Max-min Fair Resource Allocation in SD-RAN

Fidan Mehmeti
Technische Universität München
Munich, Germany
fidan.mehmeti@tum.de

Wolfgang Kellerer
Technische Universität München
Munich, Germany
wolfgang.kellerer@tum.de

ABSTRACT

Software-Defined Radio Access Networks (SD-RANs), introduced for the first time in 5G networks, symbolize a paradigm shift in the way the allocation of cellular network resources is performed. The main feature of SD-RAN is the possibility to decouple the control plane from the data plane, and associating the former with a controller that is located away from the Base Stations (BSs). This enables an increased flexibility in the allocation of network resources, resulting in considerable performance improvements compared to the classical pre-5G resource allocation in cellular networks. However, it is not yet clear to what extent this enhancement ranges in terms of different metrics of interest and optimization objectives. One such objective is to allocate resources so that the minimum value of the data rate in the network is maximized, i.e., to provide max-min fairness. Therefore, in this paper, we consider analytically the problem of max-min fairness in an SD-RAN environment by deriving the policy which accomplishes that along with the achievable performance, of interest to users and cellular operators. We do this for two scenarios; one in which we provide max-min fairness across all users in the network, and the other in which the goal is to provide max-min fairness in throughput across BSs, and within users of each BS. We evaluate the performance with input parameters from real data sets. Results show that the introduction of SD-RAN improves the minimum rate by up to 4× compared to the case with no SD-RAN controller.

CCS CONCEPTS

• **Networks** → **Network protocol design**; **Network performance modeling**.

KEYWORDS

SD-RAN; 5G; Performance optimization; Max-min fairness

ACM Reference Format:
Fidan Mehmeti and Wolfgang Kellerer. 2022. Max-min Fair Resource Allocation in SD-RAN. In *Proceedings of the 18th ACM International Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet '22), Oct. 24–28, 2022, Montreal, QC, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3551661.3561359

1 INTRODUCTION

In the previous generations of cellular networks, in Radio Access Networks (RANs) both the data plane and control plane operations were performed jointly. With the emergence of Software Defined Networks (SDNs) [1], and its adaptation in RANs, known as SD-RAN [2], the separation of control from data plane became possible for the first time in RANs of 5G networks, as a paradigm shift on how the assignment of network resources is handled in particular, and how cellular networks operate in general. The control is transferred to *centralized* units known as SD-RAN controllers. This brings a lot of benefits into the mobile network [2–4], since it detaches the monolithic RAN control and enables co-operation among RAN components, i.e., Base Stations (BSs), improving the network performance along different dimensions.

This increased level of flexibility arises from having a broader view of the network, which is provided by the centralized SD-RAN approach. In that way, depending on the current spread of the users (UEs) across BSs, and their channel conditions for which the UEs periodically update their serving BSs [5], and BSs send those information to the SD-RAN controller, the latter can reallocate resources to BSs accordingly. BSs then perform the resource allocation across their corresponding UEs. As a consequence, exploiting the wide network knowledge leads to an overall improved performance as it allows for optimal allocation decisions. As opposed to SD-RAN, in a classical RAN approach, each BS has its own fixed set of resources, and allocates them to the UEs within its operational area.

One of the improvements SD-RAN brings to both the cellular operator and UEs is in terms of the increased throughput. However, while there exist some open-source SD-RAN prototypes, like FlexRAN [2] and 5G-EmPOWER [3], it is not yet clear to what extent the delegation of traditional RAN functions to centralized controllers increases the throughput. Furthermore, not only the overall throughput is of interest to the different entities in a network, but there are also other figures of merit with corresponding objectives. One such is to improve the performance of the worst-performing user in the network, i.e., to guarantee *max-min fairness* [6]. This optimization approach is particularly important since it can provide insights on the best performance a user with the worst channel conditions can expect to experience, which could serve as an indicator for the type of applications/services she can run on her smartphone. To the best of our knowledge, this problem has not been addressed before. Therefore, we consider it in this paper.

Predicting the throughput in general, and the worst-case achievable data rate in particular, in a cellular network accurately is quite challenging mostly because of the dynamic nature of wireless channels, stemming from UEs' mobility and effects like shadowing [7]. This channel variability drives forward the need to possess accurate statistics of channel conditions and, in terms of worst-case

---

Q2SWinet '22, October 24–28, 2022, Montreal, QC, Canada

ACM ISBN 978-1-4503-9481-9/22/10…$15.00
https://doi.org/10.1145/3551661.3561359

performance optimization, to assign different amount of resources to each UE over time.

Several very important research questions arise related to providing max-min fairness in an SD-RAN environment:

- What is the allocation policy that provides max-min fairness in an SD-RAN-enabled cellular network with a given number of BSs, where the number of UEs per BS is known *a priori* together with the channel conditions at a given time? What is the highest achievable throughput of the worst performing user in such a network?
- If the goal is to provide max-min fairness among BSs, what is the allocation policy that leads to that objective?
- How does an SD-RAN-led network perform against a system where resource allocation is done only at BSs?

To answer the aforementioned questions, in this paper, we formulate two optimization problems. In the first, the objective is to find the maximum achievable data rate of the user with the lowest value (and hence with the worst channel conditions), given the limited number of resources, but having the flexibility of allocating them adaptively to the BSs (in line with the SD-RAN main feature), depending on the number of UEs they serve and the channel conditions of the latter. We show that if max-min fairness is to be guaranteed across all users, the policy that achieves this objective is the one which provides the same data rate to all the users, i.e., UEs with higher Channel Quality Indicator (CQI) receive fewer resources than those with worse channel conditions (lower CQIs). [1] In the second optimization problem, we aim to provide max-min fairness among BSs, for which another allocation policy should be used. With the latter policy as well, an inversely proportional amount of resources to the channel conditions is allocated to the users. Further, we derive the maximum achievable values. The results we provide in this work are particularly helpful for the cellular operator as they can provide an exact prediction of the expected maximized worst-case performance and can also help in network resource planning. The main message of this paper is that the use of SD-RAN can improve the throughput of the worst performing UE. Specifically, our main contributions are:

- We derive the maximum achievable throughput for the worst-performing entity, i.e., following a max-min fair allocation policy, and we derive that policy for two cases of interest.
- We validate our results with realistic input data gathered from a measurement campaign [9].
- We show concrete performance improvements when using SD-RAN compared to the traditional RAN approach in terms of the lowest data rate.

The remainder of this paper is organized as follows. In Section 2, we discuss some related work. The system model and the problem formulation are presented in Section 3. The analysis for the max-min fairness among all UEs and BSs is provided in Section 4. Further, Section 5 introduces a benchmark model with no SD-RAN. Section 6 describes the simulator used in this work, and provides evaluation results, including validations of the theoretical models. Finally, Section 7 concludes the paper.

2 RELATED WORK

Since its introduction, the concept of SD-RAN has attracted a considerable attention in the past few years [10], [11]. Among the first works that propose transferring the control decisions from the BS to a centralized controller are [12] and [13], which also discuss the advantages in increased flexibility when using SD-RAN. However, the gains in terms of the increased throughput or the improvements in terms of the lowest rate in the network are not discussed neither in [12] nor in [13].

To the best of our knowledge, the data rate (from different aspects) maximization problem in SD-RAN environments has not been considered so far, including the max-min fairness problem. The first prototype implementations of SD-RAN are FlexRAN [2] and 5G-EmPOWER [3], both suffering from the limited number of UEs that can be served by a single controller. On the other hand, as opposed to [2] and [3], with our analysis we can predict the performance for any number of BSs, number of UEs, channel conditions, and any amount of available resources.

In [14], the problem of minimizing the number of assigned resources has been considered in an SD-RAN environment, by taking into account two types of network slices; those that pertain to delay-sensitive traffic, and the second type with throughput-critical traffic. The other contribution in [14] is that it is shown that slice isolation is maintained. However, there is no discussion on the resource allocation policy that would provide max-min fairness.

The general problem of providing max-min fairness in a wireless network has been considered in [15]. However, the setup in [15] is different from ours as the authors focus there on the rate control. Also, the problem setup in [15] is not compatible with the SD-RAN environment, losing this way the additional flexibility in resource allocation.

On a related note, the authors in [16] consider the problem of resource allocation where network slices can be spread across multiple BSs. The objective is to allocate resources to maximize the overall throughput across all UEs, by guaranteeing a minimum data rate to everyone first. However, the solution in [16] is based on a non-closed form approximation approach, which does not allow to see the explicit dependency of throughput on different input parameters. Also, the max-min fair resource allocation is not considered in [16]. Instead, in our work, we solve the problem over the entire network in its most general form for any number of UEs, BSs, and heterogeneous channel statistics while providing a closed-form expression for the max-min throughput.

The work most related in spirit to ours is [8], in which the authors consider the problem of of max-min fairness after providing the same minimum data rate to everyone in a single BS, and reallocating afterwards the unused resources to the same UEs. While the data rates vary from one frame to another, all the users receive the same data rate in a given frame. However, the SD-RAN controller is not used there, meaning that the amount of resources belonging to each BS is fixed, loosing the additional flexibility of adaptive resource allocation to the BSs, which as will be shown in Section 6 of our work increases the data rate considerably compared to the no-SD-RAN setup.

---

[1] CQI is an information each UE sends to its BS to describe the channel conditions. The value ranges from 1 (very poor channel conditions) to 15 (excellent channel conditions) [8].

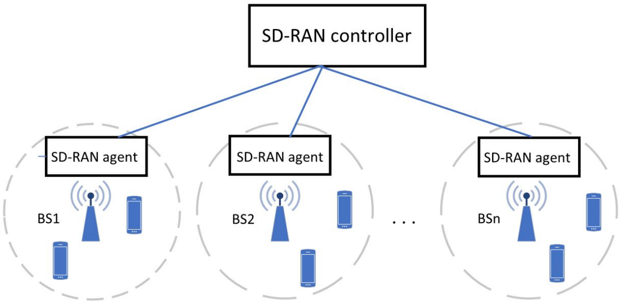

Figure 1: Illustration of the SD-RAN environment.

3 PERFORMANCE MODELING

In this section, we first introduce the system model, and then define the two optimization problems that we are solving in this paper.

3.1 System model

We consider an SD-RAN-enabled setup (Fig. 1) with a single controller responsible for allocating resources to BSs. For each BS there is an SD-RAN agent that communicates with the controller [17]. This communication is performed using the Transport Control Protocol (TCP). We denote the BSs by $i \in \mathcal{N}$. There are in total $n = |\mathcal{N}|$ BSs in the operational area of the controller. We denote by $j \in \mathcal{M}_i$ the UEs within the coverage area of BS i, where $m_i = |\mathcal{M}_i|$ is the number of UEs in BS i. So, the total number of UEs in the system of interest is $\sum_{i=1}^{n} m_i$.

5G uses *Physical Resource Blocks (PRBs)* as the unit of allocation per slot [18]. Each PRB consists of 12 subcarriers. The slot duration is a function of the subcarrier spacing. Specifically, if the subcarrier spacing is 15 KHz (implying PRB width of 180 KHz), the slot duration is 1 ms. If the subcarrier spacing is 30 KHz (PRB width of 360 KHz), the corresponding slot duration is 0.5 ms. A further decrease in the slot duration (2×) is realized when switching to subcarrier spacings of 60 KHz, and another 2× when switching to 120 KHz [5]. Within a slot, different PRBs are assigned to different UEs. In general, the assignment varies across slots. Consequently, scheduling is to be performed along two dimensions, *time* and *frequency*. In our setup, there are K available PRBs for n BSs.

UEs experience different channel conditions (CQIs) in different PRBs even within the same slot. Due to the UE mobility and time-varying channel characteristics, per-PRB CQI (which is a function of Signal-to-Interference-Plus-Noise-Ratio (SINR)) changes from one slot to another, whose value depending on the Modulation and Coding Scheme (MCS) used sets the per-PRB rate. To maintain the problem analytically tractable, we make a simplifying assumption in this paper. Namely, we assume that the BS splits the transmission power equally among all the PRBs it transmits on, and that the channel characteristics for a UE remain unchanged across all PRBs (identical CQI over all PRBs for a given UE), but change randomly (according to some distribution) from one slot to another, and are mutually independent among UEs. These assumptions reduce the resource allocation problem to the number of allocated PRBs and not to which PRBs are assigned to a UE.

Table 1: Notation

$\mathcal{N}$	The set of all BSs
$n = \lvert\mathcal{N}\rvert$	Number of BSs
$\mathcal{M}_i$	The set of all UEs in BS i
$m_i = \lvert\mathcal{M}_i\rvert$	Number of UEs in BS i
K	Total number of PRBs
K_i	Number of PRBs allocated to BS i
$K_{i,j}$	Number of PRBs allocated to UE j in BS i
$R_{i,j}$	Per-block rate of UE j in BS i in a slot
$C_{i,j}$	Data rate of UE (i, j) in a slot
$p_{R_{i,j}}(x)$	PMF of per-PRB rate of UE (i, j)
UE	User Equipment, i.e., user

The previous assumptions imply that in every slot, UE (i, j)[2], where $i \in \mathcal{N}$ and $j \in \mathcal{M}_i$, will have a per-PRB rate (also known as per-block rate) that can be modeled with a discrete random variable, $R_{i,j}$, with values in $\{r_1, r_2, \dots, r_{15}\}$, such that $r_1 < r_2 < \dots < r_{15}$, with a Probability Mass Function (PMF) $p_{R_{i,j}}(x)$, which is a function of UE (i, j)'s CQI over time.[3] Different UEs have different per-block rate distributions.

Before proceeding any further, Table 1 summarizes the notation used throughout this paper.

3.2 Problem formulation

Every UE sends periodically the information about its CQI to the corresponding serving BS. Each BS then collects all the CQI information from the UEs in its area and forwards them to the SD-RAN controller (see Fig. 1). Based on the CQI information from all BSs (and hence from all UEs), the controller then, depending on the allocation policy used, decides on the number of PRBs to allocate to each BS in each slot. Further, from the PRBs it receives, each BS "forwards" those resources to the UEs in its coverage area. Hence, using SD-RAN, *the resource allocation is performed in two levels.* First, among BSs, and then each BS allocates the PRBs it received from the controller to its respective UEs.

Let $K_{i,j}, \forall j \in \mathcal{M}_i$, denote the number of PRBs UE j gets from BS i.[4] If $K_i, \forall i \in \mathcal{N}$, denotes the number of PRBs that BS i receives from the controller in a slot, then it holds $K_i = \sum_{j=1}^{m_i} K_{i,j}$. Let $C_{i,j}$ express the data rate of UE (i, j), for which it holds $C_{i,j} = K_{i,j} R_{i,j}$.

In this paper, the goal is to provide max-min fair resource allocation along two dimensions: (i) across UEs, and (ii) across BSs (and across users within the same BS).

3.2.1 Max-min fairness across UEs. First, our aim is to provide max-min fair resource allocation across all users in the entire SD-RAN-led network, which is equivalent to maximizing the lowest data rate in the network. We have the following optimization problem in that case:

$$\mathcal{P}_1: \quad \max_{K_{i,j}} \min \quad K_{i,j} R_{i,j} \tag{1}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{i,j} \le K, \tag{2}$$

---

[2] We denote every UE with the ordered pair (i, j), where i stands for the BS, while j denotes the UE receiving service by that BS.

[3] To simplify the notation, we omit the reference to time throughout this paper.

[4] Each UE can receive resources from a single BS only.

$$K_{i,j} \geq 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{3}$$

Constraint (2) expresses the maximum number of PRBs that can be allocated across all BSs, whereas constraint (3) captures the fact that the number of allocated PRBs to UEs cannot be negative. The decision variables are $K_{i,j}$.

3.2.2 Max-min fairness across BSs. In the second scenario, the goal is to provide max-min fair allocation across BSs, and for the assigned resources to BSs, max-min fair allocation among UEs of a given BS. Essentially, the idea is to maximize the lowest BS throughput. The following optimization problem describes this scenario:

$$\mathcal{P}_2: \quad \max_{K_{i,j}} \min_{i} \quad \sum_{j=1}^{m_i} K_{i,j} R_{i,j} \tag{4}$$

$$K_{i,j} R_{i,j}, \tag{5}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{i,j} \leq K, \tag{6}$$

$$K_{i,j} \geq 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{7}$$

Note that $\mathcal{P}_2$ is a multi-objective optimization problem [19] because besides maximizing the minimum throughput across all BSs (4), within UEs of each BS max-min fair resource allocation should be maintained as well (5). Hence, the presence of subscript i (denoting BSs) under the max-min operator. The constraints (6) and (7) are identical to those of $\mathcal{P}_1$. The decision variables are again $K_{i,j}$.

In the next section, we solve optimization problems $\mathcal{P}_1$ and $\mathcal{P}_2$ by obtaining the corresponding optimal policies and providing the objective values.

4 PERFORMANCE OPTIMIZATION

In this section, first we determine the optimal policy and derive the corresponding data rate by solving $\mathcal{P}_1$. Then, we solve $\mathcal{P}_2$.

4.1 Max-min fairness across all UEs

Before solving $\mathcal{P}_1$, we need to introduce what max-min fair resource allocation policy is. The *max-min fairness* [6] is defined as:

Definition 1. *A resource allocation policy is max-min fair for a vector of rates $(C_1^*, \ldots, C_n^*)$ if for any set of other feasible rates $(C_1, \ldots, C_n)$ it holds that if $C_i > C_i^*$, for some $i = 1, \ldots, n$, then there exists a $j \in \{1, \ldots, n\}$, such that $C_j^* \leq C_i^*$ and $C_j < C_j^*$.*

We proceed with solving $\mathcal{P}_1$. Essentially, Definition 1 states that under max-min fair allocation it is not possible to increase the data rate of a user (say, user i) without decreasing the rate of some other user, who already has a lower (or equal) data rate than user i. Therefore, in our setup, a max-min resource allocation *is the one in which all the users have the same rate.* Consequently, we have the following:

Result 1. *A max-min fair policy across all UEs for allocating the resources is the one for which it holds*

$$K_{i,j} R_{i,j} = const, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{8}$$

where $\sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{i,j} = K$.

Proof. If the data rate is the same for all users, then increasing any of the $K_{i,j}$ would result in decreasing the $K_{t,s}$ of some other user for which it already holds $K_{i,j} \geq K_{t,s}$. So, when all the data rates are identical for every user, we can conclude that the allocation is max-min fair. Therefore, we have

$$K_{1,1} R_{1,1} = K_{1,2} R_{1,2} = \ldots = K_{n,m_n} R_{n,m_n}, \tag{9}$$

resulting in Eq.(8). □

Essentially, with max-min fairness, UEs with bad channel conditions (low CQI) in a slot are "rewarded" by getting more resources from the SD-RAN controller, as opposed to the UEs experiencing good channel conditions (high CQI), which receive a lower number of PRBs. Note that constraint (2) holds with equality. This is obvious as the principle goal is to maximize something (hence all the resources should be used).

From Eq.(8), we have $K_{1,2} = K_{1,1} \frac{R_{1,1}}{R_{1,2}}, \ldots, K_{i,j} = K_{1,1} \frac{R_{1,1}}{R_{i,j}}, \ldots, K_{n,m_n} = K_{1,1} \frac{R_{1,1}}{R_{n,m_n}}$. Combining these expressions with $\sum_{i=1}^{n} \sum_{j=1}^{m_i} K_{i,j} = K$, we obtain $K_{1,1} R_{1,1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{R_{i,j}} = K$, resulting in

$$K_{1,1} = \frac{\frac{K}{R_{1,1}}}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{R_{i,j}}}, \tag{10}$$

or written in the general form we have:

Result 2. *A max-min fair policy across all UEs in the network with SD-RAN is achieved if the number of assigned PRBs to UE (i, j) follows the policy*[5]

$$K_{i,j} = \frac{\frac{K}{R_{i,j}}}{\sum_{k=1}^{n} \sum_{l=1}^{m_k} \frac{1}{R_{k,l}}}, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{11}$$

Substituting Eq.(11) into Eq.(8), we obtain:

Result 3. *The maximum lowest data rate in a slot in a network with SD-RAN, and hence the solution to $\mathcal{P}_1$, is the one all UEs experience and is given by*

$$C = \frac{K}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{R_{i,j}}}. \tag{12}$$

The data rate from Eq.(12) is for a given slot and depends on the configuration of per-block rates of all UEs in that slot. Therefore, the computation of the number of PRBs to be assigned to every user has to be done on per-slot basis. The average data rate over time for this scenario is

$$\mathbb{E}[C] = K \mathbb{E}\left[\frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{R_{i,j}}}\right], \tag{13}$$

which yields

$$\mathbb{E}[C] = K \sum_{x_{1,1}=r_1}^{r_{15}} \cdots \sum_{x_{n,m_n}=r_1}^{r_{15}} \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{x_{i,j}}} \prod_{i=1}^{n} \prod_{j=1}^{m_i} p_{R_{i,j}}(x_{i,j}). \tag{14}$$

In Eq.(14), there are $\sum_{i=1}^{n} m_i$ sums, one for each UE. Needless to say, as the minimum data rate is maximized in every slot with our policy, $\mathbb{E}[C]$ provided in Eq.(14) is the maximum expected value

---

[5] We have changed i to k and j to l in the denominator of Eq.(11) for practical purposes because we express K for user (i, j).

of the lowest data rate that can be obtained with any policy for a given set of users with corresponding per-block rate distributions.

4.2 Max-min fairness across BSs

When it comes to solving $\mathcal{P}_2$, a slightly different approach needs to be taken because this is a multi-objective optimization problem. Nevertheless, for the same reasons as those stated when solving $\mathcal{P}_1$, all the resources must be allocated, i.e., constraint (6) must hold with strict equality.

First, let us focus on the second objective, i.e., (5). As this objective states that max-min allocation has to be valid within UEs of the same BS, then in line with the discussion from Section 4.1, we can deduce that the data rates of the users receiving service from the same BS have to be identical.

The aforementioned discussion implies that for BS i, it holds

$$K_{i,1}R_{i,1} = K_{i,2}R_{i,2} = \ldots = K_{i,m_i}. \tag{15}$$

Without loss of generality, for BS i we express the data rate of the general term $K_{i,j}R_{i,j}$ through that of the first UE, i.e., $j = 1$, leading to $K_{i,j}R_{i,j} = K_{i,1}R_{i,1}$. The total throughput in BS i is

$$\sum_{j=1}^{m_i} K_{i,j}R_{i,j} = \sum_{j=1}^{m_i} K_{i,1}R_{i,1} = K_{i,1}R_{i,1}m_i. \tag{16}$$

So, for the max-min fairness in this scenario, the number of users in a BS comes explicitly into play.

We proceed with the first objective, i.e., (4). Similarly, in line with Definition 1, and the discussion preceding Result 1, we can conclude:

Result 4. *A max-min fair policy for allocating the resources across all BSs in an SD-RAN environment is the one for which it holds*

$$\sum_{j=1}^{m_i} K_{i,j}R_{i,j} = const, \quad \forall i \in \mathcal{N}, \tag{17}$$

where $\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_{i,j} = K$.

Similarly, w.l.o.g., we express the general term of Eq.(17) in terms of UE $(1,1)$, through UE $(i,1)$. The procedure is as follows. First, combining Eq.(17) with Eq.(16), and expressing the latter through BS 1, we get

$$K_{i,1}R_{i,1}m_i = K_{1,1}R_{1,1}m_1, \tag{18}$$

resulting in

$$K_{i,1} = K_{1,1} \cdot \frac{R_{1,1}}{R_{i,1}} \cdot \frac{m_1}{m_i}. \tag{19}$$

From Eq.(15) we have

$$K_{i,j}R_{i,j} = K_{i,1}R_{i,1}, \tag{20}$$

and after substituting Eq.(19) into Eq.(20), we obtain

$$K_{i,j} = K_{1,1} \cdot \frac{R_{1,1}}{R_{i,j}} \cdot \frac{m_1}{m_i}. \tag{21}$$

The next step is to determine $K_{1,1}$. Replacing Eq.(21) into $\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_{i,j} = K$, and rearranging, we get

$$K_{1,1} = \frac{\frac{K}{m_1 R_{1,1}}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m_i R_{i,j}}}. \tag{22}$$

Substituting Eq.(22) into Eq.(21), we have:

Result 5. *A max-min fair reaource allocation across all BSs, and within the UEs of each BS, in the network with SD-RAN is achieved if the number of assigned PRBs to UE* (i, j) *follows the policy*

$$K_{i,j} = \frac{\frac{K}{m_i R_{i,j}}}{\sum_{k=1}^{n}\sum_{l=1}^{m_k}\frac{1}{m_k R_{k,l}}}, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{23}$$

There are interesting observations that can be made from Result 5. The first, as expected, is that the number of allocated PRBs is inversely proportional to the channel conditions of the UE in that slot. The second outcome is that the number of allocated PRBs for a UE should also be inversely proportional with the number of UEs in the slot. So, UEs with good channel conditions receive fewer resources, and UEs which are receiving service from BSs with fewer users receive more PRBs.

Substituting Eq.(23) into $C_{i,j} = K_{i,j}R_{i,j}$, we get:

Result 6. *The maximum lowest data rate in a slot in BS i in a network with SD-RAN, and hence the solution to* $\mathcal{P}_2$*, is the one all UEs in BS i experience and is given by*

$$C_i = \frac{\frac{K}{m_i}}{\sum_{k=1}^{n}\sum_{l=1}^{m_k}\frac{1}{m_k R_{k,l}}}, \quad \forall i \in \mathcal{N}. \tag{24}$$

Result 6 indicates that the data rate of UEs in a BS, following the max-min fair policy across BSs, is inversely proportional to the number of UEs in that BS, which is to be expected.[6]

Further, we have:

Result 7. *The maximum lowest BS throughput in the network with SD-RAN (and hence the objective of* $\mathcal{P}_2$*) is*

$$m_i C_i = \frac{K}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m_i R_{i,j}}}, \tag{25}$$

which is the same for every BS.

In this case as well, the computation of the number of PRBs to be assigned to every user has to be done on per-slot basis. The average BS throughput over time with this policy is

$$\mathbb{E}[m_i C_i] = K\mathbb{E}\left[\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m_i R_{i,j}}}\right], \tag{26}$$

which yields

$$\mathbb{E}[m_i C_i] = K \sum_{x_{1,1}=r_1}^{r_{15}} \cdots \sum_{x_{n,m_n}=r_1}^{r_{15}} \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m_i x_{i,j}}} \prod_{i=1}^{n}\prod_{j=1}^{m_i} p_{R_{i,j}}(x_{i,j}). \tag{27}$$

Obviously, this is the maximum expected lowest BS throughput for any allocation policy. This will be shown in Section 6.

5 BENCHMARK MODEL

We need a benchmark model in order to be able to assess the performance achieved by the SD-RAN-enabled network in terms of max-min fairness when it comes to resource allocation. To that end, the most suitable model is the one in which there is no SD-RAN, i.e., the network operates in a "classical" way, where every BS is

---

[6] Note that here we have made a slight abuse in the notation, denoting by C_i the data rate for all the users in BS i, instead of using $C_{i,j}$.

pre-assigned its set of PRBs. If K is the total number of PRBs in the system, then we assume that each BS operates on $\frac{K}{n}$ PRBs, where n, as already defined, is the number of BSs.

In no-SD-RAN setup, the optimization formulation for BS i, whose solution guarantees max-min fair resource allocation to its UEs is

$$\mathcal{P}_0(i): \quad \max_{K_{i,j}} \min \quad K_{i,j}R_{i,j} \tag{28}$$

$$\text{s.t.} \quad \sum_{j=1}^{m_i} K_{i,j} \leq \frac{K}{n}, \tag{29}$$

$$K_{i,j} \geq 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}_i. \tag{30}$$

Essentially, for each BS we would need to solve $\mathcal{P}_0(i)$ separately. From Section 4.1, we know that the amount of resources a UE gets (in this case from its BS) is inversely proportional to its CQI. Therefore, similar to Result 2, in order max-min fairness to be established in BS i, it is required that the resource allocation underlies the following policy:

$$K_{i,j} = \frac{\frac{K}{nR_{i,j}}}{\sum_{l=1}^{m_i} \frac{1}{R_{i,l}}}, \quad \forall j \in \mathcal{M}_i. \tag{31}$$

Apparently, identical to the solution of $\mathcal{P}_1$, the UEs within a BS which suffer from bad channel conditions will receive more PRBs than those experiencing high CQI.

Substituting Eq.(31) into $C_{i,j} = K_{i,j}R_{i,j}$, for the maximized minimum UE rate in BS i, we obtain

$$C_i = \frac{\frac{K}{n}}{\sum_{j=1}^{m_i} \frac{1}{R_{i,j}}}, \quad \forall i \in \mathcal{N}, \tag{32}$$

i.e., it is the same for all UEs in BS i, but it is different at different BSs (see the index i in the denominator of Eq.(32)). This implies that across the entire network the UE with the lowest rate in a slot can have the maximum value of $\min(C_i)$, which as will be shown in Section 6, is always lower than the value obtained from $\mathcal{P}_1$.

For apparent reasons (re-scheduling of PRBs within BSs is not enabled), it is not possible to establish max-min fairness among the BSs in a no-SD-RAN network.

Having the benchmark model against which we are going to compare the results obtained with our approaches, we also need to mention that we are going to perform this comparison in terms of the minimum data rate in the whole network in relation to $\mathcal{P}_1$. The comparison in relation to $\mathcal{P}_2$ will be conducted in terms of the minimum BS throughput in the whole network.

6 PERFORMANCE EVALUATION

In this section, we first describe the simulation setup. Then, we validate our theoretical results on a 5G publicly-available trace. This is followed by performance comparisons between our two approaches, the benchmark, and another policy (Round-robin) for different scenarios.

6.1 Simulation setup

For input parameters we have used a 5G trace with data measured in the Republic of Ireland. These traces can be found in [20], with a detailed description in [9]. The parameter of interest from the trace is CQI (Channel Quality Indicator) with 15 levels, which serves to determine the per-block rate of a user in a slot. These measurements were conducted for one user, but at different days, for different applications, and when the user was static and moving around. To mimic the dynamic nature of these users, we have picked 8 users that are moving around, and assume they are all in the same cell. Based on the frequency of occurrence of a per-block rate for every user, we obtained the corresponding per-PRB rate probabilities (Table 2).

The slot duration is 0.5 ms. The subcarrier spacing is 30 KHz, with 12 subcarriers per block, making the PRB width 360 KHz. The total number of PRBs is $K = 273$ [5]. The simulations are conducted in MATLAB R2021b.

In the simulator, in each slot, every BS sends the information on the CQI of its UEs to the controller. With the full picture of all CQIs in the network, the controller, according to the allocation policy used, distributes the resources (PRBs) to BSs together with the information on how to further distribute them to UEs, and the BSs then assign the resources they have obtained to UEs within their coverage area. Depending on the amount of resources assigned, and its per-PRB rate, we determine the data rate each UE experiences in a slot.

6.2 Validation

We validate two of our analytical results (Eq.(14) and Eq.(27)) against simulations conducted on the aforementioned trace. We show results for four cases, which are:

- Case 1: 4 BSs; 3 UEs per BS.
- Case 2: 5 BSs; 2 UEs for BSs 1, 2, and 3, 3 UEs for BSs 4 and 5.
- Case 3: 6 BSs; 4 UEs per BS.
- Case 4: 8 BSs; 2 UEs for BS 1, 3 UEs for BSs 2 and 3, 4 UEs for BSs 4, 5 and 6, and 5 UEs for BSs 7 and 8.

The outcomes related to Eq.(14), i.e., to $\mathcal{P}_1$, are depicted in Fig. 2. Note that in all the cases, a UE is chosen randomly from one of the eight types of Table 2. Then, the CQI values in a slot are chosen according to the corresponding PMF distributions for each user from Table 2. As can be observed from Fig. 2, in all the cases, our analytical predictions match the simulation results exactly. The second thing to observe is that the objective value decreases with the number of UEs, which is to be expected as there will be fewer resources for all the users. Note that both in Case 1 and Case 2 there are 12 UEs in total.

The results obtained related to Eq.(27), i.e., to $\mathcal{P}_2$, are shown in Fig. 3 for the same input parameters and setup as those corresponding to the scenario of Fig. 2. Again, there is an exact match between theory and simulations, corroborating the accuracy of our analytical approach. As opposed to the previous results, when it comes to the maximized lowest BS throughput, it is more sensitive to the number of BSs than UEs. The rationale behind this is that the throughput in all BSs needs to be equal.

6.3 Performance comparisons

Having validated the correctness of our theoretical results, we proceed with comparing the performance of our approaches against the benchmark models. In the first case, we compare our approach for the maximized lowest data rate in an SD-RAN-enabled network (the solution to $\mathcal{P}_1$), to which we refer as SD-RAN in the

Table 2: Per-PRB rates and the corresponding probabilities for every user from the Republic of Ireland trace [9]

R (kbps)	48	73.6	121.8	192.2	282	378	474.2	712	772.2	874.8	1063.8	1249.6	1448.4	1640.6	1778.4
$p_{1,k}$	0	0	0	0	0	0	0.01	0.05	0.11	0.13	0.14	0.18	0.06	0.11	0.21
$p_{2,k}$	0	0	0	0	0	0.01	0.02	0.06	0.13	0.14	0.2	0.21	0.07	0.09	0.07
$p_{3,k}$	0.01	0	0	0	0	0.01	0.01	0.02	0.06	0.13	0.17	0.18	0.08	0.18	0.15
$p_{4,k}$	0	0	0	0	0	0.02	0.03	0.13	0.06	0.2	0.32	0.11	0.01	0.09	0.03
$p_{5,k}$	0	0	0	0	0	0	0.04	0.07	0.13	0.17	0.22	0.2	0.05	0.06	0.06
$p_{6,k}$	0	0	0	0	0.01	0.03	0.11	0.12	0.19	0.15	0.15	0.12	0.05	0.04	0.03
$p_{7,k}$	0	0	0	0	0	0	0.04	0.07	0.13	0.17	0.22	0.2	0.05	0.06	0.06
$p_{8,k}$	0	0	0	0	0.01	0.03	0.11	0.12	0.19	0.15	0.15	0.12	0.05	0.04	0.03

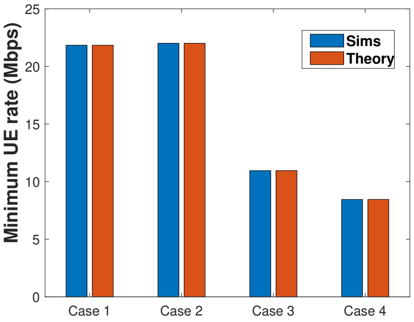

Figure 2: The average minimum data rate among all UEs with $\mathcal{P}_1$ for the four cases.

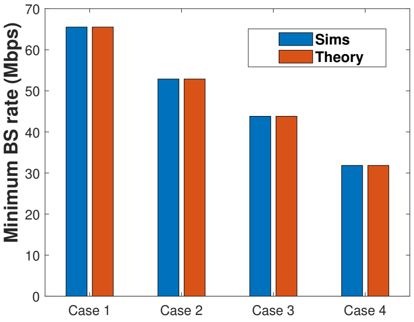

Figure 3: The average minimum BS throughput with $\mathcal{P}_2$ for the four cases.

figures, with the benchmark and another allocation policy in both of which there is no SD-RAN, i.e., each BS "owns" a fixed number of PRBs. The first is the previously described benchmark in Section 5, whose results are referred to as *no SD-RAN*. The second policy for comparison is Round-robin (RR) [21].

We show results for two cases: Case 1 and Case 4, introduced in Section 6.2. Fig. 4 portrays the results for the lowest data rate in the network over time with the three policies for Case 1. As can be observed, the solution to $\mathcal{P}_1$ always outperforms that of the benchmark $\mathcal{P}_0$, and RR. E.g., the solution to $\mathcal{P}_1$ provides a higher minimum rate than that of $\mathcal{P}_0$ in the range $10 - 60\%$. RR provides the worst results. This is even more emphasized in cases when most of the users in a slot experience good channel conditions.

Fig. 5 shows the results for Case 4. As there are more UEs in Case 4, the corresponding data rates are lower than in Case 1. Note that we are showing the results only across 30 slots. The reason is that if we increase the number of slots, the plots become overcrowded, and it is more difficult to discern the outcomes. Nevertheless, the same trend is observed (SD-RAN outperforms the other two policies in all slots.)

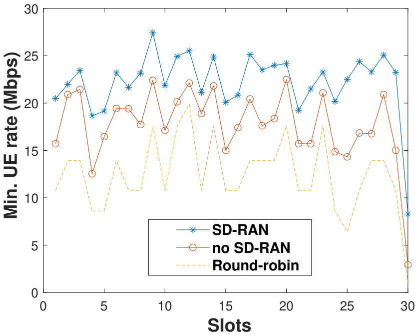

Figure 4: The evolution of minimum UE rate in Case 1.

Table 3: The coefficient of variation of per-PRB rates for users of Table 2

user	1	2	3	4	5	6	7	8
$c_{V,R}$	0.31	0.31	0.3	0.32	0.31	0.38	0.31	0.38

Fig. 6 shows the results for the lowest BS throughput in the network over time with two policies for Case 1. Our approach now uses the solution to $\mathcal{P}_2$ (and the results are referred to as *SD-RAN*). The results obtained with $\mathcal{P}_0$ are referred to as *no SD-RAN*. The other parameters remain unchanged from the previous scenario. Fig. 7 shows the result for Case 4. Both in Case 1 and Case 4 our approach outperforms the benchmark by at least 35%, and in some cases reaching almost 300%, which is a considerable gain. Also, the lowest cell throughput is much lower in Case 4 due to the higher number of BSs (in line with our previous discussion).

The effects shown in the previous results can be observed for other cases (different input parameters). Common to all these is that SD-RAN outperforms the other policies under all circumstances.

6.4 Impact of channel variability

Next, we look at the impact of channel variability (expressed through the variability of the per-block rate) on the variability of the assigned number of PRBs to UEs. We pick Case 4 to demonstrate the results for this scenario because there are more UEs in that case, and as we choose uniformly the user types from Table 2, all of them must be represented by the UEs. Our focus here is to look how

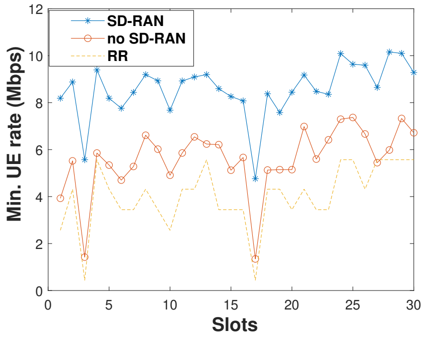

Figure 5: The evolution of minimum UE rate in Case 4.

Table 4: The coefficient of variation of the inverse of per-PRB rates for users from Table 2

user	1	2	3	4	5	6	7	8
$c_{V,\frac{1}{R}}$	0.36	0.37	1.88	0.37	0.34	0.43	0.34	0.43

much varies the number of assigned PRBs to users with different channel variability. To capture the latter quantitatively, we use the coefficient of variation (c_V), defined as the ratio of the standard deviation and the mean. Table 3 shows the coefficient of variation of the per-PRB rates for users of Table 2. As can be observed, all of them have roughly similar channel variability. We pick three of the users: user 1, user 2, and user 3, and introduce a "phony" user who only experiences a CQI of either 8 or 9 with equal probability. We do the latter to have a representative with very low channel variation. So, how varying is the number of PRBs assigned to users?

Fig. 8 depicts the values (marked as $\mathcal{P}_1$) of the coefficient of variation for the number of PRBs allocated to those four user types, where the allocation is performed in line with the solution of $\mathcal{P}_1$, i.e., to maximize the minimum value in the network with SD-RAN. On the same plot (marked as $\mathcal{P}_2$), we also show the results for the coefficient of variation of the number of assigned PRBs to those users, but following the policy for maximizing the minimum BS throughput in an SN-RAN environment, i.e., according to the solution of $\mathcal{P}_2$.

The interesting observation from Fig. 8 is that the number of assigned PRBs to user type 3 is much more varying than that of the other users. So, the coefficient of variation of the number of assigned PRBs does not really depend on the variability of the per-PRB rate. However, looking at Eqs.(11) and (23), it can be observed that *they both depend on the inverse of the per-PRB rate.*, i.e., on $\frac{1}{R_{i,j}}$.

Let us look next at the coefficient of variation of the inverse of the per-PRB rates of users from Table 2. Table 4 summarizes these results. From Fig. 8 and Table 4, it can be observed that users with higher variability *in the inverse of the per-PRB rate* experience higher variability (i.e., user type 3) in the number of assigned PRBs per slot. Note that the coefficient of variation of the inverse of the per-PRB rate of the phony user is only 0.12; hence a very low variability in the number of the assigned PRBs to her. The outcomes of this scenario apply equally to both the solution of $\mathcal{P}_1$ and $\mathcal{P}_2$.

6.5 Policy comparisons: Corner cases

So far, we have compared the allocation policies for various configurations, considering UEs with roughly similar distributions of CQI. Next, we demonstrate the difference in performance among the allocation policies in corner cases in terms of the channel conditions of UEs. To that end, we consider the following four scenarios (in all of them there are 4 BSs with 4 UEs each):

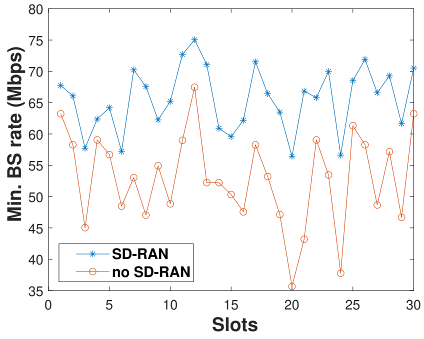

Figure 6: The evolution of minimum BS throughput for Case 1.

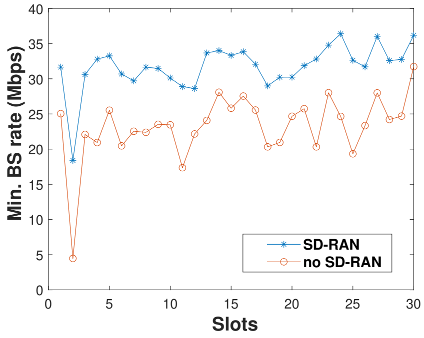

Figure 7: The evolution of minimum BS throughput for Case 4.

- Scenario A: All UEs in all BSs experience similar channel conditions. The CQI of every UE in a slot is drawn uniformly over the entire set.
- Scenario B: All UEs have excellent channel conditions (CQIs are drawn uniformly in the range $13-15$).
- Scenario C: All UEs have very bad channel conditions (CQIs are drawn uniformly in the range $1-3$).
- Scenario D: UEs in two of the BSs experience excellent channel conditions (CQIs are drawn uniformly in the range $13-15$), and UEs in the other two BSs suffer from bad channel conditions (CQIs are drawn uniformly in the range $1-3$).

Fig. 9 shows the average lowest data rate (over time) in the network with SD-RAN (solving $\mathcal{P}_1$), and no-SD-RAN (the outcome from the solution of $\mathcal{P}_0$) for the four scenarios. In all cases, the SD-RAN outperforms the no-SD-RAN. The difference in performance increases when there is a mix of UEs with good and bad channels (Scenario D), where the average lowest rates with $\mathcal{P}_1$ is more than 2× higher than with $\mathcal{P}_0$.

Fig. 10 depicts the average lowest BS throughput (over time) for the same setup as previously, but with the results obtained from $\mathcal{P}_2$. Again, in all scenarios, the SD-RAN outperforms the no-SD-RAN network. The difference between the two approaches increases in

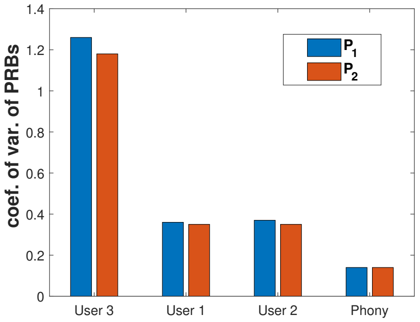

Figure 8: The coefficient of variation of the number of PRBs allocated to different users over time.

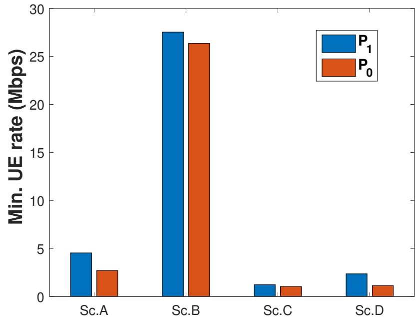

Figure 9: Minimum UE average data rate for special cases with $\mathcal{P}_1$ and $\mathcal{P}_0$.

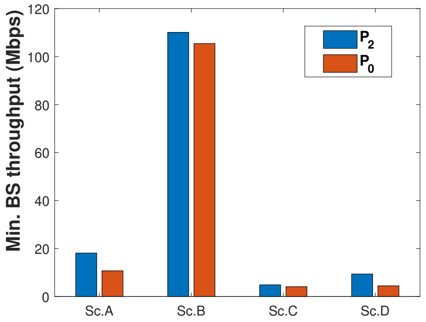

Figure 10: Minimum BS average throughput for special cases with $\mathcal{P}_2$ and $\mathcal{P}_0$.

Scenario D, surpassing 2×, as previously, which is a considerable improvement.

Therefore, we can conclude that using SD-RAN is always beneficial in terms of the lowest data rate across all UEs, and in terms of the lowest throughput across all BSs.

7 CONCLUSION

In this paper, we considered the problem of maximizing the lowest user throughput in an SD-RAN environment. We derived in closed form the allocation policies that should be performed to provide max-min fair resource allocation together with the achievable throughput. We did this for two scenarios. In the first, the goal was to provide max-min fair resource allocation across all the users in the network, whereas in the second scenario the objective was to guarantee max-min fairness among base stations, and within users of each base station separately. We evaluated the performance on real traces, and compared the performance with no-SD-RAN network and another allocation policy, showing the significant improvements the introduction of SD-RAN brings in terms of the performance of the worst-performing user in the network. In the future, we plan to consider other objectives, such as providing proportionally-fair resource allocation in a setup with SD-RAN, as well as to consider the general case of α-fairness.

ACKNOWLEDGEMENT

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF), in part under the project "6G-Life", with project identification number 16KISK002, and in part under the project "6G-ANNA", with project identification number 16KISK107.

REFERENCES

[1] L. Cui, R. Yu, and Q. Yan. When big data meets software-defined networking: SDN for big data and big data for SDN. *IEEE network*, 30(1), 2016.
[2] X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. Kontovasilis. FlexRan: A flexible and programmable platform for software-defined radio access networks. In *Proc. of ACM CoNEXT*, 2016.
[3] E. Coronado, S. N. Khan, and R. Riggio. 5G-EmPOWER: A software-defined networking platform for 5G radio access networks. *IEEE Transactions on Network and Service Management*, 16(2), 2019.
[4] A. Papa, R. Durner, L. Goratti, T. Rasheed, and W. Kellerer. Controlling Next-Generation Software-Defined RANs. *IEEE Communications Magazine*, 58(7), 2020.
[5] ETSI. 5G NR overall description: 3GPP TS 38.300 version 15.3.1 release 15. www.etsi.org, 2018. Technical specification.
[6] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.
[7] A. Goldsmith. *Wireless communications*. Cambridge University Press, 2005.
[8] F. Mehmeti and T. La Porta. Reducing the cost of consistency: Performance improvements in next generation cellular networks with optimal resource reallocation. *IEEE Transactions on Mobile Computing*, 21(7), 2022.
[9] D. Raca, D. Leahy, C. J. Sreenan, and J. J. Quinlan. Beyond throughput, the next generation: A 5G dataset with channel and context metrics. In *Proc. of ACM MMSys*, 2020.
[10] Z. Zaidi, V. Friderikos, and M. A. Imran. Future RAN architecture: SD-RAN through a general-purpose processing platform. *IEEE Vehicular Technology Magazine*, 10(1), 2015.
[11] Q. Qin, N. Choi, M. R. Rahman, M. Thottan, and L. Tassiulas. Network slicing in heterogeneous Software-defined RANs. In *Proc. of IEEE INFOCOM*, 2020.
[12] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng. OpenRAN: A software-defined RAN architecture via virtualization. *ACM SIGCOMM Computer Communication Review*, 43(4), 2013.
[13] A. Gudipati, D. Perry, L. E. Li, and S. Katti. SoftRAN: Software defined radio access network. In *Proc. of ACM SIGCOMM workshop on Hot topics in Software Defined Networking*, 2013.
[14] A. Papa, M. Klugel, L. Goratti, T. Rasheed, and W. Kellerer. Optimizing dynamic RAN slicing in programmable 5G networks. In *Proc. of IEEE ICC*, 2019.
[15] L. Zheng, D. W. H. Cai, and C. W. Tan. Max-min fairness rate control in wireless networks: Optimality and algorithms by Perron-Frobenius theory. *IEEE Transactions on Mobile Computing*, 17(1), 2018.
[16] A. Papa, A. Jano, S. Ayvaşık, O. Ayan, H. M. Gürsu, and W. Kellerer. User-based Quality of Service aware multi-cell radio access network slicing. *IEEE Transactions on Network and Service Management*, 19(1), 2022.
[17] A. Papa, P. Kutsevol, F. Mehmeti, and W. Kellerer. Effects of SD-RAN control plane design on user Quality of Service. In *Proc. of IEEE Netsoft*, 2022.
[18] G. Ku and J. M. Walsh. Resource allocation and link adaptation in LTE and LTE Advanced: A tutorial. *IEEE Communications Surveys & Tutorials*, 17(3), 2015.
[19] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
[20] https://github.com/uccmisl/5Gdataset.
[21] B. Li, A. Eryilmaz, and R. Srikant. Emulating round-robin in wireless networks. In *Proc. of ACM MOBIHOC*, 2017.