On Short Term Fairness and Throughput of User Clustering for Downlink Non-Orthogonal Multiple Access System

Mohanad M. Al-Wani, A. Sali, Asem A. Salah, Borhanuddin M. Ali, Nor K. Noordin, S. J. Hashim
Wireless and Photonic Networks Research Centre of Excellence (WiPNet), Dept. of Computer and Communication Systems Engineering, Faculty of Engineering, UPM, 43400 Serdang, Selangor
moheng84anad@gmail.com
{aduwati, asem_a, borhan, nkordin, sjh}@upm.edu.my

Keivan Navaie*, Chee Yen Leow**
*School of Computing and Communications, Lancaster University, LA1 4WA, UK
k.navaie@lancaster.ac.uk
**Wireless Communication Center, Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Skudai, 81310 Johor, Malaysia
bruceleow@fke.utm.my

Abstract—Non-orthogonal multiple access (NOMA) with power-domain user multiplexing has been considered as one of the potential candidates of the fifth Generation (5G) systems. In this paper, a two-stage user selection algorithm is proposed based on proportional fairness for downlink NOMA with zero forcing beamforming (PF-NOMA-ZFBF) in order to improve the throughput-fairness trade-off for NOMA system. We focus on short term fairness, where short term refers to the minimum time window in which a specified fairness is guaranteed and evaluated using Jain’s index. A transmit power allocation then applied to enhance the throughput of NOMA system. Simulation results show that the proposed PF-NOMA-ZFBF significantly improves user fairness index with at least 38.96% over conventional NOMA-ZFBF, while maintaining the total system sum-rate (near maximum).

Index Terms—NOMA, short term fairness, proportional fairness, zero forcing beamforming, power allocation, 5G.

I. INTRODUCTION

With the new emerging concepts and services such as, internet of things (IoT), cloud computing services, dense small cells, the fifth generation (5G) wireless networks, need not only to support higher data traffic, but also new network capabilities that can support massive numbers of connected devices with diverse quality of service (QoS), higher spectral efficiency, and lower-latency, which are difficult to be satisfied with orthogonal multiple access (OMA) schemes that are currently being employed in 4G networks [1], [2].

Recently, non-orthogonal multiple access (NOMA) has been considered an alternative technology that can improve the spectral efficiency and increasing number of connections, by taking advantages of user grouping and multiplexing in the same time/frequency resources. NOMA exploits power domain to generate superimposed signals at the transmitter side and applying successive interference cancellation (SIC) for inter-user interference cancellation at the receiver side [3], [4].

The earlier works show that NOMA can offer a considerable performance gain over OMA in terms of spectral efficiency and outage probability [4], [5]. The work in [4] investigates the performance of NOMA in terms of overall cell throughput and cell-edge user throughput using proportional fairness (PF) scheduler. In [6] three power allocation methods are proposed to achieve a better tradeoff between total system throughput and cell-edge user throughput using proportional fairness. In [7] a two-step method for user pairing and power allocation was proposed for downlink NOMA based on PF.

Most of the aforementioned works focus on the optimal resource allocation for sum rate maximization. However, in some cases, user fairness also considered as an important objective that must be considered for resource allocation. In [8] an efficient algorithm was proposed, to maximize the achieved NOMA throughput while guaranteeing fairness in resource allocation by incorporating proportional fairness scheduler for long-term fairness.

The performance of NOMA can be further enhanced in terms of throughput and number of served users by using beamforming (NOMA-BF) that allows at least two users to share a single beamforming vector, known as a cluster. User selection (grouping) for each cluster is a very important issue and can highly affect on NOMA performance in terms of system throughput, cell-edge user throughput, and fairness. The authors in [9] proposed NOMA based zero-forcing beamforming (NOMA-ZFBF), in which highly correlated users are firstly selected as candidates to make a cluster. After that, two users with the largest channel gain difference were chosen for a cluster. However, their results showed that weak users sum-rate decreases as the number of users increases.

In [10] NOMA-ZFBF is improved in terms of system and weak user throughputs by using two methods of user selection for each cluster. The selection process must be sequentially performed as follows: first, strong users are selected using semi-orthogonal user selection (SUS) algorithm [11], then, beamforming vectors are calculated based on the channels of strong users, and finally, weak users are selected using matching user algorithm based on minimum inter-cluster interference. In [12] fairness oriented user selection (FOUS) algo-
Algorithm was proposed for user grouping based on their NOMA data rates. However, FOUS algorithm attempt to enhance both throughput and fairness, which can only be achieved with long-term fairness (long time intervals) especially for large number of users, whereas, for short-term fairness (short time intervals), high level of fairness index is not easy to be satisfied, and there is a clear trade-off between fairness and throughput.

In this paper, to achieve a better trade-off, we propose a two-stage user selection algorithm for both strong and weak users based on proportional fairness for downlink NOMA-ZFBF (referred to as PF-NOMA-ZFBF), adopting the basic system model in [10]. Unlike the previous works that studied fairness over long time scales, this proposed scheme aims to increase fairness of resource allocation in short time scales with maintaining a high level of system throughput. To achieve this, we made a selection for \( t_c \) parameter \( (t_c \) is responsible for setting balance between throughput and fairness for short time scales), this selection guarantees a highest possible system throughput with a better enhancement for user fairness. Finally, we find the required time interval to attain a specified short term fairness. Numerical results show that the proposed PF-NOMA-ZFBF can achieve a higher short-term fairness while maintaining a high level of system sum-rate (near maximum) compared to the previous method [10].

The rest of this paper is organized as follows; the system model is described in Section II. Section III presents the received signal for NOMA-ZFBF users. The multi-user selection using PF algorithm is presented in Section IV and the proposed algorithm is discussed in Section V. In Section VI transmit power allocation is disused. Section VII presents the simulation results and Section VIII Conclusions are discussed.

II. SYSTEM MODEL

In this paper, a single-cell downlink multi-user beamforming (BF) NOMA system is considered. The base station (BS) is equipped with \( N \) antennas and can simultaneously send \( N \) beams to \( N \) clusters using ZFBF. Each cluster contains two single-antenna mobile users as shown in Fig. 1. The total number of candidate users is \( K \) \((K \geq 2N)\). The candidate users set is defined as \( U \) \([|U| = K]\). It is assumed that the channel is uncorrelated Rayleigh flat fading and the noise is independent and identically distributed (i.i.d) additive white Gaussian with zero mean and unit variance. Power domain NOMA system with SIC detection is performed in each cluster to detect superimposed signals by grouping two users with different channel gains. The user with a high channel gain is named as a strong user (also called ZFBF user) and the other with smaller channel gain is known as a weak user (also called matched user). ZFBF is designed based on the channel of strong user. The selection sets for strong and weak users are denoted by \( k_1 \) and \( k_2 \) respectively, where \( k_1 = S_{zf}(k) \) and \( k_2 = S_{wu}(k) \).

In the downlink NOMA-ZFBF, the transmitted superposition coded signal for the strong and weak users in the \( k^{th} \) cluster is:

\[
x_k = \sqrt{P} w_k (\sqrt{\alpha_{k1}} s_{k1} + \sqrt{\alpha_{k2}} s_{k2}),
\]

where \( s_{k1} \) and \( s_{k2} \) are the information symbols for strong and weak user respectively, \( \alpha_{k1} \), \( \alpha_{k2} \) are the NOMA power allocation factors for the strong and the weak user respectively, with \( \alpha_{k1} + \alpha_{k2} = 1 \) and \( \alpha_{k2} > \alpha_{k1} \). \( P \) and \( w_k \) are the total cluster transmits power and the \( N \times 1 \) ZFBF vector for the cluster \( k \), respectively.

III. RECEIVED SIGNAL FOR NOMA-ZFBF USERS

In each NOMA-ZFBF cluster, there are two types of interferences, they are, inter-cluster interference created by users of other beams and intra-cluster interference which occurs between users sharing the same beam in a cluster. Inter-cluster interference can be perfectly pre-cancelled at the transmitter using ZFBF with full channel state information at the transmitter (CSIT). For intra-cluster interference, the strong user subtracts the weak user signal from the superposed signal by applying SIC detection technique, and then decodes its own message. Weak user does not perform interference cancellation and directly decodes its own message with interference from strong user. Therefore, two models of the received signal for strong and weak users are presented in the following subsections.

A. Received Signal and Sum Rate: Strong User

Before performing SIC at the strong user side, the received superposed signal at the strong user with noise and interferences in the \( k^{th} \) cluster is given by:

\[
y_{k1} = \sqrt{P} \alpha_{k1} h_{k1} w_k s_{k1} + \sqrt{P} \alpha_{k2} h_{k1} w_k s_{k2} + \sqrt{P} \sum_{j=1}^{N} h_{k1} w_j x_j + n_{k1}.
\]

where \( k_1 \in S_{zf} \), \( h_{k1} \) and \( n_{k1} \) are the \( 1 \times N \) channel vector and the received noise for the strong user in the \( k^{th} \) cluster. In NOMA-ZFBF with SIC, the ZFBF is used to achieve a higher sum-rate and to remove inter-cluster interference. In addition, the ZFBF weight \( W \) is designed based on the channels of the strong users in order to perform the SIC process successfully. The ZFBF matrix \( W \) is given by [4], [10]:

\[
W = [w_1, \ldots, w_N] = H^* (H H^*)^{-1},
\]
where \( H = [h_{11}, \ldots, h_{1K}]^T \) denotes the channel matrix of strong users, \( H^* \) is the conjugate transpose of \( H \) and \((\cdot)^T\) denotes the transpose operation. The preceding vector \( w_k \) for each cluster is generated at the transmitter by normalizing the \( k^{th} \) column of \( W \) so that:

\[
h_{k1}w_jx_j = 0 \text{ for } j \neq k. \tag{4}
\]

Therefore, inter-cluster interference is eliminated from the received signal using ZFBF and, \( \sum_{j=1,j \neq k}^N h_{k1}w_jx_j = 0. \)

After performing SIC, intra-cluster interference is also removed from the received signal, and therefore, it reduced to:

\[
y_{k1} = \sqrt{P \alpha_{k1}} h_{k1} w_k s_{k1} + n_{k1}. \tag{5}
\]

And accordingly, the achievable data rate for strong user \( R_{k1} \) in the \( k^{th} \) cluster is:

\[
R_{k1} = \log \left( 1 + \frac{\alpha_{k1} P |h_{k1} w_k|^2}{\sigma_n^2} \right). \tag{6}
\]

### B. Received Signal and Sum Rate: Weak User

After selecting the strong users, another user with lower channel quality must be added for each cluster to perform NOMA-ZFBF scenario. The BS allocates low power to the strong users and high power to the weak users. Consequentially, weak user does not need to perform SIC and its message is directly decoded. Therefore, the received signal by the weak user, \( y_{k2} \), has both inter-cluster and intra-cluster interferences and is given by:

\[
y_{k2} = \sqrt{P \alpha_{k2}} h_{k2} w_k s_{k2} + \sqrt{P \alpha_{k1}} h_{k2} w_k s_{k1} + \sqrt{P} \sum_{j=1,j \neq k}^N h_{k2} w_j x_j + n_{k2}. \tag{7}
\]

Consequently, the data rate of weak user \( R_{k2} \) in the \( k^{th} \) cluster is:

\[
R_{k2} = \log \left( 1 + \frac{\alpha_{k2} P |h_{k2} w_k|^2}{\alpha_{k1} P |h_{k2} w_k|^2 + P \sum_{j=1,j \neq k}^N |h_{k2} w_j|^2 + \sigma_n^2} \right). \tag{8}
\]

### IV. MULTI-USER SELECTION USING PF SCHEDULING

Most of the users scheduling algorithms have been designed to maximize throughput, but unfortunately they ignore fairness which is also an important criterion. This will reduce the probability of selecting users with poor channels quality; which will adversely impact on their respective quality of service (QoS). Scheduling with proportional fairness (PF) can overcome this problem and can allow a fine trade-off between throughput and fairness; this is achieved by considering the users instantaneous channel qualities and their average throughputs, respectively. However, for multi-user transmissions, more than one user are selected and served simultaneously. Therefore, an extended version of PF selection metric for multiuser scheduling described in [4], [13] is used here and is given by:

\[
S = \arg \max_{S \in \mathcal{U}} \sum_{i \in S} \mu_i(t), \tag{9}
\]

where \( S \) is a set of scheduled users at time slot \( \mu_i(t) \) and \( \mu_i(t) \) are the average throughput, and the actual data rate for user \( i \) at time slot \( t \), respectively, and \( \mu_i(t) \) is updated at every time slot \( t \) as the following:

\[
\mu_i(t+1) = \left\{ \begin{array}{ll}
1 - \frac{1}{t_c} & \mu_i(t) + \frac{1}{t_c} R_i(S,t), \quad i \in S, \\
1 - \frac{1}{t_c} & \mu_i(t), \quad i \notin S,
\end{array} \right. \tag{10}
\]

where \( t_c \) is the fairness parameter of PF, and its value highly determines the required time scale to satisfy a high level of fairness, i.e. with small values of \( t_c \), a high level of fairness can be satisfied in short time intervals and vice versa. Theoretically, a large \( t_c \), makes the user average throughput decrement term \( (1 - \frac{1}{t_c}) \mu_i(t) \) very slow, thus, the algorithm schedules a user with the strongest channel quality relative to its own average throughput. This leads to maximizing system throughput and decreasing fairness level among users. On the other hand, for small \( t_c \) values the user average throughput will rapidly changed, making the scheduling decision more dependent on user average throughput \( \mu_i(t) \), which improves fairness. Therefore, there is a trade-off between system throughput and fairness and \( t_c \) value must be appropriately chosen in order to improve this trade-off.

### A. Fairness Index

To measure how well a specified level of fairness is guaranteed among users over a short time scales, Jain’s fairness index (JFI) is used [15], which is a well-known quantitative metric that is widely adopted in wireless communications and it is defined as [16]:

\[
J = \frac{\left( \sum_{i=1}^K T_i \right)^2}{K \sum_{i=1}^K T_i}, \tag{11}
\]

Where \( K \) is the number of users, \( T_i \) describes the average throughput allocated to user \( i \) calculated over a finite time window of length \( W_i \). JFI values are bounded in the range \( 1/K \) (worst fairness) to 1 (ideal fairness). The unity (ideal) fair gives an indication of equal distribution of system resources between \( K \) users. In contrast, \( J = 1/K \) denotes that all the resources are given to one user only.

### V. PROPOSED ALGORITHM

The authors in [14], proposed semiorthogonal user selection (SUS) algorithm which proved that it can achieve the asymptotic optimal sum-rate growth under full and partial CSIT. In order to improve the fairness performance of NOMA-ZFBF, we modify and extend the SUS algorithm in [14] to serve multiple users in NOMA clusters based on proportional fairness PF-NOMA-ZFBF.
The proposed algorithm is a two-stage user selection algorithm as described in Algorithm 1. In the first stage, strong users are selected. In order to maximize throughput, the first user in this stage is selected based on the best user channel quality $||h_t||^2$, and the remaining users are selected based on the maximum semi-orthogonality $||g_i||^2$ [11]. Strong users selection criterion is also relied on $\mu_i(t)$, to ensure fairness between users as in PF selection criterion given in (9).

Weak users are selected in the second stage from the remaining $U - S_z$ users, using the designed beams from strong user channels. We extend PF selection criterion given in (9) to be relied on user gain matrix $|h_iw_k|^2$ to inter-cluster interference $\sum_{j=1, j\neq k}^N|h_iw_j|^2$ ratio, in order to find the best matched user for each cluster.

VI. TRANSMIT POWER ALLOCATION

The best NOMA sum-rate performance can be achieved by full search power allocation (FSPA) and exhaustive search method for user selection [3]. However, FSPA is considered computationally complex. The downlink NOMA power control basically aims to create enough difference in power for the superimposed signals of users in each cluster to enable signal separation and detection using SIC. Because of high channel gain difference between the strong and weak users, a suboptimal power allocation scheme in [17] can be used effectively. Thus, for $k^{th}$ cluster,

$$p_{k1} = \alpha_{k1}P,$$

$$p_{k2} = \alpha_{k2}P,$$

where $p_{k1}, p_{k2}$ are the power allocated to the strong and weak user, respectively and $P$ is the total cluster power. Because the strong user has a better channel gain compared to the weak user, a higher transmitted power should be allocated to the weak user, therefore, $\alpha_{k2} > \alpha_{k1}$ and the power allocation coefficients for the strong and weak users are given by:

$$\alpha_{k1} = \frac{\beta}{\beta + 1},$$

$$\alpha_{k2} = 1 - \alpha_{k1},$$

where $\beta$, is the power allocation fraction factor, $(0 < \beta < 1)$, and the system can be optimized with careful selection for this parameter.

VII. SIMULATION RESULTS

In this section, simulation results are presented to show the performance of the proposed user clustering algorithm. The basic simulation setting for the numerical results in this section are as follows: The total cluster power $P = 15$ dB, power allocation parameter $\beta$ is set to 0.15 compliant with [16], noise variance is one for all users, i.e., $\sigma^2 = 1$.

Different numbers of total candidate users $U$ are used in the simulation from 10 to 170. However, for any system, the fairness performance decreases with increasing number of users. Therefore, the simulation tests focused on a relatively high number of users ($U = 150$), to measure the achieved throughput-fairness trade-off enhancement of the proposed system over the conventional NOMA-ZFBF [10].

The simulation results in Fig. 2 shows sum-rate performance of strong and weak users vs. $t_c$, which is used to select the best $t_c$ value that can allow the total system sum-rate to reach the

Algorithm 1 Proposed PF-NOMA-ZFBF Algorithm

**Stage 1: Strong Users Selection:**

1: At time $t$, initialize user set $U = \{1, \ldots, K\}$, $S_z = \emptyset$ and $k = 1$, and select the first user in user group $S_z$ from the initial user set $U$ by using this criterion:

$$S_z(1) = \arg \max_{i \in U} \left( \frac{\log(1 + P ||h_i||^2)}{\mu_i(t)} \right),$$

$$S_z \leftarrow S_z \cup \{S_z(1)\} \quad \quad k \leftarrow k + 1$$

2: After selecting the first user, the $k^{th}$ user is selected based on subspace semi-orthogonality with respect to the first user, which is determined by semiorthogonal component $g_i$ [11]. This $k^{th}$ user is selected from the remaining users set $C = U - S_z$ as:

$$S_z(k) = \arg \max_{i \in C} \left( \frac{\log(1 + P ||g_i||^2)}{\mu_i(t)} \right),$$

$$S_z \leftarrow S_z \cup \{S_z(k)\} \quad \quad k \leftarrow k + 1$$

3: If $k \leq N$, then go to 2. Otherwise, strong users selection is completed.

4: Generate ZFBF matrix $W = \{w_1, w_2, \ldots, w_N\}$ by using (3), which will be used in the selection of weak users.

**Stage 2: Weak Users Selection:**

5: Initialize the weak users set as $S_w = \{\emptyset\}$ and $k = 1$.

6: Search within the remaining users set $C = U - S_z - S_w$ for selecting weak user for the $k^{th}$ cluster as:

$$S_w(k) = \arg \max_{i \in C} \left( \frac{\log(1 + \sum_{j=1, j\neq k}^N |h_iw_j|^2)}{\mu_i(t)} \right),$$

$$S_w \leftarrow S_w \cup \{S_w(k)\} \quad \quad k \leftarrow k + 1$$

7: If $k \leq N$, go back to 6. Else, weak users selection is completed.

8: Compute the actual supported data rates for strong and weak users using (6) and (8), respectively.

9: Update the average throughputs for all the users using $\mu_i(t+1)$s given in (10).

10: Go to the strong user selection stage to make new clusters for the next time slot $(t+1)$. 

throughput-fairness trade-off enhancement of the proposed system over the conventional NOMA-ZFBF [10].
maximum capacity as much as possible (maximum capacity is specified by NOMA-ZFBF). It can be seen that the total system throughput of the proposed scheme increases with increasing $t_c$ until it approximately reaches NOMA-ZFBF performance when $t_c = 200$. Therefore, this value is selected as the best $t_c$ that can maximize the proposed system sum-rate. It is worth to mention here that the maximum system sum-rate can be reached with a larger $t_c$, but this will lead to a noticeable reduction in fairness, which is another important performance criterion that we want to consider.

On the other hand, the sum-rate performance of weak users can reach to the maximum at $t_c = 30$ only. This performance is due to the small channel gain values of weak users, which in turn make maximization in throughput became faster with a small $t_c$ value compared to the whole system. However, the $t_c$ value should be selected according to the total system capacity.

Fig. 3 shows that fairness is a time dependent and it increases with increasing time window $W_t$. From this figure a minimum time window $W_t$ will be selected according to a specified short term fairness. We specify the minimum fairness index (fairness threshold) of 0.85. It is observed that when $W_t = 225$, the minimum achieved fairness index for PF-NOMA-ZFBF is 0.8618 (exceeds fairness threshold) when $t_c = 200$, whereas, NOMA-ZFBF fairness index is only 0.62. Therefore, $W_t = 225$ will be used for subsequent simulation tests, as a minimum time window that satisfies a specified short-term fairness.

Fig. 4 and Fig. 5 show the total system sum-rate and fairness index, respectively with respect to different number of users. In Fig. 4, it can be seen that the system throughput of the proposed scheme increases with increasing $t_c$ until it approximately reaches NOMA-ZFBF performance when $t_c = 200$. On the other hand, Fig. 5, shows that that fairness index decreased with increasing number of candidate users for all $t_c$ values. The decrease is less for small $t_c$ than with the higher $t_c$. Nevertheless, in all cases, our proposed PF-NOMA-ZFBF algorithm significantly outperforms NOMA-ZFBF.
Table I: Throughput-fairness trade-off enhancements for $U = 150$ users and $W_2 = 225$.

<table>
<thead>
<tr>
<th>$t_c$</th>
<th>Throughput Decrease (bps/Hz)</th>
<th>Fairness Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-1.0036 (6.92%)</td>
<td>+32.97 (53.16%)</td>
</tr>
<tr>
<td>100</td>
<td>-0.3865 (2.67%)</td>
<td>+29.88 (48.18%)</td>
</tr>
<tr>
<td>200</td>
<td>-0.1287 (0.89%)</td>
<td>+24.16 (38.96%)</td>
</tr>
</tbody>
</table>

Fig. 6 shows weak users sum-rate which also increases with increasing number of users, $U$. However, it is noticed that for different number of users, the sum-rate of the proposed system can reach the maximum capacity with $t_c = 30$, which is considered a small value compared to the total system capacity that approximately reaches the maximum performance with $t_c = 200$.

Fig. 6. Sum-rate performance of weak users versus the number of users

Fig. 7. Performance of weak users with NOMA-ZFBF and PF-NOMA-ZFBF.

VIII. CONCLUSION

In this paper, a two-stage user clustering algorithm has been proposed based on proportional fairness for downlink NOMA-ZFBF system PF-NOMA-ZFBF in order to improve the overall system fairness and maintaining the highest possible throughput at short time intervals. To satisfy this, a suboptimal selection for $t_c$ is introduced in which the total system capacity can reach the maximum possible value with respect to its maximum performance represented by NOMA-ZFBF. Simulation results showed a superior system short-term fairness with relatively high data-rates levels, i.e., highly improvement for system throughput-fairness trade-off, which verified the advantages of the proposed user selection method over the conventional NOMA-ZFBF.

IX. ACKNOWLEDGMENT

The authors would like to acknowledge the research grant ATOM Advancing the State of the Art of MIMO: The Key to Successful Evolution of Wireless Networks (Project No.: 690750-ATOM-H2020-MSCA-RISE-2015, UPM: 6388800-10801) for funding the project.

REFERENCES