Network Slicing in Fog Radio Access Networks. *IEEE Transactions on Vehicular Technology*, 2019, vol. 68, no. 4, pp. 3866-3881.
16. Liubogoshchev M., Zudin D., Krasilov A., Krotov A., Khorov E. DeSlice: An Architecture for QoE-Aware and Isolated RAN Slicing. *Sensors*, 2023, vol. 23, n. 9, p. 4351.
17. Khawam K., Kofman D., Altman E. The weighted proportional fair scheduler. *Proceedings of the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2006.
18. Network simulator 3 (NS-3), <https://www.nsnam.org/>. Accessed on 14/04/2023.
19. Pico Neo 2, <https://www.picoxr.com/uk/products/g2-4k>. Accessed on 14/04/2023.
20. Stolyar A. L. On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation. *Operations research*, 2005, 53(1), 12-25.
21. Basukala R., Ramli H. A. M., Sandrasegaran K. Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system. *2009 First Asian Himalayas International Conference on Internet*, 2009, pp. 1-5.

Towards Efficient Network Slicing in MU-MIMO Systems

Zudin D.E., Liubogoshchev M.V., Khorov E.M.

This paper addresses the problem of network slicing in 5G networks that use Multiple User-Multiple Input Multiple Output (MU-MIMO). Network slicing is a key approach to resource allocation in 5G systems, but MU-MIMO complicates the resource allocation problem for various slices and users. To address this problem, the paper uses the DeSlice slicing architecture and develops a resource allocation method that considers the peculiarities of the network slicing problem, MU-MIMO, and different traffic types, including Web and VR traffic. With extensive simulations, the developed method is shown to significantly improve the quality of service for both traffic types compared to the legacy system. The results demonstrate the effectiveness of the proposed method for 5G network slicing with MU-MIMO.

KEYWORDS: Cellular systems, 5G, slicing, MU-MIMO, VR.