

Fairness and User Assignment in Cloud-RAN

Hadi Ghauch*, Sahar Imtiaz*, Mikael Skoglund*, George Koudouridis[†], James Gross*,

*School of Electrical Engineering and the ACCESS Linnaeus Center, Royal Institute of Technology (KTH)

[†] Radio Network Technology Research, Huawei Technologies, Kista, Sweden

Abstract—In this paper, we extend our previous work on user assignment in Cloud-RAN, where we proposed an algorithm for user assignment (UA). We motivate the inherent fairness issue that is present in the latter UA scheme, since some users in the system will never get served. To improve the fairness, we propose that the UA scheme is preceded by a user scheduling step which aims at selecting at any time the users that should be considered by the UA algorithm for scheduling (in the next time slot). Two user scheduling approaches have been studied. The first scheme improves the minimum throughput (MT), by selecting at any time the users with the lowest throughput. The second scheme is based on round-robin (RR) scheduling, where the set of potentially scheduled users for the next slot, is done by excluding all the previously served users, in that round. Moreover, the subset of actual users to be served, is determined using the UA algorithm. We evaluate their fairness and sum-rate performance, via extensive simulations. While one might have expected a tradeoff between the sum-rate performance and fairness, our results show that MT improves both metrics, when compared to the original UA algorithm (without fairness), for some choice of parameter values. This implies that both fairness and aggregate system performance can be improved, by a careful choice of the number of assigned and served users.

Index Terms—User Assignment ; Cloud Radio Access Networks ; Fairness Scheduling ;

I. INTRODUCTION

Historically, most of the gains in data rates are due to densification [1], by increasing the number of base stations, antennas, and cells. Evidently, mechanisms for coordinating the increased number of (Base Stations) BSs and antennas, are required. In that sense, Cloud Radio Access Network (Cloud-RAN) is one of the most promising architectures, for achieving densification. In Cloud-RAN systems, radio-heads have limited baseband capabilities, and are connected to central processors (CP) (also referred to as aggregation node) via high-capacity wired/wireless backhaul links. Moreover, an antenna domain (AD) is a collection of radio-heads, connected to a particular central processor. Essentially, the central processor gathers all the required (Channel State Information) CSI, performs a set of tasks (e.g., precoder optimization), and communicates the decision to the radio-heads (and users potentially).

The first ideas behind Cloud-RAN originate in investigating setups with interconnected radio heads [2]. The problem of dynamic radio-head clustering has been addressed in several earlier works, where caching and multi-cast transmission are assumed [3], [4], and energy efficiency is optimized [5]. While most previous work on Cloud-RAN focuses on systems with one central processor, in our earlier works [6], [7],

we addressed the problem of user assignment in multi-CP / multi-AD settings, in order to mitigate intra- and inter-AD interference.

More specifically, in [6], we investigated the optimal assignment of users to antenna domains (ADs), using the total interference leakage as a metric. We formulated the problem as an integer programming problem, and proposed an iterative algorithm based on the known block-coordinate descent (BCD) method for obtaining such an assignment. Moreover, for each user-to-AD assignment, we proposed a precoding mechanism (at the radio-heads of each AD) aimed at completely suppressing the intra-AD interference. Starting from an initial/input set of users, the proposed UA algorithm selects the ‘best’ subset of users. This selection feature inevitably leads to fairness problems, as some users will not get selected. Our aim in this work is to address such a limitation, by modifying the UA scheme to enforce fairness.

We propose a first scheduling method, based on minimum throughput, where users with lowest throughputs are given to the UA algorithm (to be potentially served during the next scheduling time-slot). We also propose another scheduling method based on round-robin (RR), where the previously served users are excluded from the set of potentially served users for the current time slot. The proposed UA scheme is then used to select the set of served users. While imposing fairness comes at the cost of sum-rate performance generally, we observe that this is not the case here. More specifically in the case of min-throughput scheduling, interestingly we observe that this is not the case in the studied 5G Cloud-RAN scenario, for specific parameter values. The rest of the paper is structured as follows: We present the system model in Sec. II, recap the previously proposed scheme from [7] in Sec. III, detail the two proposed fairness schemes in Sec IV, and present the numerical results in Sec. V.

II. SYSTEM MODEL

Notation: We use bold upper-case letters to denote matrices, bold lower-case to denote vectors, and calligraphic letters to denote sets. Furthermore, for a given vector \mathbf{a} , $\|\mathbf{a}\|_2^2$ denotes its l_2 norm, \mathbf{a}^T its transpose, \mathbf{a}^\dagger its conjugate transpose, and $[\mathbf{a}]_i$ its i th element. \mathbb{B} denotes the binary set, and \mathbb{Z}_+ denotes the set of natural numbers. Given a set \mathcal{X} , $|\mathcal{X}|$ denotes its cardinality, and $\mathcal{X} - \mathcal{Y}$ the set difference operation.

We investigate a typical Cloud-RAN setup comprising of A central processors, N_T radio-heads, and U_T users. We follow the same naming conventions introduced in our earlier

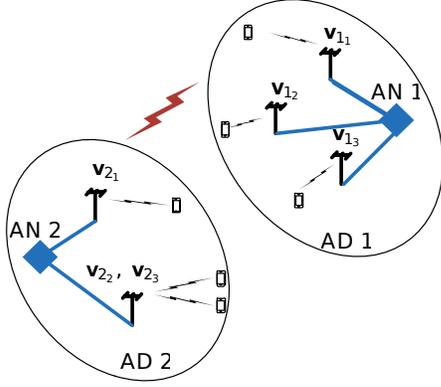


Fig. 1: System Model

works [6], [7]. Each central processor is connected to N radio-heads, and each radio-head is serving a set of users. Moreover, an *antenna domain* consists of the set of radio-heads connected to a particular central processor. Each antenna domain comprises N radio-heads and U users. We can thus abstract the operation of the radio-heads within each antenna domain: in that sense, one can say that users are assigned to antenna domains directly (rather than radio heads). This system model is illustrated in Fig. 1.

For simplicity, we denote by j_n the index of the n th user, in the j th antenna domain, $j_n \in \mathcal{U}_T$ (henceforth referred to as user j_n). We assume a *downlink transmission* scenario, where the received signal at user j_n is given by,

$$y_{j_n}^{(t)} = \sum_{q \in \mathcal{U}} \mathbf{h}_{j,j_n}^{(t)} \mathbf{v}_{j_q}^{(t)} s_{j_q}^{(t)} \sqrt{p} + \sum_{\substack{i_m \in \mathcal{U}_T \\ i_m \neq j_n}} \mathbf{h}_{i,j_n}^{(t)} \mathbf{v}_{i_m}^{(t)} s_{i_m}^{(t)} \sqrt{p} + n_{j_n}^{(t)} \quad (1)$$

where $\mathbf{h}_{i,j_n}^{(t)} \in \mathbb{C}^{1 \times MN}$ is the Multiple-Input Single Output (MISO) channel from antenna domain i to user j_n at time slot t , $\mathbf{v}_{i_m}^{(t)} \in \mathbb{C}^{MN \times 1}$ the beamforming vector to user $i_m \in \mathcal{U}_T$ at time slots t , $s_{i_m}^{(t)}$ the data symbol for user $i_m \in \mathcal{U}_T$ in time slot t , $n_{j_n}^{(t)}$ the AWGN noise for user $j_n \in \mathcal{U}_T$, during time slot t (with variance $\sigma_{j_n}^2$), and p the transmit power for each antenna domain. We consider a block fading channel model, where channels remain essentially constant, for the duration of communication.

Moreover, the precoding is done to suppress all intra-AD interference, as follows,

$$\mathbf{h}_{j,j_n}^{(t)} \mathbf{v}_{j_q}^{(t)} = \begin{cases} \beta_j^{(t)}, & \forall n = q \\ 0, & \forall n \neq q \end{cases}, \quad \forall j \in \mathcal{A} \quad (2)$$

where $\beta_j^{(t)} > 0$ is a free parameter that is chosen to satisfy the maximum transmit power constraint (per antenna domain), i.e., $\sum_{q \in \mathcal{U}} \|\mathbf{v}_{j_q}^{(t)}\|_2^2 \leq U$, $\forall j \in \mathcal{A}, \forall t$. Consequently, it is easy to verify that the received signal and SINR, for user $j_n \in \mathcal{U}_T$

during time slot t , is given by:

$$y_{j_n}^{(t)} = \beta_j s_{j_n}^{(t)} \sqrt{p} + \sum_{\substack{i_m \in \mathcal{U}_T \\ i_m \neq j_n}} \mathbf{h}_{i,j_n}^{(t)} \mathbf{v}_{i_m}^{(t)} s_{i_m}^{(t)} \sqrt{p} + n_{j_n}^{(t)} \\ \gamma_{j_n}^{(t)} = \frac{\beta_j^{(t)^2} p}{\sum_{\substack{i_m \in \mathcal{U}_T \\ i_m \neq j_n}} p |\mathbf{h}_{i,j_n}^{(t)} \mathbf{v}_{i_m}^{(t)}|^2 + \sigma_{j_n}^2} \quad (3)$$

A. Recap of User Assignment Algorithm

Starting from an initial assignment (i.e., users to radio-heads, and radio-heads to central processors), the UA problem is formulated as the optimal assignment of users to antenna domains, w.r.t. minimizing the total interference leakage in the system [7]. While maximizing the sum-rate of the network is the primary interest, finding the user assignment that maximizes the sum-rate is known to be NP-hard [8]. Moreover, future C-RAN deployments are likely to be dense, and thus interference-limited. Consequently, we opt to focus on minimizing the interference leakage. We summarize below the main steps of the UA algorithm, for completeness:

- Start with an initial input set of users, \mathcal{U}_T .
- Compute the matrix of coupling coefficients.
- Find user-to-AD assignment that minimizes the interference leakage, \mathcal{U}_s .
- Ensure that ρ users are served per AD, i.e., $|\mathcal{U}_s| = \rho$ where ρ is the loading factor for each AD.

For the purpose of this work, we can abstract the UA algorithm's operation, as the following function:

$$\mathcal{S}^{(t)} = UA(\mathcal{U}_T) \quad (4)$$

where \mathcal{U}_T is the input set (e.g., the set of all users in the network), $\mathcal{S}^{(t)}$ the subset of users selected by the algorithm at time-slot t ($|\mathcal{S}^{(t)}| = \rho$), and we neglect other input parameters such as CSI, for simplicity. Once the assignment is performed, the achievable rate of user $j_n \in \mathcal{S}^{(t)}$ for time-slot t , is given as,

$$r_{j_n}^{(t)} = \log_2(1 + \gamma_{j_n}^{(t)}), \quad (\text{bps/Hz}) \quad (5)$$

The above expression holds for the scheduled users, $j_n \in \mathcal{S}^{(t)}$, thereby implying that non-schedule users, $j_n \in \mathcal{U}_T - \mathcal{S}^{(t)}$, have zero rates. Essentially, during each time-slot, the UA algorithm selects a subset of the total number of users, that minimizes the total interference leakage. The algorithm does not take into account whether the remaining unscheduled users are ever served or not: this naturally raises concerns about its fairness.

III. PROPOSED SCHEMES FOR FAIRNESS

In this section, we investigate two different mechanisms for stimulating fairness, of the above UA method. The first scheme consists of preselecting users that have small throughput during previous time-slots, and providing that as input to the UA algorithm to select the subset that is served. In the second scheme, the pre-selection is based on round robin, and later provided as input to the UA algorithm (that selects the subset to be served).

A. Min-Throughput Scheduling

We investigate a scheduling scheme based on the achieved throughput of each user, during the previous time-slots. In the context of this algorithm, the *throughput* of user j_n for time-slot $t - 1$, is defined as its average rate until time-slot $t - 1$,

$$\bar{r}_{j_n}^{(t-1)} = \left(\sum_{l=1}^{t-1} r_{j_n}^{(l)} \right) / (t - 1) \quad (6)$$

where $r_{j_n}^{(l)}$ is the rate of user j_n (defined in (5)).

During time slot $t - 1$, the scheme essentially consists of ranking all users, based on their throughput from time-slot $t - 1$ ($\bar{r}_{j_n}^{(t-1)}$ in (6)), and discarding users with highest throughputs for the next scheduling time slot, t . Let $\mathcal{R}^{(t-1)}$ denote the (ordered) set of users indexes, whose corresponding throughput values (for time-slot $t - 1$), are sorted in increasing order. Stated differently $\mathcal{R}^{(t-1)}$ consists of the indexes of all users in the systems, sorted in increasing order of throughput ($\bar{r}_{1_1}^{(t-1)} \leq \bar{r}_{1_2}^{(t-1)} \leq \dots$). We also define $\mathcal{W}_K^{(t-1)}$ as the set of worst users (i.e., users with lowest rates) in time slot $t - 1$: it consists of the first K in $\mathcal{R}^{(t-1)}$, and is expressed as,

$$\mathcal{W}_K^{(t-1)} = \{\mathcal{R}^{(t-1)}\{1\}, \dots, \mathcal{R}^{(t-1)}\{K\}\} \quad (7)$$

We reiterate the aim of the algorithm, that is to support the weakest terminals, in terms of throughput during previous time-slots. From that perspective, $\mathcal{W}_K^{(t-1)}$ can be seen as the set of *potentially* scheduled users for the next slot. We use the UA algorithm to find the set of users that will be served in the next slot. Then, $\mathcal{W}_K^{(t-1)}$ is given as input to the UA algorithm, as well as ρ , channel information $\{\mathbf{h}_{k,j_n}\}$, and other system parameters. Finally, the UA algorithm is used to select the best subset of users, to be served in the next slot, \mathcal{S}_t , i.e.,

$$\mathcal{S}^{(t)} = UA \left(\mathcal{W}_K^{(t-1)} \right) \quad (8)$$

Recall that ρ users are served in each time slot, i.e. $|\mathcal{S}^{(t)}| = \rho$. The selected users are served, and the corresponding sum-rate at time-slot t is given by,

$$R_\Sigma^{(t)} = \sum_{j_n \in \mathcal{S}^{(t)}} r_{j_n}^{(t)} \quad (9)$$

The scheme is detailed in Algorithm 1.

Remark 1. In some cases, more than K -users can have the same throughput: to break such a tie, we pick the bottom K -users that have the lowest user index n (for simplicity).

B. Round-Robin Scheduling

In th part, we propose another scheme to improve the fairness of the UA algorithm. A pre-selection of users is performed first, based on Round-Robin (RR). The resulting set is then given as input to the UA algorithm, to select the users to be served. Recall that UA has a built-in user selection feature: starting with an input of K users, UA selects the best subset ρ (where $\rho < K$) that minimizes the interference leakage, and optimizes the corresponding precoders. Our proposal is to exploit this built-in user selection feature, to serve non-overlapping subsets of the total user set, in consecutive time slots. Let G denote the number of such user groups (we assume

Algorithm 1 Min Throughput Scheduling (MT)

// Input from time-slot $t - 1$

Compute the throughput of all users, $\bar{r}_{j_n}^{(t-1)}$

// At time-slot t

Sort throughputs to obtain set of ordered user indexes, $\mathcal{R}^{(t-1)}$

Find users with lowest throughput, as first K elements in $\mathcal{R}^{(t-1)}$:

$$\mathcal{W}_K^{(t-1)} = \{\mathcal{R}^{(t-1)}\{1\}, \dots, \mathcal{R}^{(t-1)}\{K\}\}$$

Run UA algorithm with $\mathcal{W}_K^{(t-1)}$ as input, and find subset of ρ best users from $\mathcal{W}_K^{(t-1)}$:

$$\mathcal{S}^{(t)} = UA \left(\mathcal{W}_K^{(t-1)} \right)$$

that the total number of users is divisible by G). During each round, the users in each time slot are determined by excluding users that have been served, in the current round. Moreover, within each round, the set of scheduled users are non-overlapping. Let $\mathcal{S}^{(t)}$ denote the set of users served during time-slot t .¹ Recall that \mathcal{U}_T is the total number of users in the system. Then the proposed scheme can be described as follows. For instance, the first round is shown below:

- At $t = 1$:
run UA with \mathcal{U}_T as input to obtain $\mathcal{S}^{(t)}$, i.e., $\mathcal{S}^{(t)} = UA(\mathcal{U}_T)$. Serve users in $\mathcal{S}^{(t)}$.
- At $t = 2$:
run UA with $\mathcal{U}_T - \mathcal{S}^{(t-1)}$ as input to obtain $\mathcal{S}^{(t)}$, i.e., $\mathcal{S}^{(t)} = UA(\mathcal{U}_T - \mathcal{S}^{(t-1)})$. Serve users in $\mathcal{S}^{(t)}$.
- At $t = G$:
run UA with $\mathcal{U}_T - (\cup_{l=1}^{G-1} \mathcal{S}^{(l)})$ as input to obtain $\mathcal{S}^{(t)}$, i.e., $\mathcal{S}^{(t)} = UA(\mathcal{U}_T - (\cup_{l=1}^{G-1} \mathcal{S}^{(l)}))$. Serve users in $\mathcal{S}^{(t)}$.

where $\mathcal{X} - \mathcal{Y}$ denotes the set difference. After $t = G$ time-slots the first round is over, and the algorithm goes back to selecting from the full set of users (due to its round-robin nature). The detailed steps are shown in Algorithm 2, for round r ($r = 1, 2, \dots$) of the algorithm. The sum-rate at the end of time slot t (needed for the numerical results), is given by,

$$R_\Sigma^{(t)} = \sum_{j_n \in \mathcal{S}^{(t)}} r_{j_n}^{(t)} \quad (10)$$

Remark 2. In contrast to ‘pure’ round robin, where the selected users are predetermined and fixed across different rounds, the sequence of scheduled users is potentially different for different rounds in Algorithm 2, due to the additional UA selection.

¹Although $\mathcal{S}^{(t)}$ denotes the set of selected users for both schemes, the elements in $\mathcal{S}^{(t)}$ are obviously different for the two schemes (and their respective sum-rates as well).

Algorithm 2 Round-Robin Scheduling (RR)

```

// Round number  $r$  (where  $r = 1, 2, 3, \dots$ )
At time-slot  $t = (r - 1)G + 1$  :
  Find  $\mathcal{S}^{(t)} = UA(\mathcal{U}_T)$ . Serve users in  $\mathcal{S}^{(t)}$ 
At time-slot  $t = (r - 1)G + 2$  :
  Find  $\mathcal{S}^{(t)} = UA(\mathcal{U}_T - \mathcal{S}^{(t-1)})$ . Serve users in  $\mathcal{S}^{(t)}$ 
:
At  $t = rG$  :
  Find  $\mathcal{S}^{(t)} = UA\left(\mathcal{U}_T - \left(\cup_{l=(r-1)G+1}^{t-1} \mathcal{S}^{(l)}\right)\right)$ . Serve
  users in  $\mathcal{S}^{(t)}$ 
// End of round
  
```

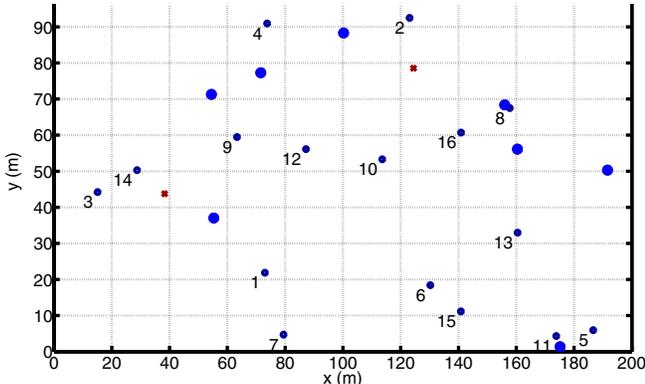


Fig. 2: Location of users (small dots), radio-heads (large dots) and aggregation nodes (crosses)

IV. NUMERICAL RESULTS

A. Simulation Methodology

In this section we compare the performance of both schemes in terms of fairness and sum-rate performance. The simulation setup and parameterization is described in full details, in [9]. We consider a system with $A = 2$ AD (central processor), $N = 4$ radio heads per AD, $M = 4$ antennas per radio-head, $K = 8$ users per AD, for a total of $U_T = 16$ users. Positions of users, radio-heads, and central processors are generated at random, and fixed throughout the simulation. The corresponding geometry is shown in Fig. 2, where all the users are numbered from 1 to 16. All results are averaged over 500 different channel realizations. Each channel realization follow a Rician LoS model (Refer to [9] for the parameterization). As a reference scheme, we use the original UA scheme (without any fairness), while setting $\rho = 4$.

B. Simulation Results

We initially investigate the sum-rate CDF of the proposed schemes for several values of ρ (in the case of the Min-Throughput (MT) scheme), and G (in the case of the Round-Robin (RR) scheme). The results are shown in Fig. 3 (we also show the average sum-rate value for each of the schemes). Note that we see a huge improvement by going from MT with $\rho = 2$, to MT with $\rho = 4$, which implies that the choice of loading factor is extremely important. We also observe a

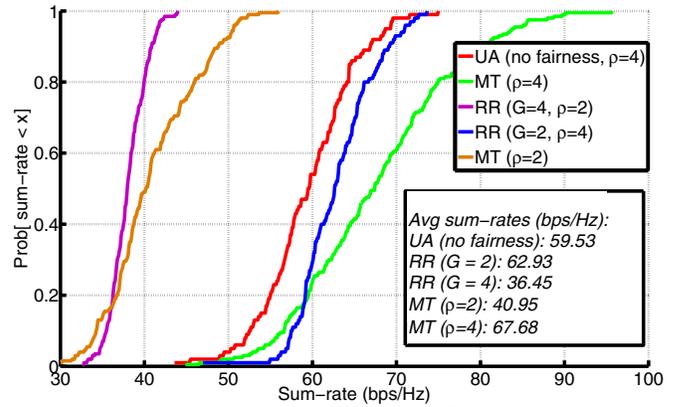


Fig. 3: CDF of sum-rate (system with $A = 2, M = 4, N = 4, K = 8, K_T = 16$)

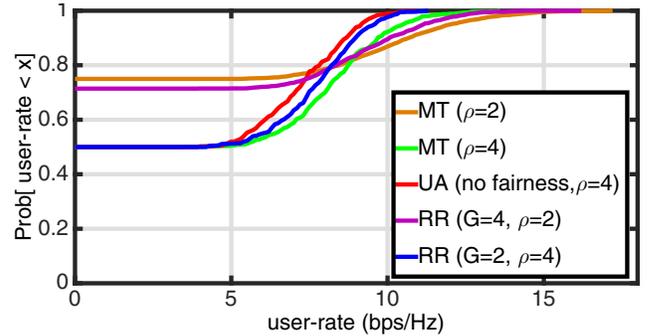


Fig. 4: CDF of user-rate (system with $A = 2, M = 4, N = 4, K = 8, K_T = 16$)

degradation in sum-rate performance by going from RR with $G = 2$, to RR with $G = 4$: this is due to the non-overlapping access nature of the scheme, that results in a loss in degrees-of-freedom. We note that this trend is still valid beyond $G = 4$, though not shown here.

We also compute the per-user CDF, for the same setup, where the averaging is performed across different channel realizations, and users. The result in Fig. 4 clearly reveals that schemes such as RR ($G = 4, \rho = 2$) and MT with $\rho = 2$, suffer from a significantly elevated probability, that the user rate is zero (due to the fact that only $\rho = 2$ users/per AD are served in each time slot, thus implying that the remaining 6 will have zero rates). Moreover, schemes such as RR ($G = 2, \rho = 4$) and MT with $\rho = 4$, significantly improve this issue, as they serve only 4 users get zero rates during the each time slot. In addition, we underline that user rate CDF is slightly better for MT with $\rho = 4$, compared to RR ($G = 2, \rho = 4$). We include in Fig. 5 the CDF of average user-rate, by first computing the average rate for each user and finding the corresponding CDF.

The results in the above figure show that MT with $\rho = 4$ provides the best sum-rate performance, out of all the other schemes. We also ‘zoom-in’ on the spatial distribution of the sum-rate, by considering the average rate for each of the users in the system. This provides insight into fairness across the users. The simulation setup is still the same as above, and the

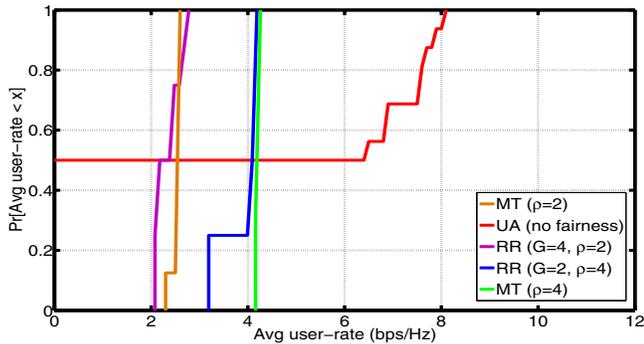


Fig. 5: CDF of average user-rate (system with $A = 2$, $M = 4$, $N = 4$, $K = 8$, $K_T = 16$)

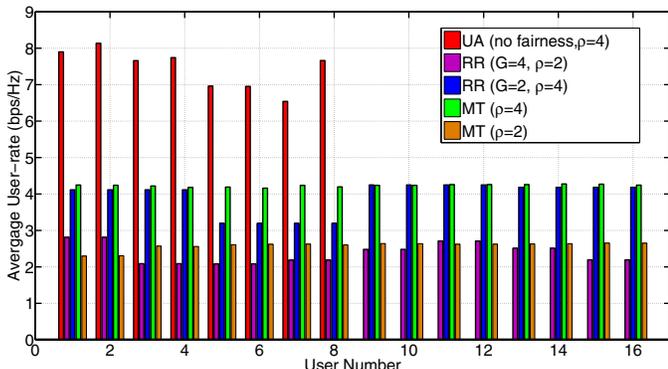


Fig. 6: Average rate for each user (system with $A = 2$, $M = 4$, $N = 4$, $K = 8$, $K_T = 16$)

results are shown in Fig. 6. It becomes clear that MT with ($\rho = 4$) provides perfect fairness. Moreover, the case of MT with $\rho = 2$ is a close second, in terms of fairness. However, it comes at the cost of lower user rates.

C. Discussions

Looking at the sum-rate CDF in Fig 3 we see that MT with $\rho = 4$ improves the performance of the UA algorithm (w/out fairness). Moreover, we also observe in Fig. 6 and Fig. 4, that this scheme also provides the best fairness. Interestingly, the latter results reveal a counter-intuitive observation: while one would have expected a tradeoff between the sum-rate performance and fairness, we were able to improve on both. This stems from the existence of an ‘optimal’ value for the size of served users, ρ , which cannot be too small - as it leads to a loss in degrees-of-freedom that degrades the sum-rate performance, and not too large - as it leads to significant interference that also degrades the sum-rate performance. This in turn implies that aggregate system performance and ‘cell-edge’ fairness, can simultaneously be improved, by carefully selecting the number of scheduled users ρ and the number of assigned users, in each time slot. More importantly, when looking at the original UA scheme in Fig. 6, we see that it suffers from severe fairness problems, in the sense that some users never get served. This reinforces the main motivation of the present work.

V. CONCLUSION

We have extended in this work our previous UA algorithm, to deal with the lack of fairness. We have investigated two schemes, to be used in conjunction with the UA algorithm, for imposing fairness: While the first scheme is based on improving the minimum throughput, the second one builds on round-robin, where the set of potentially served users is determined by excluding all the previously served users, in the current round. We relied on simulating several performance metrics (e.g., sum-rate CDF, user-rate CDF, average user rate), to evaluate their performance. A main take-home message is that both fairness and sum-rate performance can be improved, by careful choice of the number of assigned and served users (refer to above discussion on balancing number of selected users), while keeping the complexity of the algorithm virtually the same. While one would have expected a tradeoff between fairness and sum-rate performance, MT with $\rho = 4$ delivers gains over the original UA algorithm (w/out fairness), along both metrics.

REFERENCES

- [1] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, “Is the phy layer dead?,” *IEEE Communications Magazine*, vol. 49, pp. 159–165, April 2011.
- [2] J. Zhao, T. Q. S. Quek, and Z. Lei, “Coordinated multipoint transmission with limited backhaul data transfer,” *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2762–2775, June 2013.
- [3] H. Zhou, M. Tao, E. Chen, and W. Yu, “Content-centric multicast beamforming in cache-enabled cloud radio access networks,” in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2015.
- [4] M. Tao, E. Chen, H. Zhou, and W. Yu, “Content-centric sparse multicast beamforming for cache-enabled cloud RAN,” *CoRR*, vol. abs/1512.06938, 2015.
- [5] B. Dai and W. Yu, “Energy efficiency of downlink transmission strategies for cloud radio access networks,” *CoRR*, vol. abs/1601.01070, 2016.
- [6] H. Ghauch, M. Rahman, S. Imtiaz, and J. Gross, “Coordination and antenna domain formation in Cloud-RAN systems,” *IEEE International Communications Conference ICC*, 2016.
- [7] H. Ghauch, M. M. U. Rahman, S. Imtiaz, J. Gross, M. Skoglund, and C. Qvarfordt, “User assignment in cloud-ran: Algorithms and performance bounds,” *IEEE Transactions on Wireless Communications*, Submitted, available at <http://arxiv.org/pdf/1606.08401.pdf>, 2016.
- [8] M. Sanjabi, M. Razaviyayn, and Z.-Q. Luo, “Optimal joint base station assignment and beamforming for heterogeneous networks,” *IEEE Transactions on Signal Processing*, vol. 62, pp. 1950–1961, April 2014.
- [9] M. M. U. Rahman, H. Ghauch, S. Imtiaz, and J. Gross, “RRH clustering and transmit precoding for interference-limited 5G CRAN downlink,” in *IEEE Globecom Workshops (GC Wkshps)*, pp. 1–7, Dec 2015.