User-centric Pre-selection and Scheduling for Coordinated Multipoint Systems

Annika Klockar*, Mikael Sternad†, Anna Brunstrom*, Rikke Apelfröjd†
*Computer Science, Karlstad University, Karlstad, Sweden, {annika.klockar, anna.brunstrom}@kau.se
†Signals and Systems, Uppsala University, Uppsala, Sweden, {mikael.sternad, rikke.apelfrojd}@signal.uu.se

***Abstract*—The data traffic volumes are constantly increasing in cellular networks. Furthermore, a larger part of the traffic is generated by applications that require high data rates. Techniques including Coordinated Multipoint transmission (CoMP) can increase the data rates, but at the cost of a high overhead. The overhead can be reduced if only a subset of the users is served with CoMP. In this paper, we propose a user selection approach, including pre-selection of CoMP users and short term scheduling, that takes user requirements into account. Users that require a high data rate to reach an acceptable level of service satisfaction are selected to use coherent joint processing CoMP in some of their downlink transmission bandwidth. Simulation results show that both the number of satisfied users and fairness are improved with the proposed user selection as compared to user selection that does not consider individual user requirements.**

I. INTRODUCTION

The data traffic in cellular networks has increased enormously over the last years, and is expected to continue to increase. A large part of the traffic volume in cellular networks is generated by applications that require high data rates, such as video. Inter-cell interference is the main limiting factor for high data rates in cellular networks. To reduce inter-cell interference Coordinated Multipoint transmission/reception (CoMP) techniques have been proposed [1], [2], [3].

In this paper, we propose and evaluate a user selection approach for linear coherent joint processing CoMP in the downlink. Coherent joint transmission is also denoted Network MIMO. It uses antennas at different cells to form a joint downlink beamformer. This technique reduces the interference within a cooperating cluster of cells by phase cancellation. Coherent joint processing CoMP requires accurate channel state information (CSI) which in turn requires the users to frequently predict their channels and send feedback to the base stations (BSs). User data need to be distributed between the BSs, since the BSs transmit user data simultaneously to the users. In order to reduce the feedback overhead in the uplink and also the load on backhaul links due to the distribution of user data between the BSs, we select a subset of the users for CoMP.

In contrast to most studies of user selection for CoMP, we take user experience into account. Most previous studies focus mainly on maximizing sum rate per cell, e.g., [4], [5]. Users typically have different requirements, a fact that we use to select users with high requirements for CoMP, and to exclude others, for which transmission from a single BS is sufficient to meet their requirements. Some users run hard real-time applications, such as VoIP, which require a low constant data rate and almost no delay or jitter, while others are running less sensitive elastic applications, such as e-mail and file transfer, which accept relatively large variations in data rate and delay. Yet others run delay or rate adaptive applications, which have requirements in between hard real-time and elastic applications, such as streaming video and conferencing applications. The level of user satisfaction of the application types can be represented by utility functions [6].

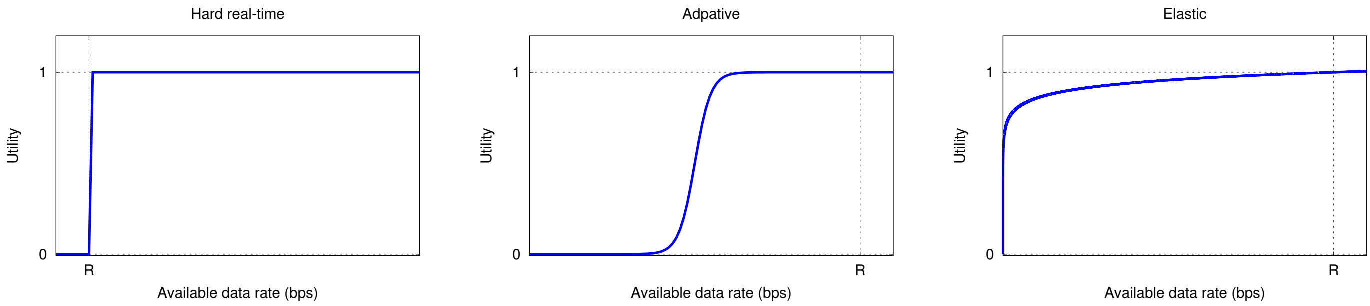
The user selection approach proposed in this paper selects users for CoMP based on application utility. User selection is performed in two steps: First, by a pre-selection process, and, then, by a joint time and frequency domain scheduler. The algorithm proposed for pre-selection of users is a refined version of the algorithm that we proposed in [7]. The pre-selection of users for CoMP is intended as a radio resource management (RRM) process that operates on relatively long time scales, in the order of hundreds of milliseconds. Application utilities of the users are evaluated, and the users that need CoMP most are selected for CoMP transmission. Application utility is naturally coupled to the longer time scale on the pre-selection level, since it is a long term measure of user experience. For many Internet applications, short term measures, such as instantaneous data rate and delay in the range of a few milliseconds are not so relevant, except for some class of hard real-time applications with very strict delay requirements. The joint time and frequency domain scheduler operates on data rate and channel quality on the millisecond time scale. Due to the short time scale and in order to keep the complexity low, application utility is not used as a scheduling metric. The time domain scheduler instead uses a target bit rate that is based on the application utilities calculated by the pre-selection process. Simulation results of the utility-based user selection show that more users reach acceptable utility levels and that fairness is improved, since resources are redistributed from the users with very high utility values to users that require more resources.

The remainder of the paper is organized as follows. In Section II, the system model is outlined. The proposed pre-selection and scheduling are explained in Section III. The simulation setup used to evaluate the scheme and the simulation results are presented in Section IV. Conclusions are provided in Section V.

II. SYSTEM MODEL

The downlink of a CoMP enabled cluster of K BSs with M users is considered, as illustrated in Fig. 1. The BSs have one transmit antenna, and the users have a single receive

978-1-4799-5863-4/14/$31.00



(a) (b) (c)

Fig. 2. Utility of application types: (a) hard real-time (b) adaptive, and (c) elastic

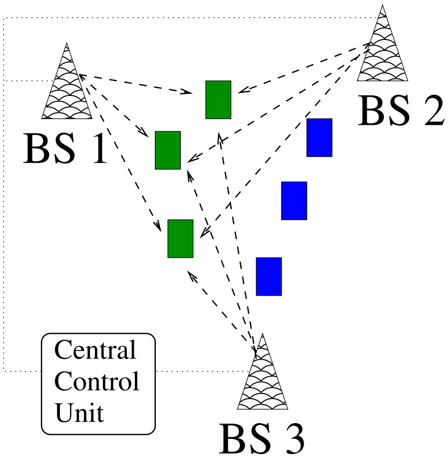


Fig. 1. CoMP cluster

antenna. Joint processing and pre-selection of users for CoMP are performed by a central control unit, which could be a separate network node or located in one of the BSs.

The users are assumed to run one Internet application each. In this paper, hard real-time, adaptive, and elastic traffic is considered. This could correspond to traffic that in LTE would use a default bearer with no guaranteed bit rate (non-GBR) [8]. The applications considered belong to one of the following application types:

- Hard real-time applications, e.g. VoIP
- Adaptive applications, e.g., streaming media
- Elastic applications, e.g., e-mail or file transfer

The user satisfaction with a service is quantified by utility functions. In this paper, we use utility functions of data rate, which are similar to the ones presented in [6], which we used also in [7]. The typical shapes of the utility functions of the application types considered are shown in Fig. 2. The y-axis indicates utility values between 0 and 1. On the x-axis, the available data rate is indicated. The data rate marked with R is the data rate for which the user is assumed to be satisfied.

For hard real-time applications, the following utility function, illustrated in Fig. 2a, is used

$$u(r) = \frac{sgn(r-R)+1}{2}, \tag{1}$$

where r is the available data rate and R is the data rate required for the maximum utility $u(R) = 1$.

The utility function, shown in Fig. 2b, used for adaptive applications is

$$u(r) = \frac{1}{1+(1/\epsilon-1)^{(1-2r/R)}}, \tag{2}$$

where R is the data rate required for the maximum utility $u(R) = 1-\epsilon$. In this paper, we evaluate an adaptive utility function with $\epsilon = 10^{-9}$, which gives a function that clearly differs from the ones used for hard real-time and elastic applications.

The following utility function, shown in Fig. 2c, is used for elastic applications

$$u(r) = \frac{\ln(r+1)}{\ln(R+1)}, \tag{3}$$

where R is the data rate required for $u(R) = 1$. Elastic applications can, in contrast to adaptive applications, use an even higher data rate, but with only marginally increased utility.

All users admitted to the system need to be served. Two transmission modes are offered:

- Coherent joint processing CoMP
- Conventional single BS transmission

Coherent joint processing CoMP requires detailed feedback information, and, since we want to keep the amount of feedback information low, only a limited number of the users are selected for CoMP. By limiting the number of users for CoMP, the load on backhaul links is reduced, which would also be an advantage if the capacity of the backhaul links is limited. Conventional single BS transmission, on the other hand, is not limited due to overhead, but can be used to serve all admitted users. All users are served with single BS transmission, and selected users are also served with CoMP. Hence, CoMP is used as an add-on service.

Frequency selective downlink channels are considered as in an OFDMA system. The bandwidth is divided into F resource blocks (RBs). In each RB, one user per cell is scheduled. The resources are divided into two groups in the time domain, with all the RBs of the time slots in the one group reserved for CoMP, rbs_c, and all the RBs of the time slots in the other group reserved for single BS transmission, rbs_{s}. Hence, the CoMP service is available at predefined time intervals, as, for example, in [9]. In Fig. 1, an example of a CoMP transmission is shown. The users that are selected for CoMP and scheduled in the current CoMP time slot are, first, precoded for CoMP transmission by the central control unit, and, then, served by all the BSs in the cluster.

Single BS transmission is performed for the users scheduled in the RBs in rbs_{s}. There is inter-cell interference resulting from other users scheduled in the same RB in the other cells. Intra-cell interference is, however, avoided by scheduling the users in separate RBs. For the users served with CoMP in the RBs in rbs_{c}, on the other hand, inter-cell interference is mitigated with linear coherent joint processing CoMP using zero-forcing precoding [10].

The total data rate of a user m is

$$r_m = \sum_{f=1}^{F} r_{f,m}, \tag{4}$$

where $r_{f,m}$ is the data rate of the user in RB f, which is

$$r_{f,m} = \beta_1 \mathrm{B}_f \log_2(1 + \beta_2 \mathrm{SINR}_{f,m}), \tag{5}$$

where B_f is the bandwidth of a RB, and the $\mathrm{SINR}_{f,m}$ is the signal to interference plus noise ratio for the mth user in resource block f. The constants β_1 and β_2 account for overhead due to protocol headers, training sequences, transmission mode, coding and modulation, etc [11], [12]. In order to reduce the risk of overestimating the system performance, a relatively high overhead is assumed in this paper, $\beta_1 = 0.57$ and $\beta_2 = 0.30$, corresponding to $-5.23\,\mathrm{dB}$.

The $\mathrm{SINR}_{f,m}$ for a user served with CoMP is calculated as follows:

$$\mathrm{SINR}_{f,m} = \frac{|\boldsymbol{h}_{f,m}^T \boldsymbol{w}_{f,m}|^2 p_{f,m}}{\sum\limits_{n \in \mathcal{M}_{rb}, n \neq m} |\boldsymbol{h}_{f,m}^T \boldsymbol{w}_{f,n}|^2 p_{f,n} + \sigma^2}, \tag{6}$$

where, $\boldsymbol{h}_m^T \in \mathbb{C}^{1 \times K}$ represents the channels between the mth user and the K BSs and $\boldsymbol{w}_m \in \mathbb{C}^{K \times 1}$ are the beamforming weights related to the BSs in the cluster. The interference remaining after precoding comes from the transmissions to the other users in $\mathcal{M}_{rb}$, which is the set of users served in the same RB. The remaining interference is zero for ideal zero-forcing. The power spectral density allocated to the user across the K BSs in each RB is $p_{f,m}$, and σ^2 is the noise spectral density in which inter cluster interference is included.

The transmission power spectral density at each BS is assumed to be limited by a maximum value, P_{max}, and equal user power allocation, as in [10], is applied within each RB. This implies that $p_{f,m} = p_{f,n}$, for all users m, n. To fulfill this constraint, the power spectral density allocation matrix $\mathbf{P} = \mathrm{diag}\{p_j\}$ is scaled as

$$\sqrt{\mathbf{P}} = \left\{ \min_{k=1,\cdots,K} \sqrt{\frac{P_{max}}{||\mathbf{W}^{(k)}||_F^2}} \right\} \cdot \mathbf{I}_{[M \times M]}, \tag{7}$$

where $\mathbf{W}^{(k)}$ is the beamforming weights of the kth BS, $(\cdot)_F$ the Frobenius norm, and $\mathbf{I}_{[M \times M]}$ an identity matrix of size $[M \times M]$. More advanced power allocation could have an impact on the result, but this is not considered in this paper.

Algorithm 1 User selection for CoMP

for Each CoMP epoch **do**
  Pre-select M_{c} users for CoMP
  for Each time slot **do**
    if Time slot reserved for CoMP transmission **then**
      Schedule $M_{\mathrm{td}} < M_{\mathrm{c}}$ selected users
    end if
    if Time slot reserved for Single BS transmission **then**
      Schedule M_{td} users of all users
    end if
    for Each RB in the time slot **do**
      Schedule one user per cell
    end for
  end for
end for

For the users served with single BS transmission, no precoding is performed to eliminate the inter-cell interference between the users served in the same RB. The $\mathrm{SINR}_{f,m}$ of a user served with single BS transmission is calculated as follows:

$$\mathrm{SINR}_{f,m} = \frac{|h_{f,m}^{(k)}|^2 p_{f,m}}{\sum\limits_{l \neq k}^{K} |h_{f,m}^{(l)}|^2 p_{f,n} + \sigma^2}, \tag{8}$$

where $h_{f,m}^{(k)} \in \mathbb{C}$ represents the channel between the mth user and its serving BS, k, and $h_{f,m}^{(l)} \in \mathbb{C}$ the channels between the other BSs, l, in the cluster and a user m. One user is served in each cell, and each BS transmits with full power to the scheduled user.

III. User selection for CoMP

In this section, we propose a heuristic approach to select users for CoMP. The proposed user selection for CoMP is outlined in Algorithm 1. On a longer time scale, typically hundreds of milliseconds, denoted CoMP epoch, a subset of the users, M_{c}, are selected as a long term pre-selection of users to be served with CoMP. Scheduling on a shorter time scale then allocates times slots and RBs to the users. CoMP is used as an add-on service. This implies that all users are served with single BS transmission, and that the M_{c} users pre-selected for CoMP are served both with CoMP and with single BS transmission. Single BS transmission, therefore, provides a baseline service. The main advantage of the add-on CoMP service, as compared to letting some users use CoMP exclusively, is that the data rates of the individual users vary less due to which users that are selected for CoMP and which are not selected.

Short term scheduling is performed first in the time domain (TD) and then in the frequency domain (FD). With a joint time and frequency domain scheduler, the scheduler in the TD selects the users to serve for the duration of the next time slot and the FD scheduler determines which user to serve in which RB. The advantages are that the burden on the FD scheduler is reduced, and that different scheduling policies can be applied in TD and FD [13]. In the typical case and in this paper, the TD scheduler considers user requirements, while the FD scheduler mainly focuses on spectral efficiency. In the rest of this section, the heuristic user selection in Algorithm 1 is described in more detail.

A. Pre-selection

The main purpose of the pre-selection of users on a longer time scale, in the range 100ms - 1s, is to reduce the feedback overhead due to CoMP, as more detailed channel information is required for CoMP than for single BS transmission. On this level, long term objectives and performance measures provide a feasible basis for the user selection, such as application utility, average data rate, or user priority.

We propose a pre-selection algorithm, utility-based user selection, that takes user requirements into account. The pre-selection algorithm is an extended version of an algorithm that we proposed in [7]. The algorithm selects users based on the utility of the data rate that is expected to be achieved if a user is served also with CoMP. The idea is to select the users for which the application utilities would increase the most in comparison to only single BS transmission. For this, an estimation of the utility gain over the next CoMP epoch is required.

The objective of the algorithm is to

$$\begin{aligned} \text{maximize} \quad & \sum_{m=1}^{M_{\mathrm{c}}} \Delta u_m \\ \text{subject to} \quad & \Delta u_m = u_m(r_{m,\mathrm{c}} + r_{m,\mathrm{s}}) - u_m(r_{m,\mathrm{s}}) \\ & u_m(r_{m,\mathrm{c}} + r_{m,\mathrm{s}}) \le u_{m,\mathrm{target}} \\ & u_m(r_{m,\mathrm{s}}) \le u_{m,\mathrm{target}} \\ & u_{m,\mathrm{target}} \le u_m(R), \end{aligned} \tag{9}$$

where M_{c} is the number of users that can be served with CoMP, Δu_m the expected utility gain if user m is selected, $u_{m,\mathrm{c}}(r_{m,\mathrm{c}} + r_{m,\mathrm{s}})$ and $u_{m,\mathrm{s}}(r_{m,\mathrm{s}})$ the utilities of rates for add-on CoMP and single BS transmission, respectively. The utilities are calculated as in Eq. (1), Eq. (2) or in Eq. (3) depending on the application type. Resources are assumed to be scarce, since otherwise user selection would not be required. The maximum utility may not be reached for most users. Therefore, to impose some level of fairness the calculated utilities are limited by a $u_{m,\mathrm{target}}$, the corresponding rate of which is used by the TD scheduler, as described in Sec. III-B. In the case of adaptive or elastic utility functions $0.7\ u_m(R) \le u_{m,\mathrm{target}} \le 0.9\ u_m(R)$ could provide an acceptable user experience. For hard real-time applications, on the other hand, the maximum utility is required, $u_{m,\mathrm{target}} = u_m(R)$, since the utility function is a discrete function, which only evaluates to 0 or 1.

In order to determine which users to select for the next CoMP epoch, the data rates for the next CoMP epoch need to be estimated. The resulting data rates will depend on user requirements, in this case on $u_{m,\mathrm{target}}$, on which M_c users that are selected for CoMP, and on which users that are scheduled in the same RB. As we want to avoid to check all possible user groups for all time slots and RBs, the following approximation is used for CoMP:

$$\mathrm{SINR}_{f,m} \approx \frac{|\boldsymbol{h}_{f,m}^T \boldsymbol{w}_{f,m}|^2 p_{f,m}}{\sigma^2}, \tag{10}$$

where m is assumed to be the only user served with CoMP. It is, however, not likely to be the case that m will be the only CoMP user in a RB, but the user groups for the RBs in the next CoMP epoch are not known and hard to estimate. Single BS transmission is approximated by:

$$\mathrm{SINR}_{f,m} \approx \frac{|h_{f,m}^{(k)}|^2 p_{f,m}}{\sum\limits_{l \neq k}^{K} |h_{f,m}^{(l)}|^2 p_{f,n} + \sigma^2}. \tag{11}$$

For the approximation of $\mathrm{SINR}_{f,m}$ for CoMP and single BS transmission, assumptions of the average $|h_{f,m}|$ values are required, which in turn depends on the channel model. With the channel model used in this paper, described in Sec. IV, pathloss and shadow fading are assumed to be known by the central control unit. Furthermore, pathloss and shadow fading are assumed to be fixed for the duration of a CoMP epoch. The short term fading is, however, not known at the time when users are pre-selected. The user that is selected for transmission in a RB is assumed to have a relatively high channel quality, which implies that the average short term fading gain for a selected user is higher than the average of the Rayleigh distribution. This problem is similar to the problem of average SNR for selection combining, as described in [14]. The average short term fading gain for user m when scheduled is $\sum_{m=1}^{M_{\mathrm{k}}} 1/m$, where M_{k} is the number of users having the same serving BS k. For the short term fading gains of the channels in the interference term, the average value of the Rayleigh distribution, $\sqrt{(\pi/2)}$, is used instead.

The data rates, $r_{m,\mathrm{c}}$ and $r_{m,\mathrm{s}}$ are calculated as in Eq. (4) and in Eq. (5). The number of RB:s and time slots in which a user will be served depends on the scheduling policies of the short term schedulers and the channel conditions. For simplicity, it is assumed that the users selected by the pre-selection process will be served equally often during the next CoMP epoch.

B. Time domain scheduling

For the proposed user selection approach, it is important that the TD scheduling policy is consistent with the assumptions used in the pre-selection policy. Application utility could be used as TD scheduling metric, which has been proposed in many papers, e.g., [6]. We have, however, chosen to use application utility only in the pre-selection algorithm, in order to keep the complexity of the TD scheduler low. We evaluate a target bit rate (TBR) scheduler similar to the TD scheduling algorithm proposed in [15]. Instead of applying the same TBR to all users, our TD scheduler uses TBR values corresponding to $u_{m,target}$ for the individual users. TBR values corresponding to $u_{m,target}$ act in place of the utility functions in the TD scheduler. All users are considered for being scheduled in time slots reserved for single BS transmission. The scheduling of time slots reserved for single BS transmission is performed without taking CoMP results into account. As an add-on service model is applied, CoMP users, on the other hand, are scheduled based on results both from single BS transmissions and from CoMP transmissions. The users with the lowest scheduling values are selected for the next time slot.

$$V_m = r_{m,\mathrm{avg}} / \mathrm{TBR}_m, \tag{12}$$

TABLE I. TRAFFIC MIXES

Traffic type	R value (kbps)	Mix 1	Mix 2	Mix 3	Mix 4
Hard real-time	87.2	0%	0%	0%	25%
Adaptive	2000	75%	50%	25%	50%
Elastic	2000	25%	50%	75%	25%

where $r_{m,\,\mathrm{avg}}$ is the exponential moving average of past rates in which a time slot history window, corresponding to the delay requirements of the applications, is included, and TBR_m is the rate that is required to reach the $u_{m,\,\mathrm{target}}$. For simplicity, in order to avoid estimating the rate for the next time slot, the $r_{m,\,\mathrm{avg}}$ is one time slot behind. This is not expected to influence the results significantly, since in this paper block fading is used, and, hence there is no dependency of the fading between time slots. Furthermore, the expected data rates in the next time slot depend on the user group, which is unknown when the scheduling decision is taken.

C. Frequency domain scheduling

For the short term scheduling in the FD we apply the cellular user grouping proposed in [16]. In this cellular user grouping, scheduling decisions are taken only with respect to single cell transmission, as in conventional systems. Then, a central control unit performs joint processing for the users scheduled in the same RB in FD. One advantage of cellular user grouping is that CoMP can be introduced without the need to replace the efficient scheduling algorithms that are used in conventional systems. The cellular user grouping evaluated in [16] uses the score-based scheduler (SB) presented in [17] in the FD for user grouping for CoMP. This user grouping results in more efficient user groups for CoMP than random user grouping, mainly since the probability that the users have different strongest BSs is higher than with random user grouping. The cellular user grouping with SB applied in [16] performed very close to greedy central scheduling, at a lower complexity and without imposing extra backhaul requirements. Therefore, it is considered a suitable FD scheduler for this study.

IV. SIMULATION SETUP AND RESULTS

The simulations are performed for a cluster of $K = 3$ single antenna BSs. The inter-BS distance is 500 meters. There are $M = 80$ single antenna users in the system. The users are dropped in a circular area with a radius of 360 meters from the cluster center. An OFDMA system with a bandwidth of $20\,\mathrm{MHz}$ is considered. The bandwidth is divided into 100 resource blocks, each consisting of 12 subcarriers of $15\,\mathrm{kHz}$. Data scheduled in the resource blocks are transmitted in time slots which have a duration of $1\,\mathrm{ms}$. The channel coefficients between the mth user and the kth BS are modeled as $\boldsymbol{h}_{mk} = \boldsymbol{h'}_{mk}\sqrt{\gamma_s\gamma_p}$, where the shadow fading is a random variable described by a log-normal distribution, $\gamma_s \sim \mathcal{N}(0, 8\,\mathrm{dB})$, the pathloss follows the 3GPP Long Term Evolution (LTE) model, $\gamma_p[\,\mathrm{dB}] = 148.1 + 37.6\log_{10}(d_{mk})$, where d_{mk} is the distance between the kth BS and the mth user, and $\boldsymbol{h'}_{mk}$ includes the small scale fading coefficients, which are i.i.d. complex Gaussian values according to $\mathcal{CN}(0, 1)$. The noise floor is $-125.22\,\mathrm{dBm}$ including a receiver noise figure of $7\,\mathrm{dB}$. The maximum transmission power per BS is $30\,\mathrm{W}$ equally divided over the RBs. For the simulations presented in this paper, the user locations and the shadow fading values are drawn once and held fixed for the duration of 500 time slots. The short term fading values, on the other hand, are randomized for the channels in each RB.

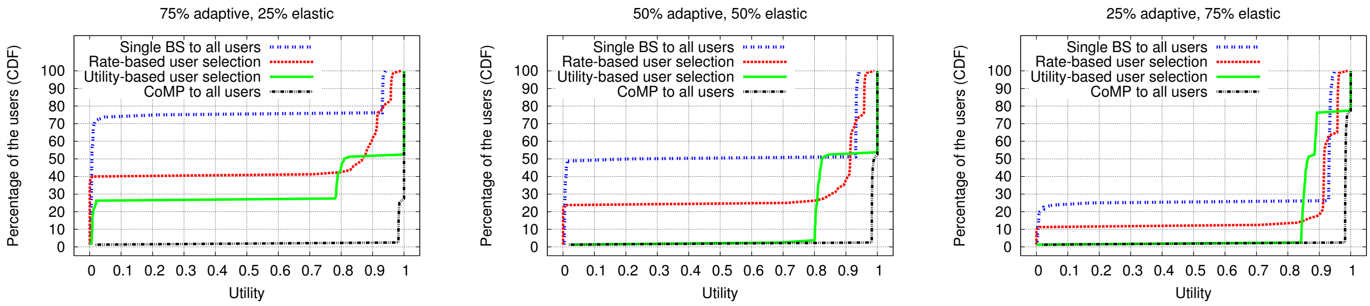
The $M = 80$ users are assumed to run one application each. The applications have hard real-time, adaptive or elastic utility functions, as described in Sec. II. The traffic mixes shown in Table I, with various percentages of the users of each traffic type, are used to illustrate the performance of the proposed utility-based user selection. The hard real-time applications generate a packet at $20\,\mathrm{ms}$ intervals. The adaptive and elastic applications are modelled as always having full buffers. The utility-based pre-selection process selects $M_{\mathrm{c}} = 40$ users for CoMP as add-on service. All $M = 80$ users are served with single BS transmission. In every time slot $M_{\mathrm{td}} = 20$ users are scheduled by the TD scheduler. CoMP transmission is used at regular CoMP intervals that occur every 4th time slot and has a duration of one time slot. The rest of the time slots are used for single BS transmission. As comparison, a rate-based user selection approach has been evaluated. The rate-based approach is similar to the utility-based approach, but it is unaware of application utility. Users are instead selected for CoMP based on expected rate gains. In the rate-based user selection, all users are seen as having the same requirements. The user selection approaches are also compared to the case when all M user are served with only single BS transmission, and to the case when all M users are served with only CoMP. In these cases, no user selection is performed, and all time slots are used for the applied transmission scheme.

In Fig. 3, cumulative distribution functions of the number of users vs. application utility is shown for traffic mixes 1-3. All users have $R = 2\,\mathrm{Mbps}$. For the utility-based user selection, the TBR values correspond to $u_{\,\mathrm{target}} = 0.8$, which gives $\mathrm{TBR} \approx 0.1\,\mathrm{Mbps}$ for elastic applications and $\mathrm{TBR} \approx 1.1\,\mathrm{Mbps}$ for adaptive applications. The rate-based user selection has $\mathrm{TBR} = 2\,\mathrm{Mbps}$ for all users.

The results for traffic mix 1 is presented in Fig. 3a. The percentage of the users with $u < 0.1$ is close to 30% for the utility-based user selection, which is significantly lower than than the 40% for the rate-based approach. Both approaches perform much better than if only single BS transmission is used for which about as much as 75% of the users have $u < 0.2$. Even with CoMP a few users have $u \approx 0$. These users have channels with low quality due to shadow fading or interference, or both.

The results for traffic mix 2 and 3 illustrated in Fig. 3b and Fig. 3c, respectively, indicate that the percentage of the users with $u < 0.8$ decreases as the division of adaptive and elastic users shifts toward more elastic users and fewer adaptive. The reason is that a lower rate is required to reach $u > 0.8$ for elastic utility than for adaptive. For traffic mix 2, shown in Fig. 3b, $u > 0.8$ is reached for almost all users for the utility-based user selection. The results are even further improved for traffic mix 3, as illustrated in Fig. 3c, to $u > 0.85$ for most of the users.

For traffic mix 4, hard real-time users are included. The results are shown in Fig. 4. Also here, utility-based user



(a) (b) (c)

Fig. 3. Cumulative distribution of users vs. utility for (a) traffic mix 1, (b) traffic mix 2, and (c) traffic mix 3

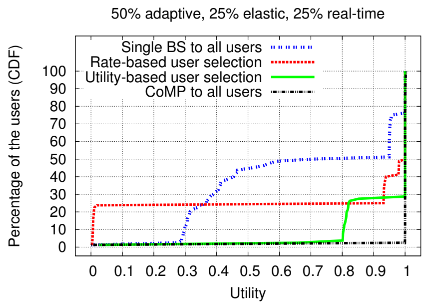


Fig. 4. Cumulative distribution of users vs. utility for traffic mix 4

selection provides a higher utility to more users than rate-based user selection. To increase the probability that hard real-time traffic reaches $u = 1$ for utility-based user selection, the TD scheduler is instructed to use $\mathrm{TBR} = 1.5R$, which is $\mathrm{TBR} \approx 131\,\mathrm{kbps}$. The gain of utility-based user selection over rate-based user selection for $u = 0.8$ is about 20%. All the compared schemes are capable of providing the maximum utility, $u = 1$, for the users of hard real-time applications, also single BS transmission to all users. The shape of the curve for single BS transmission to all users depends on the fact that the hard real-time users reach $u = 1$ even for low data rates. Therefore, there are more resources left for the users of adaptive applications, and their utility values are increased as compared to the other traffic mixes. Another difference that is visible in the figures is that there are more users with $u > 0.8$ for the utility-based user selection. No hard real-time users were selected for CoMP by the utility-based user selection, which resulted in higher utilities for the users of adaptive applications. The rate-based user selection, on the other hand, selected some of the users of hard real-time applications for CoMP.

For the tested traffic mixes, the proposed utility-based user selection increases the fraction of users for which $u \approx 0.8$ by $10-20\%$ as compared to rate-based user selection. By lowering the user requirement from $u = 1$ to $u = 0.8$ for adaptive and elastic applications and selecting the users that would gain more utility-wise from CoMP, fairness is also improved. Fairness is an important issue to consider when resources are scarce, which is typically the case in wireless networks [18].

In Table II, Jain's index [19] is indicated to illustrate fairness of utility for the tested traffic mixes. If all users would have the same utility value, then Jain's index would be 1. For CoMP transmission to all users, the highest fairness is achieved, and for single BS transmission to all users the lowest. When CoMP is used, the available data rate is high enough to reach high utilities for almost all users, except for the ones with very low channel qualities. When all users are served with single BS transmission, instead, fewer users reach acceptable levels of utility. The proposed utility-based user selection improves fairness considerably as compared to rate-based user selection for the tested traffic mixes, since the users are selected and scheduled depending on their utility functions. For traffic mix 1, CoMP transmission to all users is required to achieve a high fairness index. For the other traffic mixes, utility-based user selection provides almost as high fairness indices as when all users are served with CoMP. Resources are redistributed to the users with low utilities at the cost of fewer users with $u \approx 1$.

TABLE II. FAIRNESS OF UTILITY, JAIN'S INDEX

	Mix 1	**Mix 2**	**Mix 3**	**Mix 4**
Single BS to all users	0.27	0.51	0.76	0.83
Rate-based user selection	0.60	0.76	0.89	0.77
Utility-based user selection	0.73	0.98	0.98	0.98
CoMP to all users	0.99	0.99	0.99	0.99

V. CONCLUSION

In this paper, we have proposed a utility-based user selection approach for limited CoMP. User selection is performed on two separate time scales. On the long term time scale, a pre-selection process selects users for CoMP based on their utility requirements and on which users that would gain more from CoMP transmission. On a short time scale, joint time and frequency domain scheduling is applied. The time domain scheduler is unaware of utility, but operates based on target bit rate values that have been set by the pre-selection process according to the utility requirements. The frequency domain scheduling is performed per cell, as in conventional systems. In case of CoMP transmission, the users scheduled in the same resource block are precoded for CoMP transmission by a central control unit. The advantages of the utility-based user selection are that the CSI feedback overhead required for CoMP is reduced, the total utility of the users is increased, and fairness is improved compared to if utility is not taken into account when users are selected.

ACKNOWLEDGMENT

This work is part of Project 2009-4555 Dynamic Multipoint Wireless Transmission which is financed by the Swedish Research Council.

REFERENCES

[1] 3GPP, "TR 36.819 v11.2.0, 3rd Generation Partnership Project; Technical specification group radio access network; Coordinated multi-point operation for LTE physical layer aspects, (Release 11);," September 2013.

[2] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, pp. 148–155, February 2012.

[3] W. Lee, I. Lee, J. S. Kwak, B.-C. Ihm, and S. Han, "Multi-BS MIMO cooperation: challenges and practical solutions in 4G systems," *IEEE Wireless Communications*, vol. 19, pp. 89–96, February 2012.

[4] F. Diehm and G. Fettweis, "Centralized scheduling for joint decoding cooperative networks subject to signalling delays," *2011 IEEE 74th IEEE Vehicular Technology Conference: VTC2011-Fall, San Francisco, CA, USA*, pp. 1–5, September 2011.

[5] S. Mehryar, A. Chowdhery, and W. Yu, "Dynamic cooperation link selection for network MIMO systems with limited backhaul capacity," *IEEE International Conference on Communications, ICC, Ottawa, Canada*, June 2012.

[6] L. Shi, C. Liu, and B. Liu, "Network utility maximization for triple-play services," *Computer Communications*, vol. 31, pp. 2257–2269, June 2008.

[7] A. Klockar, C. Botella, M. Sternad, A. Brunstrom, and T. Svensson, "Utility as a user selection criterion for coordinated multi-point systems," *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), London, United Kingdom*, pp. 3123–3128, September 2013.

[8] 3GPP, "TS 23.203 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 11)," March 2012.

[9] Balamurali, "A low complexity resource scheduler for cooperative cellular networks," *Proceedings of the 3rd IEEE international conference on Internet multimedia services architecture and applications (IMSAA'09)*, pp. 182–187, 2009.

[10] A. Papadogiannis, H. Bang, D. Gesbert, and E. Hardouin, "Efficient selective feedback design for multicell cooperative networks," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 196 – 205, January 2011.

[11] P. E. Mogensen, W. Na, I. Z. Kovcs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. E. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the Shannon bound," *The 65th IEEE Vehicular Technology Conference, VTC2007-Spring, Dublin, Ireland*, pp. 1234–1238, April 2007.

[12] Q. Zhang and C. Yang, "Transmission mode selection for downlink coordinated multipoint systems," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 465–471, 2013.

[13] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Communications Surveys & Tutorials*, no. 99, pp. 1–23, 2012.

[14] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[15] G. Monghal, K. I. Pedersen, I. Z. Kovács, and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," *2008 IEEE 67th Vehicular Technology Conference-Spring (VTC Spring), Marina Bay, Singapore*, pp. 2532–2536, May 2008.

[16] R. Apelfröjd and M. Sternad, "Measurement based evaluations of coherent JT CoMP - a study of precoding, user grouping and resource allocation with predicted CSI," *EURASIP Wireless Communications and Networking*, doi: 10.1186/1687-1499-2014-100, June 2014.

[17] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," *European Wireless, Barcelona, Spain*, February 2004.

[18] H. Shi, R. V. Prasad, E. Onur, and I. G. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 5–24, 2014.

[19] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *DEC Research Report TR-301*, 1984.