

RESEARCH

Open Access

A novel scheduling framework for QoS-aware OFDMA resource allocation in a network with small relay cells and macro users

Venkatkumar Venkatasubramanian* and Thomas Haustein

Abstract

Relaying is a convenient way to provide full coverage in cellular networks. In particular, small relay cells can be used as a cost-effective solution for indoor coverage in MIMO-OFDM systems. The small relay cells would need to cater for indoor users' quality of service (QoS) expectations. One key QoS objective is delivering stable data rates for multimedia applications, which we refer to as guaranteed data rates. In this article, we consider optimization for delivering guaranteed data rates in a network with multiple relays and a macro base station, in a scenario when there are both macro users and relay users to be served. A novel scheme called cell-guaranteed bit rate by relay scheduling is proposed, with both optimal and heuristic scheduling methods. To perform the optimization we exploit resource block allocation, and parameters such as relaying duration and relay bandwidth allocation. Interference between relays and macro is avoided through time domain orthogonalization. Another key aspect of the scheme is inter-frame scheduling, wherein relay feeder links can be flexibly scheduled in any time slot along with macro users. Performance evaluation is presented using real-time indoor measurement channels and a sample test scenario. Results show the heuristic method can improve performance by 89.47% as compared to round-robin scheduling at relays and is within a 5% gap to optimal scheduling.

Introduction

The idea of relaying has attracted high interest over the last decade as a solution for improving coverage in cellular networks. One way to realize relay-based solutions is through installation of dedicated 'helper' nodes, which are also known as fixed relays or infrastructure relays. For example, in the literature, the studies [1,2] discuss some practical deployment scenarios and methods such as time division multiple access, frequency division multiple access (FDMA), in-band relaying and out-of-band relaying for fixed relays.

Fixed relays have traditionally been deployed in cellular networks as repeaters; a device which re-transmits after boosting the signal. The repeaters are installed by a service provider and their transmission parameters are set to fulfill the network needs and standardization requirements. Such devices have been a part of the coverage solution for both global systems for mobile communications (GSM)

and wideband-code division multiple access systems, for example, as evaluated in [3].

The new generation of cellular technologies such as Worldwide Interoperability for Microwave Access (WiMAX), or long-term evolution (LTE) use orthogonal frequency division multiplexing access (OFDMA) as the air interface. At Fraunhofer Heinrich Hertz Institute (HHI) we have developed an LTE test-bed to demonstrate the features of LTE, showing capabilities such as peak downlink data rate of 160 Mbps using 2×2 multiple input multiple output (MIMO) over 20 MHz bandwidth [4].

Recent indoor LTE measurements have shown the problem of so-called coverage holes in macro cells at 2.6 GHz [5]. These are areas in the cell which receive markedly low signal power. It has been observed that attenuation from obstacles in urban environment can add to pathloss attenuation at 2.6 GHz, and thereby limit outdoor to indoor coverage. For example, building penetration loss and modern window coatings originally designed for thermal insulation add to the mean pathloss at 2.6 GHz. Field trials with indoor relays showed that very good coverage can be obtained from MIMO-OFDM air interface.

*Correspondence: venkat.kumar@hhi.fraunhofer.de
Wireless Networks, Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany

Indoor relays become important to future cellular network for the following reasons. First, some recent statistics have shown that indoor users originate 60–90% of the cellular traffic [6,7]. Thus, indoor coverage solutions are important. Second, high targets for indoor coverage have been set by International Telecommunications Union (ITU) by requiring an average cell spectral efficiency of 3 bits/s/Hz/cell and user spectral efficiency of 0.1 bits/s/Hz/cell indoors for fourth generation systems (4G) [8]. LTE-advanced [9] has been working on indoor coverage solutions to meet the target. Third, relays are a convenient way to realize indoor small cell solutions because installation of expensive cables are not necessary.

Deployment of advanced relaying by data regeneration, also known as decode and forward (DF), is under discussion by 3GPP, classified as types 1 and 2 relays. System-level investigations with DF relays for OFDM cells in [10] showed a 3-dB power gain for indoor users. Improvement in minimum data rate based on Shannon capacity formula has been shown based on the channel sounding measurements in 5 GHz from an indoor relay in [11].

In this article, we propose an approach for improving the performance of DF in-band multi-antenna relays through resource optimization. Our main resource optimization tool is diversity in channels across frequency subcarriers and multiple users. The process of utilizing this diversity at relays is called scheduling at relays. In the spirit of this approach, we conducted indoor field trials at 2.6 GHz and full 20 MHz bandwidth using multiple antenna relays (MIMO relays) to characterize the indoor channel frequency response. We also performed data rate evaluation based on channel quality feedback in frequency division duplexing (FDD) operation and interference-free scenario.

Prior works and contributions

Among related works in scheduling-based relaying, the literary works [12,13] propose fairness approaches to OFDMA relaying assuming multiple parallelly activated relays in a time slot. A similar architectural assumption is used in this article, thereby enabling spatial frequency reuse among many relays within a macro-cell. For relaying without spatial reuse, we refer to works such as [14], which performs optimization by enforcing that only one relay link is active on a subcarrier in the entire macro-cell. The authors of [12] propose resource partitioning of feeder links based on time slots by considering that the relays transmit concurrently in one time slot. They further optimize the time slot splits between relay feeder links to improve relaying efficiency, without dealing with user scheduling issues at relay access links. The study in [13] proposes a relay scheduler called multiple relay parallel activation, in a framework similar to [12] but wherein they make use of user-subcarrier scheduling at each relay. The

scheduler however does not optimize relaying time durations. More importantly, their scheduler tries to minimize the outage of *all* the active users and thus also does not exploit admission control procedures.

Our objective in relaying is to maximize the number of users who are given guaranteed bit rates (GBR). This approach is applicable to a real-time mode called GBR in LTE which allows for an admission control method to be employed.

In [15], we proposed a novel scheme called GBRS for the case of relay GBR users. GBRS protocol maximizes the guaranteed bit rates offered in an LTE cell based on joint user scheduling, relay access time optimization, relay spatial reuse, and admission control. Admission control is performed hierarchically also taking the feeder link efficiencies into account.

In this study, we go a step further and deal with a situation which will be often encountered at the system level; there are both macro and relay GBR users to be served in a cell. We provide a novel approach to improve performance in this challenging situation by building on the GBRS protocol and term this cell-guaranteed bit rate by relay scheduling (cell-GBRS).

The novel ideas behind our cell-GBRS algorithm which we present in this article are (a) decoupling resource allocation problems for relays and base station and (b) inter-frame scheduling. Inter-frame scheduling denotes the approach of flexible resource partitioning, wherein the scheduling of macro-base station users can be interspersed in dedicated relay time slots. Similarly, the base station opportunistically activates feeder link to the relays in few subcarriers within the macro-users frames. This possibility of interspersing relay and macro users is already made use of in [13]. However, fairness between the macro users and sets of relay users is an issue which needs further attention. In our cell-GBRS scheduler, we address this issue and provide a fairness-based framework for supporting the relay and macro users. In our framework, we use independent sub-schedulers in dedicated frames; one for the macro-users and one for each relay set and use inter-frame scheduling to connect the sub-schedulers. Through detailed description of the algorithm, we show how to best utilize the network resources through our scheduling framework.

In addition, our scheduling algorithm can also handle possible interference scenarios between adjacent relays via interference avoidance (orthogonalization). Thus, we remark that our approach and results in the article may be applicable for a wide variety of network situations. Through these contributions, we hope to serve the following purposes: (a) Our scheme can be utilized for providing guaranteed bit rates in a real-time system for any particular relay deployment scenario and (b) Our scheme can be used to perform system level simulations for various

node placement settings and thus predict the system performance in terms of number of supported users.

As a suitable illustration, “Example numerical results” section presents simulation-based results for a single-cell scenario considering 1 Mbps bit rate per user. For presenting these results, we use quantized 26 modulation and coding levels per frequency resource block, which has been adopted for WINNER studies [16].

One attractive feature of our relaying solution is however that any MCS feedback scheme can be applied. This means that the relaying and macro-cell algorithms need not be re-worked for a change in MCS feedback levels. For instance, as a response to uplink congestion, the MCS feedback level can be reduced to either coarse, fine, or extended feedback scheme, as we presented in [5]. Results show that our relaying scheme is robust to reduction of uplink feedback. A 91.7% reduction using extended feedback scheme results in only 5% performance loss.

We compare our results with a baseline synchronous round-robin relaying scheme based on [12]. Results show improvement of 89.47% as compared to the baseline scheduler.

The article is organized as follows. In “Relay deployment” section, the scenario for relaying is discussed. In “Scheduling framework in a cell” section, the framework of relaying and optimization is described. In “Problem definition” section, the problem statement is given. In “Scheduling steps” section, the detailed steps of the proposed cell-GBRS scheduler is provided. “Channel measurement study” section presents the indoor channel measurements which are used for our results. In “Example numerical results” section, we illustrate the performance benefit of our scheduler through simulations in an example test scenario.

Relay deployment

We consider the downlink of a large macro-cell comprising a macro base station and few relay transmitters. There are indoor users located in the cell in many residential/office buildings, and only few buildings are equipped with the relays. Thus, the cell has to cater a mixed scenario of direct and relayed transmissions. The users who are catered directly by the base station are called macro-users, and the users who depend on a relay are called relay users. The large macro-cell has dimensions in the order of hundreds of meters.

The relays are DF relays, in which case a relay fully decodes the data from the base station before forwarding to the users. The relay applies uniform transmit power on all subcarriers, schedules the users on selected resource blocks, applies modulation and coding scheme (MCS) and re-transmits on the access link. A user is assumed to be given handover to the relay only if the received signal power (RSRP) from the relay is satisfactory. The data

transfer from the macro base station to the relay is done on the same air interface, which is called in-band relaying. The coverage area of a relay is relatively smaller than a macrocell and is called a relay cell. We do not differentiate between type 1 and type 2 relays [17], or other signaling requirements but rather focus on the data rate improvement that is achievable through deployment of relays.

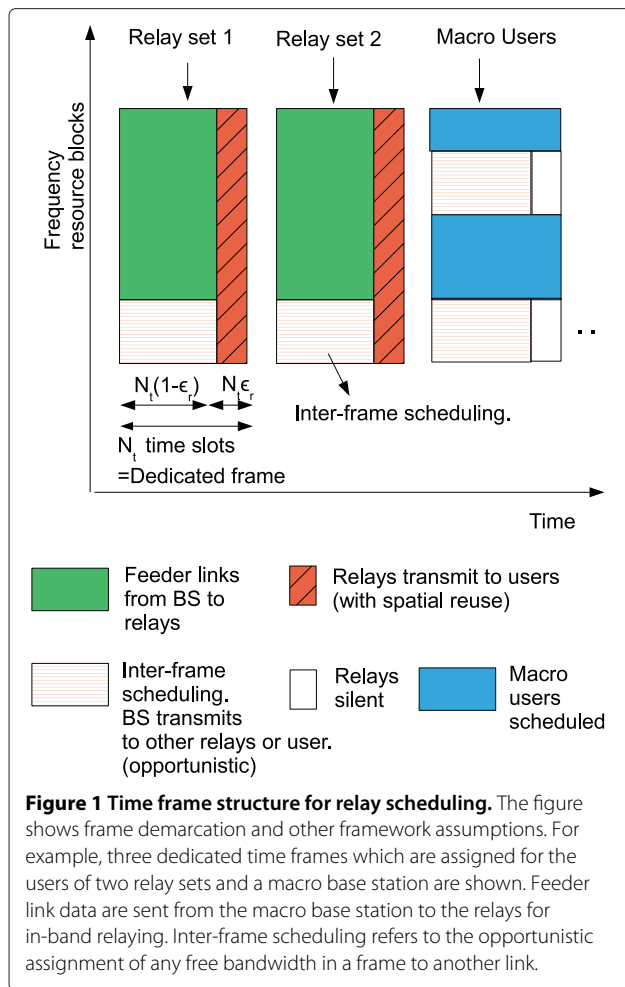
The relay can be placed at a convenient location indoors, a deployment which is similar to a Wifi access point. The relay unit consists of two parts: a feeder unit which connects to the macro base station and an access unit which connects to the user equipment. All transmitters and receivers are equipped with multiple antennas, thus providing 2×2 MIMO-OFDM air interface on three separate links: (a) the direct link from base station to macro-users, (b) the feeder link from base station to relay feeder unit, and (c) the access links from relays access unit to relay users. The access and feeder links of a particular user can be on different set of frequency sub-carriers.

Scheduling framework in a cell

The purpose of scheduling in OFDM cells is to ensure that the radio resources are utilized both efficiently and with some fairness. Our proposed scheduling framework in a macro cell consists of two aspects: relay scheduling and user scheduling. Relay scheduling concerns which set of relay nodes would transmit in a time slot. User scheduling deals with scheduling macro users and relay users on resource blocks in a time slot. To incorporate the above two aspects of scheduling, we consider a dedicated frame structure as in Figure 1.

In this frame demarcation N_t time slots make a so-called dedicated frame. A certain set of relays and users of those relays are served in each dedicated frame. Figure 1 shows an example of three dedicated frames, one for users of relay set 1, one for relay set 2 users, and one for macro-users. The introduction of dedicated frames provides fairness by pre-allocating equal number of resources to each relay set and macro-users. This notion of fairness thus avoids only few relay cells from consuming the system bandwidth. There are two types of time divisions.

- Frame splits: frame splits are parameterized to be the time durations for which each of the dedicated frames transmit out of the total time. A dedicated frame t is thus assumed to transmit for N_t time slots (0.5 ms each time slot). The configuration of the frame split ratios will depend on the system level fairness targeted for different sets of users. For example, factors such as whether the macro-users are prioritized over the relay users may influence the frame splits.
- Relay time splits: relay time splits denote the time durations for which the relay feeder link and access



link are active out of the N_t time slots. The relay time splits are denoted using an optimization variable ϵ_r , refer Figure 1, such that the relay receives for $N(1 - \epsilon_r)$ time slots and transmits for $N(\epsilon_r)$ time slots out of N time slots.

The other physical layer aspects in the proposed framework are as follows.

- Full frequency reuse: The main idea of our relaying solution is to let each relay utilize the full transmission bandwidth, which we term full frequency reuse. In our viewpoint, the major benefit of this approach is the spectral efficiency gain that is achievable from multi-user diversity and frequency diversity on the relay access link. This gain can provide better network performance via appropriate user scheduling algorithms, and can be exploited at all the relays and the macro base station. Therefore, the relay feeder link and access link are split into time phases (as shown in Figure 1) instead of frequency division. One important feature is that the access and feeder links of a particular user can be on different set

of frequency sub-carriers which provides a high degree of freedom for scheduling.

- Interference coordination: A cellular network may consist of a number of relays deployed for specific coverage needs. One example is installation of relays in an office building consisting of many floors, wherein a relay is deployed on each floor. All the relays may then utilize the same licensed frequency band for access links. Thus, interference situations may arise if the relays are closely placed or if users of one relay move to another relay cell. Interference coordination is then necessary. This coordination is handled according to the frame structure in Figure 1 by scheduling different sets of relays in each time slot. To do this, a set of non-interfering relays is defined as a relay set. Thus, a relay is first included in a relay set and then the relay set is scheduled on a time slot. This means that potentially interfering relays are orthogonalized by being scheduled on different time slots. At the same time, care should be taken to avoid possible relay to macro-user interference. Therefore, we enforce that the relay transmitters are silent during the dedicated frames (time slots) assigned to macro-users.
- Spatial reuse: In reality not all the relays would actually interfere with each other. The transmission power of relays is low (typically 23 dBm) which means that the relay cell sizes are also relatively smaller. Thus, multiple relays may be able to transmit within a large macro-cell without causing interference to each other if they are sufficiently separated. Therefore, non-interfering relays can be scheduled simultaneously which we term spatial reuse.
- Inter-frame scheduling: Users in a large cell are associated with either a nearby relay or directly to the macro-base station. For overall fairness, time slot splits are fixed among dedicated frames, i.e., between each relay set (and its set of users) and macro base station. The fixed demarcation of time slots may however result in under-utilization of resources in case there are not enough active users in one or more time frames. To overcome this problem, we propose inter-frame scheduling, wherein free bandwidth from one dedicated frame is temporarily re-assigned to another frame and to the link which can best utilize it. Inter-frame scheduling process is explained in detail in “Inter-frame scheduling” section.

Problem definition

Summary of variables and constants

The following are the list of constants and variables.

Constants:

- Ω_k : The data rate demand in terms of bits per time slot to be loaded for user k .

- M : The maximum number of resource blocks in the downlink for a macro-cell assumed to be available in a time slot.
- Q : The total number of dedicated frames.
- u_{km} : The modulation and coding value for m th resource block and the k th user to achieve a target bit error rate. These are obtained from the channel estimation in the downlink. It is represented in loaded bits per subcarrier on an OFDM symbol (data symbol) and applied throughout that resource block.
- L, T : L is the number of data subcarriers per OFDM symbol in the downlink for a macro-cell. T is the number of OFDM symbols per time slot. The product $L \times T$ is for example 144 data symbols. The product $L \times T \times u_{km}$ will give the bits loaded per resource block.
- N_t : The total number of time slots in a dedicated frame. We drop the suffix t for convenience and just call this N . Each dedicated frame consists of time slot splits for feeder and relay transmissions.
- R, R_t : The total number of relays in the macro-cell and the total number of relays in frame t , respectively.
- R_U^t : The set of relays in the macro-cell in a frame t .
- γ_r : The spectral efficiency (averaged over all the subcarriers) of the feeder link from base station to the r th relay. It is assumed that this spectral efficiency can be realized by coding the feeder link data symbols across few randomly distributed subcarriers in the frequency domain.
- K, K_r, K_M : K is the total number of active users, K_r is the number of users affiliated to the r th relay. K_M is the number of macro-users affiliated directly to the macro base station.
- F_r : Feasible set of users at the r th relay.
- U_r, U_M, U_t : U_r is the set of users affiliated to the r th relay. U_M is the set of macro-users indices affiliated directly to the macro base station. U_t is the set of all users affiliated to the t th dedicated frame at the start of scheduling.

Variables:

- a_k : Binary variable to model data rate satisfaction of user k . a_k equals 1 if data rate Ω_k has been assigned and equals zero if data rate is less than Ω_k .
- x_{km} : Allocation variables showing the amount of allocation of m th resource block to the k th user in the relay to user access link. $0 \leq x_{km} \leq 1$.
- b_r : Bandwidth allocation variables allocated to the r th relay's feeder link borrowed from the frames other than the current one t . $b_r \geq 0$.
- $\hat{\Delta}_r$: Bandwidth allocation variables in terms of resource blocks allocated to the r th relay's feeder link during the current frame t . $0 \leq \hat{\Delta}_r \leq M$.

- ϵ_k : Duplex time sharing variables showing the time allocation to the k th user's access link. This is normalized to N , are thus fractions $0 \leq \epsilon_k \leq 1$. It takes the value 1 for macro-users.
- ϵ_r : Duplex time sharing variables showing the time allocation to the r relay to transmit. This is normalized to N , and are thus fractions $0 \leq \epsilon_r \leq 1$.
- M_S : Surplus bandwidth resources available for inter-frame scheduling.

The overall cell objective is to satisfy the maximum number of users for each dedicated frame t

$$\max \sum_k a_k \quad (1)$$

$$\forall k \in U_t$$

For achieving (1) there are however various radio resource constraints as we would observe in the following sections. The detailed steps of the macro-user and relay scheduling phases are presented in the following sections.

Scheduling steps

User connection

In the first step, each user indicates the preferred transmitter for connection (which relay or macro) and is affiliated to that node. This step of user association can simply be performed based on the well-known key performance indicators, e.g., signal-to-interference-noise ratio (SNR) or RSRP. We assume that a user connects either to the closest relay or macro-base station within its coverage zone.

Relay grouping

In the second step, the relay transmitter nodes are grouped into different sets. Each relay set transmits on its corresponding dedicated frame. This grouping of transmitters is based on the condition of realizing very low interference mutually and in practice can be done in two ways. The first way is static, wherein the global positioning system coordinates of the transmitters can be used to work out the inter-transmitter distance and thereby find the relays which are sufficiently out of range. The second and more robust way is dynamic, wherein a central controller or the base station obtains interference reports from the users of each respective relay. Thereafter any two relays (r_1, r_2) are decided to be mutually non-interfering only if all the users of r_1 report very low interference from r_2 and vice versa. We provide a simple algorithm called SUFFICIENTGROUPING which can be applied for the case of interference reports.

- Enlist all relays of the macro-cell in a set $SET = \{1, 2, \dots, R\}$

- For each relay r in SET , obtain interference reports from its users which contains the ids of interfering relays. By using all interference reports at all relays, minimum number of sets of mutually non-interfering relays are to be obtained. This is done as follows.
- Start with relay $r = 1$ and incrementally add mutually non-interfering relays to the set. Call this as interference set I_1 . Thus no two relays in I_1 are interfering each other. Proceed to the next relay not in I_1 , obtain mutually non-interfering relays and name the set I_2 . Continue the process until there are no more relays left. Thus in Q' iterations, we have Q' sets of mutually non-interfering relays, each being a subset of SET .
- Now obtain the minimum number of sets out of Q' such that all the relays are covered. This is a standard set cover problem that can be solved through a greedy heuristic [18]. As a result we would have $Q - 1$ sets.
- Perform the following sequentially for relays $r = 1$ to R . If a relay r is in more than one subset, keep it only in the set with the minimum number of relays. This step ensures that relay r is not feeder link throughput constrained and also that we prevent a relay from transmitting twice in $Q - 1$ frames. Remove it from all the other subsets. Ties are broken randomly.

At the end of the grouping algorithm, we would thus have $Q - 1$ relay sets and one macro-base station, which are scheduled in Q dedicated frames. Now receiver scheduling is done in two phases: (a) phase 1 for dedicated frames, consisting of macro and relay users and (b) phase 2 for inter-frame scheduling.

Macro dedicated frames

The macro dedicated frames are frames in which macro-users are scheduled for reception. In this section, we describe the scheduler for macro user selection and resource block allocation. In this step, resource block scheduling is done for the direct link from the macro base station to the macro-users without allocating any resource blocks to the relay feeder link.

The objective function for macro scheduling is framed as

$$\max_{x_{km}} \left(\sum_{k \in U_M} a_k, - \sum_m \sum_{k \in U_M} x_{km} \right), \quad (2)$$

where a_k is 0 if data rate of user k is less than Ω_k and 1 otherwise.

This above multi-objective has been formulated intuitively keeping in mind the possibility of resource shortage in other frames for feeder links. The first objective maximizes as many user guarantees as possible which is the

key performance indicator. The second objective minimizes the bandwidth consumed for the direct link users while providing those data rates. For example, satisfying four users instead of just three is a better resource allocation as per the first objective, whereas satisfying those four users with minimal bandwidth is the target of the second objective. For convenience, we call the first objective as MAXUSER and second objective as MINBANDWIDTH.

We note that we have intentionally not formulated sum-rate maximization or proportional fairness as the second objective. The effect of (2) is that the macro scheduler is expected to free as many additional resources as possible for the relay feeder links in other frames.

Optimally solving the dual-objectives in (2) can be done sequentially via two linear programs (after relaxing the variables to be continuous) using solvers such as LIPSOL [19]. The MAXUSER problem can be solved first as a linear program to obtain the feasible set of users, and then solve the MINBANDWIDTH problem as another linear program for the feasible set of users. It may however be too complex to be beneficial in real time as it involves solving two linear programs.

Sort and greed scheduler

We now present a simple heuristic algorithm called 'Sort and greed' for scheduling macro users towards the objectives in (2). We solve the problem iteratively as follows.

- Start with a user set U_M which is the full set of macro users and a set of resource blocks $V_M = \{1, \dots, M\}$. Perform the following iterations to solve for $(k^*, x_{k^*m}^*)$ and update U_M and V_M .
- In iteration n , solve

$$k^* = \arg \max_{k \in U_M} \frac{\sum_m u_{km}}{\Omega_k} \quad (3)$$

$$\min_{x_{k^*m}} \sum_m x_{k^*m} \quad (4)$$

$$s.t \sum_m x_{k^*m} u_{k^*m} \geq \Omega_{k^*} \quad (5)$$

$$x_{k^*m} \in \{0, 1\} \quad \forall m \quad (6)$$

Equations (4)–(6) can be solved as a simple greedy selection of resource blocks for user k^* . Hence, the name 'sort and greed'.

- The user k^* and its corresponding set of allocated resource blocks are removed from U_M and V_M , respectively, after each iteration. Proceed to iteration $n + 1$. The method continues until either U_M or V_M is empty.

Scheduling algorithm at relays

This step proposes a scheduling solution for relay users to be implemented at each of the R relays.

In-band relay problem

We begin by describing the problem for a relay in dedicated frame t . It can be recollected that many relays can be expected to be co-scheduled in the same frame. The objective of maximizing the number of satisfied users for the relaying case is written as

$$\max \sum_k a_k \tag{7}$$

$$\text{s.t } a_k = 1,$$

$$\text{if } NLT \sum_m u_{km} x_{km} \epsilon_k \geq N\Omega_k \tag{8}$$

$$a_k = 0,$$

$$\text{if } NLT \sum_m u_{km} x_{km} \epsilon_k < N\Omega_k \tag{9}$$

$$\forall k \in U_t \tag{10}$$

In (8), ϵ_k denotes the fraction of time slots for which the relay access link to user k is active out of the total N time slots. This time splitting on a user-subcarrier basis denotes a *per user half-duplex operation*. This half-duplex functioning however might require tight filtering and isolation requirements at relay feeder unit to avoid inter-carrier interference.

A more practical approach is a *per relay half-duplex operation*. In this case, we require that when the access link of a relay r is activated, the feeder link altogether stops to that relay r . Because of this *per relay half-duplex operation*, the variables ϵ_k are now replaced by *per relay half-duplex* variables ϵ_r . A spectral efficiency of γ_r averaged over all the subcarriers is assumed for the feeder link to relay r . The assumption of average spectral efficiency means that the subcarriers comprising a resource block allocated to that relay are randomly distributed in frequency domain. We point out that a practical mechanism for distributed subcarrier resource allocation to a relay is available through resource allocation type 2 in LTE [20].

The relay may borrow bandwidth b_r from all other dedicated frames which have a surplus. Thus, we have

$$N\hat{\Delta}_r \gamma_r (1 - \epsilon_r) + N\gamma_r b_r = N\epsilon_r \sum_m \sum_{k \in U_r} u_{km} x_{km}, \tag{11}$$

where (11) says that amount of feeder link data is equal to the amount of access link data for r th relay. This gives the half-duplex time sharing values

$$\epsilon_r = \frac{\hat{\Delta}_r \gamma_r + \gamma_r b_r}{\sum_m \sum_{k \in U_r} u_{km} x_{km} + \hat{\Delta}_r \gamma_r} \quad \forall r \tag{12}$$

The set of equations (7)–(10) can be relaxed using continuous variables a_k . Following the relaxation, it can be transformed using weights Ω_k^{-1} as we showed in [21]. Upon plugging the value of ϵ_r from (12) into this relaxed

objective, we get the following set of equations for a given set of relays R_t^U in frame t :

$$\max_{LT} \sum_{k=1}^{K} \Omega_k^{-1} \frac{\sum_{m=1}^M u_{km} x_{km} (\hat{\Delta}_r \gamma_r + \gamma_r b_r)}{\sum_m \sum_{k' \in U_r} u_{k'm} x_{k'm} + \hat{\Delta}_r \gamma_r} \tag{13}$$

$$\text{s.t } LT \frac{\sum_{m=1}^M u_{km} x_{km} (\hat{\Delta}_r \gamma_r + \gamma_r b_r)}{\sum_m \sum_{k' \in U_r} u_{k'm} x_{k'm} + \hat{\Delta}_r \gamma_r} \leq \Omega_k \quad \forall k, \tag{14}$$

$$\sum_{k \in U_r} x_{km} = 1 \quad \forall m, \forall r \in R_t^U, \tag{15}$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m, \tag{16}$$

$$\sum_{r \in R_t^U} \hat{\Delta}_r = M \tag{17}$$

$$\sum_{r \in R_t^U} b_r = M_S \tag{18}$$

$$\hat{\Delta}_r \geq 0, \quad b_r \geq 0 \quad \forall r \in R_t^U, \quad r = H(k) \tag{19}$$

where the mapping function $H(k)$ uniquely maps a user index k to a relay r .

The solution to the problem requires solving: resource allocation variables at relays x_{km} , bandwidth allocation variables in current dedicated slots $\hat{\Delta}_r$, and bandwidth allocation variables of surplus b_r .

Straightaway we note that the variables b_r multiply with x_{km} , which means that they cannot be re-written in a linear form. We thus use a slight work around and assume that the feeder link active duration to a particular relay does not vary w.r.t the frame number of the dedicated frame.

The above assumption changes (11) into

$$N\hat{\Delta}_r \gamma_r (1 - \epsilon_r) + N\gamma_r b_r (1 - \epsilon_r) = N\epsilon_r \sum_m \sum_{k \in U_r} u_{km} x_{km}, \tag{20}$$

which gives

$$\epsilon_r = \frac{(\hat{\Delta}_r + b_r) \gamma_r}{\sum_m \sum_{k \in U_r} u_{km} x_{km} + (\hat{\Delta}_r + b_r) \gamma_r} \quad \forall r \tag{21}$$

From hereon, one could simply substitute the summation of variables $\hat{\Delta}_r + b_r$ as a new variable Δ_r which has a new upper bound $M + M_S$.

Now we just need to solve resource allocation variables x_{km} and bandwidth allocation variables Δ_r . Can (13)–(19) be solved iteratively? The answer is no. The reasoning is as follows. In iteration n , let variables x_{km} are solved by treating Δ_r as constants. Assume that for a user k' in relay cell r' , the data rate inequality in (14) is met with equality. In iteration $n + 1$ because of the constraint in (14) for user k' , $\Delta_{r'}$ cannot be increased further. Thus, for all the other users in cell r' , there is no benefit from iteration $n + 1$.

To solve the problem with low complexity, our basic idea is to decouple the two problems. Importantly, we note that if the data rates are feasible in the r th relay's access link (to its users), there exists a corresponding solution $\Delta_r > 0$ for which it is also feasible in the end-to-end link. The converse part is also true: if the data rates are not feasible in the relay's access link, there does not exist a solution $\Delta_r > 0$. Based on this fact, we use the notion of a feasible list to decouple the problem. A feasible list F_r is defined as the subset of users in a relay cell r for which a scheduling solution for data rates $\Omega_k, \forall k \in F_r$ exists in the access link. Thus, $LT \sum_m x_{km} u_{km} \geq \Omega_k, \forall k \in F_r$. Now the scheduling problem at relays is to minimize the feeder bandwidth that would be needed for a feasible user list. This scheduling problem is solved in relay resource allocation (RRA) subroutine. We thus have the following stages.

- Solve $x_{km}, \forall k \in F_r, \forall r, \forall m$ with an objective to minimize the feeder bandwidth Δ_r for relay r while guaranteeing data rate demand $\Omega_k, \forall k \in F_r$. This is called the RRA sub-routine.
- Select the best set of feasible lists across all the relay cells $\{1, 2, 3, \dots, R\}$ such that the sum of feeder bandwidths does not exceed $M + M_S$ in (18). This is called the group selection (GS) sub-routine.
- The value of M_S is not known initially to the scheduler. In view of this, the group-selection subroutine is first done for each individual dedicated frame considering a feeder bandwidth limit of M . Upon implementing this phase on all the Q frames, the surplus resource value M_S is obtained.
- Finally, the inter-frame phase is realized, wherein the group selection is repeated again on the surplus resources M_S for all the links with resource shortage.

RRA subroutine

In the RRA subroutine, the objective is to minimize the required feeder bandwidth for a feasible list F_r . Implicitly, we thus enforce $x_{km} = 0, \forall k \notin F_r, \forall m$. The problem is now essentially to minimize the maximum of the feeder bandwidths computed for each user in F_r . Thus, we deduce for each user $k \in F_r$, the feeder bandwidth Δ_r^k required to realize an end-to-end rate of Ω_k . This bandwidth Δ_r^k is the critical bandwidth, which is the bandwidth required to include user k in the feasible set F_r . Equating $LT \epsilon_r \sum_m u_{km} x_{km} = \Omega_k, \forall k \in F_r$, and using ϵ_r from (21),

$$\Delta_r^k = \frac{\sum_m \sum_{k \in F_r} x_{km} u_{km} \Omega_k}{\gamma_r [LT \sum_m x_{km} u_{km} - \Omega_k]} \quad (22)$$

To minimize the feeder bandwidth needed by the set F_r , we are required to minimize the maximum taken over the set of users in list F_r for each relay r as

$$\min_{x_{km}} \max_{k \in F_r} \Delta_r^k \quad (23)$$

From the above, we may now represent the problem for the relay cell r and given a list F_r simply as

$$\max_{x_{km}} \min_{k \in F_r} \frac{\gamma_r [LT \Omega_k^{-1} \sum_m x_{km} u_{km} - 1]}{\sum_m \sum_{k \in F_r} x_{km} u_{km}} \quad (24)$$

$$\text{s.t. } \sum_{k \in F_r} x_{km} = 1 \quad \forall m, \quad (25)$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m. \quad (26)$$

RRA optimal solution

This type of problem as in (24)–(26) is a fractional linear program [22] in variables x_{km} . The optimal solution is obtained by using substitutions $z = [\sum_m \sum_{k \in F_r} x_{km} u_{km}]^{-1}$ and $y_{km} = x_{km} z$, which basically transforms the fractional problem into a standard linear program as follows by using t as a dummy variable:

$$\max t \quad (27)$$

$$\text{s.t. } \gamma_r [LT \Omega_k^{-1} \sum_m y_{km} u_{km} - z] \geq t \quad \forall k \in F_r, \quad (28)$$

$$\sum_{k \in F_r} y_{km} - z = 0 \quad \forall m \quad (29)$$

$$0 \leq y_{km} \leq z \quad \forall k, \forall m, \quad (30)$$

$$\sum_m \sum_{k \in F_r} y_{km} u_{km} = 1. \quad (31)$$

Equations (29) and (30) are written from (25) and (26), respectively. Equation (31) appears because of the substitutions. The solution for (28)–(31) can be obtained using linear programming solvers such as LIPSOL [19]. However, it can computationally be expensive for real-time implementation.

Sub-optimal method

We now present a simplified algorithm for relay resource allocation subroutine that can be implemented with less complexity. This sub-optimal scheme is called Heuristic-GBRS.

- Step 1: First, we ascertain if an admission list $A_r \subset K_r$ is a feasible list. The requirement is to obtain a feasibility certificate via scheduling but with low complexity. For this we perform an external point descent as follows.
- Step 1.1: Sum rate maximization: Greedy resource allocation is performed to allocate the resource blocks to users with highest spectral efficiency. On a per resource block basis, this is done as

$$k^* = \arg \max_k u_{km}, \quad \forall m \quad (32)$$

From (32) we basically obtain initial binary solutions of x_{km} as $x_{k^*m} = 1 \quad \forall m$ and $x_{km} = 0 \quad \forall k \neq k^*, \forall m$.

- Step 1.2: Two user sets G_1 and G_2 are formed wherein: G_1 is the set of users for who the solutions

x_{km} in Step 1.1 satisfy the targeted data rates and G_2 , the set of users for who the solutions does not provide the data rate. Thus, some resource blocks are de-allocated from users in G_1 and allocated to G_2 .

- Step 1.3: Resource block reallocation: A set of resource blocks V are pooled and is considered transferable from G_1 to G_2 . The pool V is formed such that none of the bit rate targets of the user set G_1 would be sacrificed if any one resource block in the pool was removed.
- Step 1.4: User selection: A user is prioritized based on

$$\tilde{k} = \arg \min_{k \in G_2} \frac{\Omega_k - \sum_m u_{km} x_{km}}{\sum_{m \in V} u_{km}} \quad (33)$$

The metric in (33) exploits user diversity by doing user selection on the basis of average channel quality and the rate demand. To benefit from frequency diversity, we assign a resource block to \tilde{k} on the basis of spectral efficiency as

$$m^* = \arg \max_{m \in V} u_{\tilde{k}m} \quad (34)$$

Update $x_{\tilde{k}m^*} = 1$, and $x_{km^*} = 0 \quad \forall k \neq \tilde{k}$. The resource block m^* is removed from the set V . Steps 2 to 5 are repeated until (a) all users are given their bit rates, or (b) the resource block set V is empty. Compute the total allocated data rates to the list A_r as $\sum_m u_{km}$, $\forall k \in K_r$.

- Step 2: If the list A_r was found to be infeasible, this list is ignored, we go back to Step 1 and consider another admission list. If the admission list is feasible, we proceed further. In this case, we compute the feeder bandwidth required for each user in A_r based on (22). The computed feeder bandwidths is stored in a list 'BW'.
- Step 3: We now employ scheduling once more to minimize the maximum bandwidth in BW . To do this, resource blocks are reassigned from a user k_{\min} , who requires the least bandwidth in the list BW to a user k_{\max} who requires the highest bandwidth. The pool of resource blocks allocated with k_{\min} is denoted V_{\min} . To exploit frequency diversity, we select the best resource block using $m^* = \arg \max_{m \in V_{\min}} u_{k_{\max}m}$. Further, for a more precise computation, the resource block m^* may be subdivided into g granular blocks and optimum number of granular blocks are reassigned from m^* . Steps 2, 3 are repeated until $\max BW - \min BW < \delta$, where δ is sufficiently low.

Group selection subroutine

This subroutine is used at the base station to allocation the feeder bandwidths to relays. The problem is to decide the best group of feasible user subsets from all the relays in that dedicated frame. Let us denote by F_r^i , the i th feasible

list in a relay cell r . Now the best group of user lists in all the R relay cells, out of all feasible lists $F_r^i, \forall i, \forall r$ has to be found. Let the minimum feeder bandwidth solution of the RRA subroutine to a list F_r^i be $\Delta^{\min}(i, r)$. By using binary variables s_{ir} to indicate the choice of F_r^i , we now write the following objective

$$\max \sum_i \sum_r s_{ir} |F_r^i| \quad s.t \quad (35)$$

$$\sum_i \sum_r s_{ir} \Delta^{\min}(i, r) \leq M \quad (36)$$

$$\sum_i s_{ir} = 1 \quad \forall r \quad (37)$$

$$s_{ir} \in \{0, 1\}, \quad (38)$$

Equations (35)–(38) are the integer program for which optimal solutions can be found by standard techniques. We however propose to use a sub-optimal gradient scheme. In each step of the ascent, we merely select the 'user list-relay' pair (i^*, r^*) with the minimum gradient $\frac{\Delta^{\min}(i,r)}{|F_r^i|}, \forall i, \forall r$ and set $s_{i^*r^*} = 1$. The algorithm terminates when $\sum_i \sum_r s_{ir} \Delta^{\min}(i, r) = M$ or if all the lists have been exhausted. Upon resolving $s_{ir}, \forall i, \forall r$, the feeder link bandwidth to a relay r is obtained as

$$\Delta_r = \sum_i s_{ir} \Delta^{\min}(i, r). \quad (39)$$

Complexity and overhead issues

We recall that in a relay cell r , there are K_r users. Thus, there are $\sum_{i=1}^{i=K_r} C(K_r, i)$ number of user lists that are possible in each cell, where $C(n, x) = \frac{n!}{x!(n-x)!}$. For each of these lists, the RRA subroutine has to be implemented and the feeder bandwidth has to be informed to the base station. Conveying this information to the base station can be bandwidth expensive.

To reduce complexity and the signaling overhead, we again propose to sort the users using an admission control metric $\sum_m \frac{u_{km}}{\Omega_k}$ at each relay. This metric thus takes into account the MCS values of the access link in terms of bits per subcarrier and the data rate demand from the user. The first i users of the ordered list in relay cell r make the i th admission list. If it is a feasible list, the same is denoted as F_r^i . Note that there are K_r users and thus $i = 1, \dots, K_r$.

This effectively means that each relay cell now only needs to feedback K_r values of feeder bandwidth request, i.e., inform $\Delta^{\min}(i, r)$, for $i = \{1, 2, 3, \dots, K_r\}$. The base station applies the group selection subroutine and decides the best value $i = i^*$ for each relay. This is in fact a hierarchical implementation of admission control: each relay makes K_r admission lists and informs the corresponding K_r feeder bandwidth requests to the base station while the base station decides the best admission list out of the K_r lists of each relay. Finally, from the solutions x_{km}, Δ_r

and F_r^{i*} , we simply back substitute in (12) and obtain the optimal duration ϵ_r for relay r .

Inter-frame scheduling

Following the scheduling phase of dedicated frames, the next phase of scheduling is called inter-frame scheduling. Inter-frame scheduling works in the following ways: an available resource from one frame is opportunistically assigned to either (a) a relay feeder link, or (b) macro-users. The important point is that after the dedicated scheduling phase, we are now aware of the dedicated frames on which all the users are satisfied and the amount of bandwidth resources that are available. These frames are called frames with surplus. There may be thus M_S 'surplus resources' available from all the frames (there may be some leftover resources even in frames without surplus).

The task of scheduling is now to best allocate the M_S bandwidth units to relay feeder links and macro user links. To do this we can again implement the group-selection subroutine, using M_S as the bandwidth limit. The group selection subroutine would simply refer to the already computed RRA subroutine results to deduce the bandwidth needed to support extra users. However, this procedure might incur higher complexity in case there are large number of relay sets. To cut down complexity, an incremental mechanism is used as follows. In this mechanism, in each iteration at most one extra user is added from each relay cell. The bandwidth request to support that one extra user from each relay cell is known from the RRA subroutine computation. In a current iteration, the users are selected by sorting the bandwidth requests and then choosing the minimum bandwidth request. The iterations continue until there are no more users or bandwidth left.

In practice, it might be needed that the feeder data to a relay arrives ahead of its access link getting activated. To handle this, we propose a simple feeder accumulation schedule wherein all the frames with surplus are scheduled ahead of the frames with deficit.

Interference situations

We adopt time domain orthogonalization technique to handle the following interference situations.

Relay to relay

In the notion of inter-frame scheduling, feeder data may be sent to one or more relays in a frame which is meant for another set of relays. Thus, the access link of a nearby relay may be active while the feeder data are being sent. This will cause interference on the feeder link. To handle this case, we apply time domain orthogonalization and thus limit the time duration of inter-frame feeder link to $N_t - \max_{r \in R_t} \epsilon_r N_t$. This essentially means that we switch off the feeder link when any relay of that frame starts its transmission. The feeder link data rate will thus

scale down by this factor. Fortunately, we note that this problem does not arise between any feeder link schedules on the macro frames because in problem formulation we have already enforced all the relays to be silent on those frames.

Relay to macro user

Through inter-frame scheduling, pending macro-users are opportunistically served in relay frames. In this case, the extra bandwidth needed for the pending macro-users is deduced using the same 'sort and greed' heuristic in "Macro dedicated frames" section. Based on this bandwidth estimate, the pending macro users can thus be included in the inter-frame allocation stage.

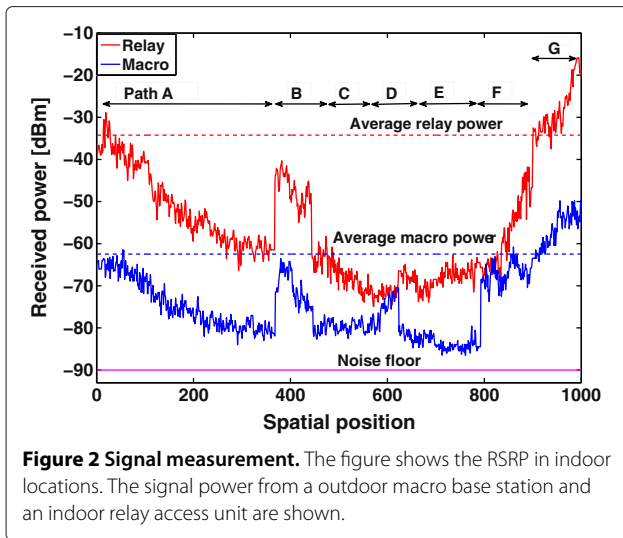
When the macro-users are opportunistically scheduled in relay frames, they may receive interference from few relays because of the fact that the relays occupy the full bandwidth on their access links. For this purpose, the macro-user scheduling is limited to time duration $N_t - \max_{r \in R_t} \epsilon_r N_t$ on relaying frame t . The data rates for opportunistic macro-user allocations will thus scale down by this factor.

Serving macro to relay user

In addition to the above two interference situations, interference could also be caused to a relay user by the serving macro base station. The serving macro may be transmitting to other relays for feeder link or directly to macro users in the cell. The first interference situation because of feeder links can arise in all the relaying frames, because the scheduler has allowed for independent half-duplex time splitting at each relay. Thus, the feeder link to one relay can in fact interfere with the access link of another relay for few overlapping time slots. But, as we presented in Figure 2, [15] this interference can effectively be handled by power control on the feeder link. Our measurements show that up to 19.5 dB of transmit power control can be applied for satisfactory results. Moreover, given the high signal noise ratio received by most indoor relay users on the relay access link, it is observed that even without power control the macro interference may affect only a few percentage of users. The second case of interference from serving macro to relay users arises when some macro users in the cell are opportunistically scheduled on relaying frames through inter-frame scheduling. We refer to the previous "Relay to macro user" section and note that this interference case is mutually avoided when the macro-user scheduling is limited to time duration $N_t - \max_{r \in R_t} \epsilon_r N_t$ on relaying frame t .

Channel measurement study

To study the potential performance benefits of relays, indoor relay measurements were conducted at an indoor cell 485 m away from a serving base station. A detailed



publication of these measurement results were done in [5]. We describe some of the key points here.

A three-terminal relay network, consisting of an outdoor macro base station, an indoor relay, and an indoor user equipment (UE), was setup at an indoor office. The link between any two nodes is 2×2 OFDM-MIMO with dedicated uplink feedback from the UE to either the outdoor base station or to the relay. All the terminals used cross-polarized antennas to make use of polarization multiplexing, wherein the two MIMO data streams are fed into a separate polarization mode. A bandwidth of 20 MHz was employed at 2.6 GHz carrier frequency. No other base station was active during the experiments. The aim of the experiment was to characterize the data rate benefit of relaying and to also collect the channel measurements. The transmission parameters are shown in Table 1, with the macro base station applying a uniform power spectral profile. The base station antennas were sectorized with 11 dBi directivity gain towards the measurement site.

The measurements were conducted in an office at the sixth floor (at height approximately 20 m) in the Heinrich Hertz Institute, Berlin which is shown in Figure 3. The office layout consisted of paths A to G, approximately of lengths 20, 5, 10, 5, 5, 10, and 10 m, respectively, as shown in the figure. Paths A, C, F, and G are corridors. Paths B, D, and E are small sections opening from the corridors. The measurements were conducted by moving the user equipment along the defined paths in a trolley. The office

Table 1 Downlink system parameters for indoor measurements

BS, RN transmit powers = 43 dBm, 23 dBm
BS downtilt = 10°
RN isolation = 20 dB

is equipped with standard office furniture and has doors at sides of the floor while one of the doors consists of a glass front. Note that in case of heavy window coating, the relay feeder unit's antennas will be placed outside the window but the access unit's antennas can still be placed indoors to avoid the penetration loss.

The relay feeder unit which connects to the base station was positioned in path G (at a distance of approximately 485 m from the base station). The feeder unit was placed close to the window to enjoy good channel conditions to the base station. The relay access unit which connects to the user equipment, was positioned just 1 m away from the feeder unit and used two isotropic antennas at a height of 3 m. An ethernet cable connects the relay feeder and access units. The isolation between the relay feeder and access for 1 m separation at the measurement site and with additional shielding was only about 20 dB. Because of this low isolation, orthogonalization between the two relay units is needed. A simple relay time split strategy was employed for this purpose. The base station to relay link was active for a fixed half of the time slots, and the relay to user equipment was active for remaining time slots within a radio frame of 10 ms.

The following downlink measurements were performed. The direct outdoor to indoor measurement was conducted with the relay node manually switched off. The second measurement was made in the same office with the DF relay operating at 23 dBm transmit power.

The signal measurements comparing the RSRP from the relay to that from the outdoor base station is shown in Figure 2. The relay provides an SNR between 15 and 70 dB for indoor users depending on their location, whereas the SNR range is 5 to 40 dB from the macro base station.

Example numerical results

To show the effectiveness of our scheme, we present example numerical results based on the channels from the above coverage measurement. The description of the scenario used for showing the numerical results is as follows.

Description of network setting

We consider a simple cellular network scenario based on a scenario as depicted in Figure 4. The layout consists of a macro base station, and a set of indoor relay stations as shown. Each indoor relay station provides coverage to its respective small indoor cell. We assume that each of the indoor cells is exactly similar to that of sample test scenario we showed in "Channel measurement study" section.

Based on the above assumptions, the channel to an indoor user from the relay is taken from the measurement tests. The simulation methodology is as follows. A certain number of users are assumed in each relay cell

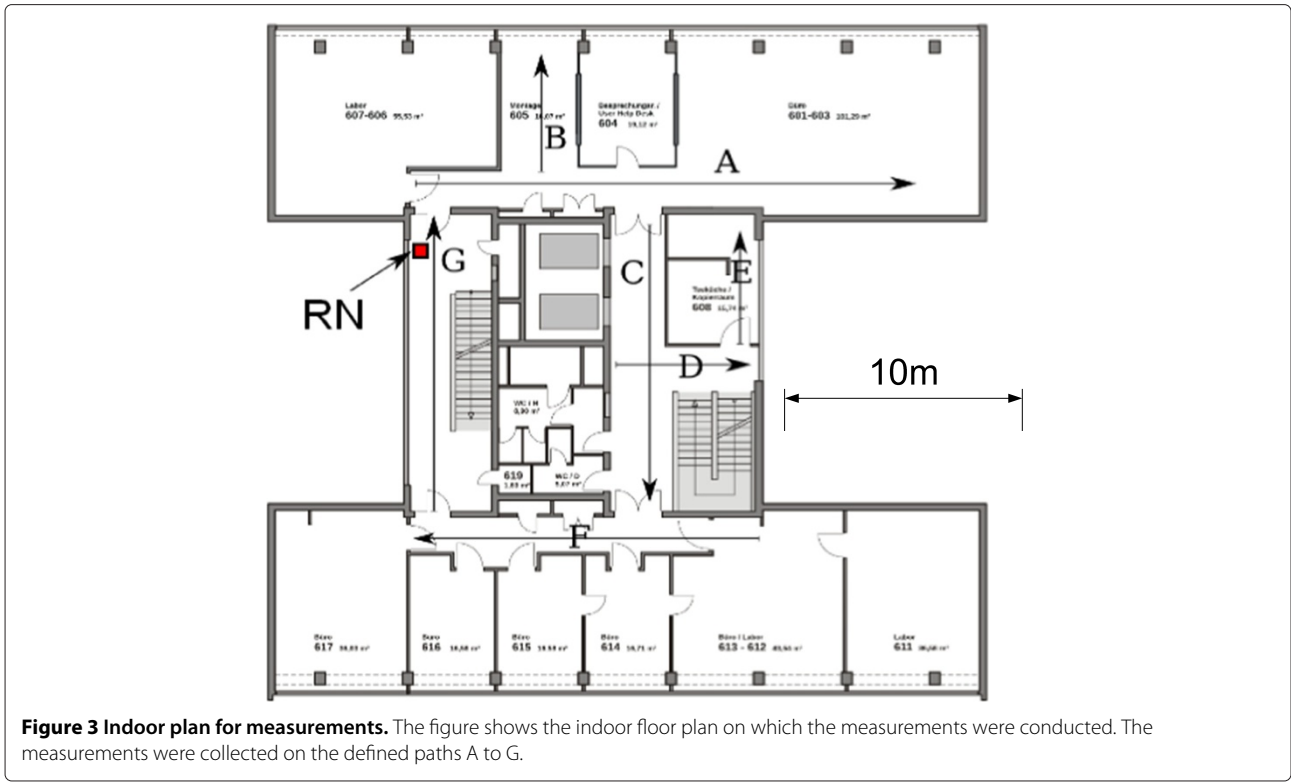


Figure 3 Indoor plan for measurements. The figure shows the indoor floor plan on which the measurements were conducted. The measurements were collected on the defined paths A to G.

and are distributed in random spatial positions indoors. The channel frequency responses at those different positions are known from the measurements and are thus used for numerical simulations. An important step is to decorrelate the channel realization of one indoor cell from another so that the channels are independent at a given time. This channel independence is realized by randomizing the user spatial positions in each of the relay cells independently.

In the layout, each indoor relay cell is located at a distance of 485 m from the macro base station. Thus, the feeder link spectral efficiency to all relay feeder units are assumed to be exactly similar and set to 10 bits per data symbol in a data subcarrier (for two MIMO streams). We expect high feeder spectral efficiency in practice, because our measurements from base station showed an SNR above 40 dB per spatial stream to be achievable around the location of the feeder unit in an interference-free scenario.

Neighboring indoor relays are assumed to be close to each other as shown in Figure 4, in which case they would interfere. As a consequence, the signal from one relay may affect the users of the neighboring relay as unwanted interference. This situation is depicted using an inter-relay distance of $d < d_{\min}$ between the relays, where d_{\min} is the distance for mutually interference free operation, e.g., 100 m.

We thus consider that the indoor cells are separated as two groups based on a static grouping discussed in “Relay

grouping” section. In Figure 4, the two groups are shown in two different colors—red and green, with a relay cell of dimension 20 m as in our test scenario. The figure shows a total of eight relays serving eight indoor cells. This static partitioning essentially means that any two neighboring relays will not be scheduled in the same dedicated frame.

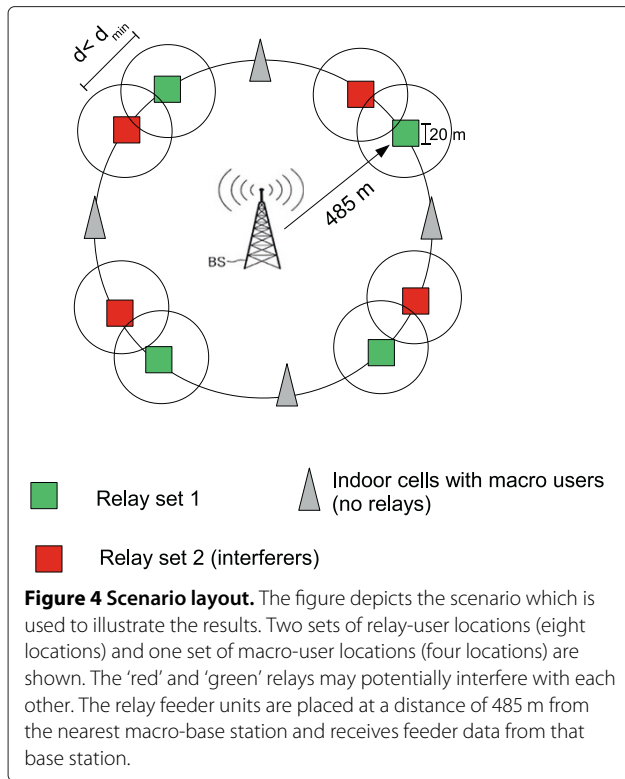
Figure 4 also shows users in four other indoor cells, who are served directly by the base station. These users do not have any installed relay in the vicinity and thus depend only on the macro base station for coverage.

Baseline comparison schemes

For baseline comparison to DF relaying in LTE, we refer to the scheme in [12] which uses a synchronized relay duplex time-sharing protocol. In this protocol, the feeder link to each relay is activated for a time duration fractions ρ_r sequentially. The feeder links use the full transmission bandwidth and finally all the relays synchronously transmit for a time duration α . Thus, $\sum_R \rho_r + \alpha = 1$.

In this baseline scheme, we assume that the relays just employ round-robin scheduling but optimized time durations instead of an optimized resource scheduler. The optimal duration for which all the relays are ‘on’ within a frame t is given as (using the same approach as in [12])

$$\epsilon(R) = \left[1 + \sum_{r=1}^{r=R_t} \frac{\sum_{k \in K_r} \sum_m u_{km}}{k_r M \gamma_r} \right]^{-1}. \quad (40)$$



Thus, a comparison to the baseline scheme is fair because the same downlink bandwidth as well as relay spatial reuse factor is employed in the baseline scheme and our scheme. To show the benefit of our novel admission control scheme, we follow the design principle of [12] (which they use for greedy rate maximization) and apply the selection of the number of relays in baseline scheme.

Thus, the following relay-based admission control scheme is used for the baseline scheme: The relays are enumerated in a random order by the base station and only as many relays that will maximize the number of guaranteed bit rates in (8) are admitted. We refer to this baseline scheme as simply 'Sync round-robin'.

The differences between the comparison scheme based on [12] and our work presented in this article are thus

- we do not limit the relays to be synchronous. Their duplexing durations are independent
- we optimize multi-user resource scheduling at relays
- we employ a hierarchical user-relay admission control instead of only relay admissions

We note that in both our proposed cell-GBRS scheme and the baseline 'Sync round-robin' scheme, inter-frame allocations can be utilized. For inter-frame allocations in 'Sync round-robin' scheme, the available extra bandwidth for the scheduler is used for both the feeder and access links. In doing so, the same relay feeder and access time

splits are applied to the extra bandwidth. The relays are further split into non-interfering relay sets as was done before. We also compare the performance of our proposed scheme in case cross-slot allocations are not made. We refer to this scheme just as 'Plain GBRS' because now the system resources are not being fully utilized.

Performance assessment

We present results for the cell scenario previously discussed as in Figure 4. The indoor 'macro users' are in four indoor cells as shown in the figure and have no relay in vicinity. We assume 6 indoor macro users per indoor cell, thereby totaling to 24 macro users. The relay users connect to a relay in one of the two relay sets; relay set 1 or relay set 2.

Relay set 2 consists of 4 relays, with each relay serving 6 users. Thus, there are a fixed total of 24 users in relay set 2.

For testing our scheme, we vary the other cell parameters as follows: (1) increase the number of relay set 1 users, (2) increase the number of macro users and observe the effect of it on relay set 1, (3) vary the data rate of just one of the users in relay set 1, and (4) increase the number of relay sets.

The access link employs frequency-dependent adaptive MCS. For this purpose, the user equipment estimates the downlink channel, decides the appropriate MCS for each resource block to achieve 10^{-2} packet error rate, and sends feedback in the uplink (using FDD). A resource block is taken to be 25 adjacent sub-carriers. This would correspond to 144 data symbols in a 0.5-ms time interval. The MCS set comprising of 26 MCS levels [16] (for $K = 1152$) which are designed for a target packet error rate of 10^{-2} . The same MCS level is applied to all data symbols in subcarriers within a resource block.

As part of feedback, the uplink also reports the precoding matrix index chosen from a set of three codebooks, and the number of streams to be sent per data symbol of a resource block.

Our primary objective in scheduling is to fulfill as many user bit rate guarantees as possible to all relay and macro users.

Increasing number of relay users

Figure 5 shows the performance results for increasing number of relay users in a relay cell. Table 2 summarizes the cell scenario parameters used for the results in Figure 5. We remark that user data rate requests of 500 kbps have been assumed keeping in mind the recommended value for satisfactory Youtube video streaming [23]. On the other hand, Youtube also provides high-quality videos which would require a higher 1 Mbps data rates. Thus, we assume that relay set 1 users request for this higher data rate to test the performance of our scheme. Figure 5 shows the plots when the number of

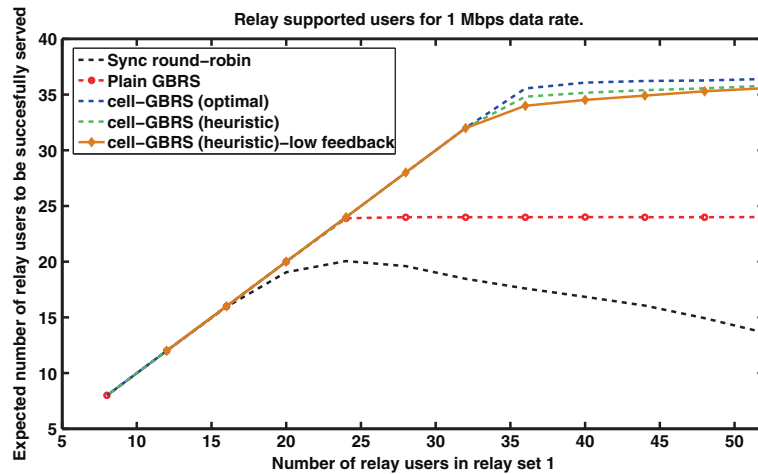


Figure 5 Performance for increasing number of relay users. The figure compares the performance of the cell-GBRS scheme to baseline relaying scheme for increasing number of relay users. The number of users in relay set 1 are assumed to increase. Relay set 2 has 24 users and there are 24 macro-users in the cell.

relay users in relay set 1 (ref Figure 4) is increased from 2 to 13 per relay. The y -axis shows the expected number of relay set 1 users who are successfully supported upon averaging over 1,000 realizations.

Thus, the total number of users in the cell (macro+2 relay sets) vary from 54 to 100. A 30% overhead would be incurred for downlink signaling. In addition, the frame durations are split evenly between the three user types: relay set 1, relay set 2, and macro. Taking the overhead and frame splits into account, we thus target a bit loading corresponding to 2304 bits in a 0.5-ms time slot for providing 1 Mbps data rate.

We compare our proposed cell-GBRS scheme to baseline ‘Sync round-robin’ scheme. As can be observed in Figure 5, the performance of ‘Sync round-robin’ scheme drops down steadily as the number of users per relay increase beyond a certain limit of users, which is 24 requesting users in this case. The steady degradation can be understood by the shortcoming of admission control in baseline scheme. The main issue is that based on relay-level admissions, *all* users in those relays are served without user level admission control. This effectively means that the round-robin access link from relays to some of those users quickly become a limiting factor. The peak

performance of baseline is 19 users when there are 24 requesting users.

In the ‘plain GBRs’ scheme, even though the access link is well optimized using the RRA subroutine, the lack of feeder link resources becomes a limiting factor. It can be noted that this limitation occurs rather quickly, for just around 24 users. However, the scheme is comparatively more robust (shows a performance saturation) as compared to the baseline because any additional user requests are simply dropped based on a hierarchical user-relay admission control.

Our proposed ‘cell-GBRS’ scheme overcomes the above-mentioned drawbacks by using optimized relay scheduling, hierarchical admissions, and inter-frame allocations for feeder link. RRA subroutine performs resource allocation for a much improved access link efficiency. Hierarchical admissions are done for user and relay selection. Free resources available after the ‘sort and greed’ macro-link scheduler are now allocated to the feeder link, which are called as inter-frame allocations. Thus, we observe a maximum of 36 users as compared to 24 users of the ‘plain GBRs’ scheme, which is a 50% performance improvement because of inter-frame allocations. The improvement is 89.47% over the baseline ‘Sync round-robin’ scheduling scheme.

Importantly, we observe that this performance benefit is obtained even while avoiding interference in the inter-frame allocations. This interference avoidance is done via silencing the base station feeder link to relay set 1 when any relay in relay set 2 starts its transmission.

It is noted that the 24 macro-users and 24 relay set 2 users were most often successfully supported served by the scheduler for their requested data rate of 500 kbps

Table 2 Increasing relay user scenario parameters for the results

User type	Number of users	Data rate per user
Macro	24	500 kbps
Relay set 2	24	500 kbps
Relay set 1	8 to 52	1 Mbps

Table 3 Increasing macro user scenario parameters for the results

User type	Number of users	Data rate per user
Macro	8–24	500 kbps
Relay set 2	24	500 kbps
Relay set 1	52	1 Mbps

which is lower than the 1 Mbps requested by relay set 1 users.

Feedback rate

In this section, we discuss the issue of channel quality feedback overhead. Channel feedback conveys the following: 1 of 26 adaptive modulation and coding levels, the number of MIMO streams, and the best precoding matrix. Overall 12 feedback bits would be needed per resource block to convey the above information to the base station or relay transmitter. This corresponds to 576 bits to be sent by each user equipment within an arbitrary channel coherence interval. Even assuming a relaxed 10 ms channel update interval and just ten users in the macro-cell, this would incur at least 576 kbps of uplink data rate per user (even without robust error protection coding). Thus, we see a situation of users in a cell requiring more than a 500-kbps uplink to enjoy a 500-kbps downlink.

We now consider an alternative reduced feedback scheme, similar to the extended feedback scheme which we presented in [5]. In this scheme, we now use only 4 MCS levels out of the 26 levels in [16]. The MCS levels indicate one of the following four options per spatial stream: 64QAM rate $\frac{24}{26}$, 16QAM rate $\frac{24}{30}$, QPSK rate $\frac{24}{30}$, and zero loading. No other coding options are considered, but three precoding options are still used. Furthermore,

only one MCS feedback is given for every six adjacent resource blocks (called a feedback block or a resource group block). The MCS of a feedback block is the minimum of the MCS levels of the six constituent resource blocks. Thus, the number of feedback bits are now only 48 bits per channel coherence interval, which is a reduction of 91.7%.

Strikingly, our results in Figure 5 show that the performance sacrifice of cell-GBRS relaying is within 5% for this 91.7% reduction of feedback. This result can be understood from the signal measurement results in Figure 2, where a high SNR above 30 dB is observed in many indoor locations. It can thus be expected that high SNR users do not require a high level of uplink feedback to achieve high spectral efficiency. Importantly, this means they can also be served on any resource block with a high spectral efficiency. The net effect is a high scheduling degree of freedom that can be exploited for other medium to low SNR users based on frequency diversity (refer Step 3 in section RRA-suboptimal method). Thus, the resource blocks are more often well utilized for the few medium to low SNR indoor users.

Trade-off with macro users

We now show results when the number of macro users increase as summarized in Table 3.

It can be deduced that as the number of macro-users increase the feeder bandwidth available for the relays would decrease. Thus, more efficient the macro-scheduler, better it is for the feeder link data rate. This effect can be observed in our results.

The results are shown in Figure 6, where we see the number of supported relay users of relay set 1 versus the number of macro users. For obtaining these results, relay set 2 users are fully served on their own frame and we

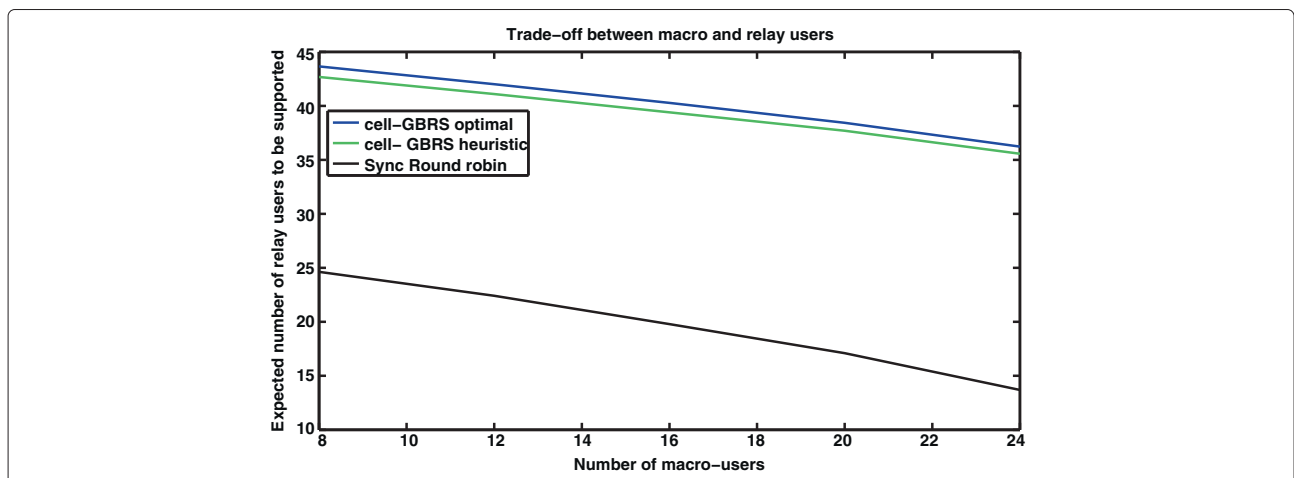


Figure 6 Trade-off between relay and macro-users. The figure shows performance of the cell-GBRS scheme when the number of macro-users vary. There are 52 relay users in relay set 1 and 24 users in relay set 2. All the relay set 2 and macro users are supported. Results show the number of relay set 1 users who are supported against the number of macro-users in that cell.

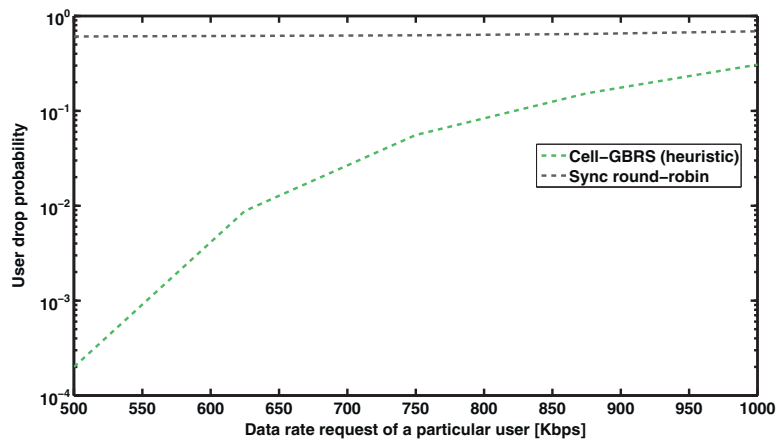


Figure 7 User drop probability versus the data rate. The figure plots the chances of a user getting dropped against the requested data rate. There are 52 relay set 1 users, 24 relay set 2 users and 24 macro-users. Only one user in relay set 1 is assumed to vary the data rate request, whereas all the other user requests remain unchanged. The results are shown for the varying user.

also do not consider any free bandwidth available from relay set 2 frames. It is seen that both the cell-GBRS and the baseline schemes benefit from additional feeder bandwidth. When there are only 8 macro-users in the cell, 42 out of the 52 relay users in relay set 1 are supported successfully. In this scenario, we again see that cell-GBRS consistently outperforms the baseline scheme.

User drop versus the requested data rate

We now consider a different scenario as follows. Assume that relay set 2 has 24 users being served at 500 kbps. There are 24 indoor macro-users being served at the same data rate 500 kbps and relay set 1 has 52 users. We now generate results when only one user in relay set 1 lowers the data rate request, while the other 51 users request 1 Mbps as usual.

Figure 7 shows the results, where we show the user drop probability versus the data rate. Five thousand channel realizations were used to generate these results. A user is defined to be dropped in two cases: (a) when the admission control does not admit the relay or user or (b) when the user is admitted but the requested data rate could not be served. Ideally, one would expect that a good scheduling scheme should support the user who lowers the data rate with higher probability. In Figure 7, we see that our scheduling scheme demonstrates this behavior quite well. In contrast, the round-robin scheme with relay admissions neglects the adaptiveness of the user. This is mainly because a relay is altogether rejected by the base station in relay-based admissions, if most other users of that relay request a higher data rate. This relay dropping would also happen when the users of another relay experience

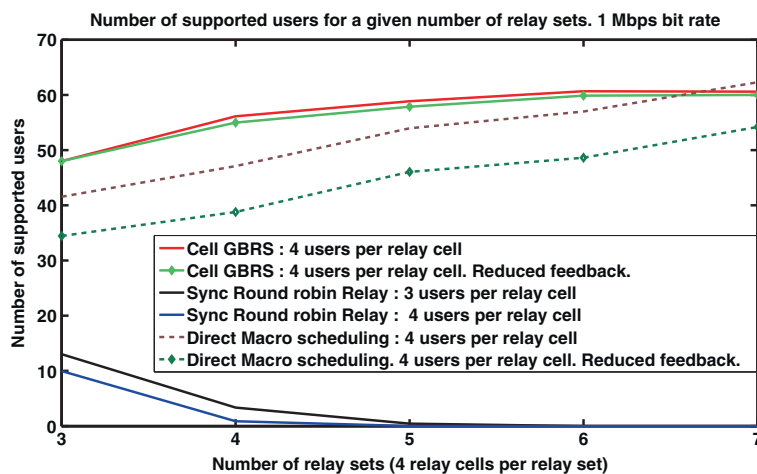


Figure 8 Increasing number of relay sets. The figure shows the performance when the number of relay sets increase. There are 16 users per relay set with each user requesting 1 Mbps data rate.

better channel conditions and therefore the other relay gets preferably admitted.

Increasing the number of relay sets

We now show results in Figure 8 when the number of relay sets increase. For this scenario, we assume there are four relay cells per relay set and four users per relay cell. The requested data rate from each user is 1 Mbps.

As the number of relay sets go up, the effective data rate required per user in a dedicated frame proportionally goes up. This proportional increase is because we have enforced equal frame split durations. Thus for four relay sets and one set of macro users, the effective data rate per user in a frame becomes 5 Mbps in order to realize 1 Mbps after time averaging. Accordingly, we observe that the performance of baseline 'Sync round robin' relaying suffers significantly because the round-robin relay access link struggles to support high effective data rates. As an indication of this access link limitation, this scheme shows better performance for a scenario with three users per relay cell as compared to four users per relay cell as seen in Figure 8.

We further note that the effect of increasing the relay sets (and hence frame time splits) would be in fact similar to increasing the required data rate for users assuming only one relay set. Thus, we see that GBRS relaying is in general more efficient than the baseline and can adapt to both the situations.

We also show the performance of direct macro link for the case of full feedback and reduced feedback (i.e., 91.7% reduction in feedback). In the direct macro link case, there are no relay nodes assumed for any indoor cell. Thus, all the indoor users in the relay cells in this case are assumed to connect directly to macro base station. The users are scheduled according to the 'Sort and Greed' scheduler in "Sort and greed scheduler" section. We see that direct macro link especially underperforms for low number of relay sets as compared to our proposed cell-GBRS scheme. The relative performance to relaying however markedly improves for high number of relay sets. In fact for the full feedback case the macro performance is even higher, with 62 users supported on average with macro-only connection as compared to 60 with cell-GBRS relaying.

This high performance of macro-only case when there are high number of relay sets is caused by the combinatorial effect of admission control and multi-user diversity. In the case of cell-GBRS relaying, even though it is more resource efficient and the combinatorial effect of admission control and multi-user diversity still exists, the lack of feeder link resources becomes a limitation for high number of relay sets. Thus, we see that the macro-only and cell-GBRS cases converge as the number of relay sets increase.

Again we observe that cell-GBRS suffers only negligible performance loss when the feedback is reduced by 91.7% as in "Feedback rate" section. In comparison, the effect of feedback reduction on the macro-only scheduler is significant. For 3 relay sets, the number of users is down from 42 to 34 users, which is a 19% performance loss when the feedback is reduced by 91.7%. For 3 relay sets, cell-GBRS relaying shows a 41.18% gain w.r.t macro-only scheduling using the same low amount of feedback.

Conclusion

The article presented a novel scheduler called cell-guaranteed bit rate by relay scheduling (cell-GBRS) for fair resource allocation to macro-base station users and relay users in MIMO-OFDMA cellular networks. The scheduler then supports as many users as possible for each dedicated category of users, denoted as macro and sets of relay users. The proposed scheduler consists of two sub-schedulers: the so-called 'sort and greed' scheduler for the macro-users and GBRS scheduling for the relay users. Both the schedulers independently minimize the bandwidth consumed for satisfying their user demands. The available bandwidth after dedicated scheduling is utilized by the base station through opportunistic inter-frame scheduling, wherein the bandwidth is assigned to other feeder links and macro-users. The scheme further manages interference situations, such as between interfering relays and between relays and macro-users, by means of orthogonalized time slots. Our optimized results in a test scenario shows that while 48 macro and relay users can be supported at a time for 500 kbps data rate, another set of 36 relay users can be supported for 1 Mbps data rate. We find that a key benefit of relays is also to significantly reduce the uplink feedback. Results with indoor relay show that the performance loss is less than 5% even for a 91.7% reduction of feedback.

Competing interests

The authors were at Nokia Siemens networks, Munich until August 2008. The authors declare that they have no other competing interests.

Received: 5 April 2012 Accepted: 13 September 2012

Published: 4 October 2012

References

1. R Pabst, BH Walke, DC Schultz, P Herhold, H Yanikomeroglu, S Mukherjee, H Viswanathan, M Lott, W Zirwas, M Dohler, H Aghvami, DD Falconer, GP Fettweis, Relay based deployment concepts for wireless and mobile broadband radio. *IEEE Commun. Mag.* **42**(9), 80–89 (2004). doi:10.1109/MCOM.2004.1336724
2. R Schoenen, R Halfmann, BH Walke, An FDD multihop cellular network for 3GPP-LTE. in *Proceedings of IEEE VTC 2008*. (Singapore, 11–14 May 2008), pp. 1990–1994
3. T Isotalo, P Lähdekorpi, J Lempinen, Improving HSDPA indoor coverage and throughput by repeater and dedicated indoor system. *EURASIP J. Wirel. Commun. Netw.* **2008**(Article ID951481), 1–11 (2008). doi:10.1155/2008/951481
4. T Haustein, A Forck, H Gabler, V Jungnickel, S Schiffermuller, Real-time signal processing for multi-antenna systems: algorithms, optimization and

- implementation on an experimental test-bed. *EURASIP Implem Aspects* (Special Issue). **2006**(Article ID 27573), 1–21 (2006).
doi:10.1155/ASP/2006/27573
5. V Venkatkumar, T Hausteine, M Faulkner, Relaying results for indoor coverage in long-term evolution and beyond. *Eur. Trans. Telecommun.* **21**(8), 770–779 (2010). doi:10.1002/ett.1450
 6. Improving 4G coverage and capacity indoors and at hotspots with LTE femtocells, Nokia Siemens Networks Whitepaper (2011). Found as *LTE_Femto_WP_110204_online.pdf*
 7. Z Yuanyuan, Boosting WiMAX indoor coverage. *Huawei Commun.* **Issue 48**, 31–32 (April 2009)
 8. ITU-RM. 2134, Requirements related to technical performance for IMT-Advanced radio interface(s), 1–8 (November 2008)
 9. 3GPP TR 36.913 v 8.0.0, Technical Specification Group Radio Access Network Requirements for Further Advancements for E-UTRA, LTE-Advanced (Release 8), (June 2008)
 10. IST-4-027756 WINNER II, Final assessment of relaying concepts for all CGs scenarios under consideration of related WINNER L1 and L2 protocol functions. D3.5.3 v1.0, (30 September 2007)
 11. K Haneda, V Kolmonen, T Riihonen, Evaluation of relay transmission in outdoor-to-indoor propagation channels. in *European Cooperation in Science and Technology Workshop (COST) 2100*. Trondheim, Norway, 3–4 June 2008, Paper ID W08114
 12. A Saleh, S Redana, B Raaf, T Riihonen, J Hamalainen, R Wichman, Performance of amplify-and-forward and decode-and-forward in LTE-advanced. in *Proceedings of IEEE VTC Fall 2009*. Anchorage, USA, 20–23 September 2009, pp. 1–5
 13. M Kaneko, P Popovski, K Hayashi, Throughput-guaranteed resource-allocation algorithms for relay-aided cellular OFDMA system. *IEEE Trans. Veh. Technol.* **58**(4), 1951–1964 (2009)
 14. M Salem, A Adinoyi, M Rahman, H Yanikomeroglu, D Falconer, Y Kim, Fairness-aware radio resource management in downlink OFDMA cellular relay networks. *IEEE Trans. Wirel. Commun.* **9**(5), 1628–1639 (2010, IEEE Press Piscataway, NJ, USA)
 15. V Venkatkumar, T Hausteine, Multi-user relaying with full frequency reuse for enhanced LTE-GBR coverage. in *Proceedings of European wireless 2011*: Vienna, Austria, 27–29 April 2011, pp. 380–387
 16. S Stiglmayr, M Bossert, E Costa, Adaptive coding and modulation in ofdm systems using BICM and rate-compatible punctured codes. in *Proceedings of European Wireless 2008*. Paris, France, 1–4 April 2007, Invited paper
 17. 3GPP TR 36.814 v 9.0.0, Technical Specification Group Radio Access Network Requirements for Further Advancements of E-UTRA Physical layer aspects, Release 9 (March 2010)
 18. V Chvatal, A greedy heuristic for the set-covering problem. *Math. Oper. Res.* **4**(3), 233–235 (1979)
 19. Y Zhang, *Solving large-scale linear programs by interior-point methods under the MATLAB environment*: Department of Mathematics and Statistics, University of Maryland Baltimore County, 1996
 20. 3GPP TR 36.213 v 10.5.0, Technical Specification Group, Evolved Universal Terrestrial Radio Access, E-UTRA Physical layer procedures, Release 10, (March 2012)
 21. V Venkatkumar, T Hausteine, M Faulkner, Joint admission control and resource allocation for multiuser loading in LTE networks. in *Conference on Smart Spaces (ruSMART)*. vol. 6294, Saint Peterburg, Russia, August 2010, pp. 421–435, doi:10.1007/978-3-642-14891-0_37
 22. S Boyd, L Vandenberghe, *Convex Optimisation*. Cambridge, MA: Cambridge University Press, 2004)
 23. What system requirements are needed to watch videos on YouTube?
<http://support.google.com/youtube/bin/answer.py?hl=en&answer=78358>

doi:10.1186/1687-1499-2012-309

Cite this article as: Venkatasubramanian and Hausteine: A novel scheduling framework for QoS-aware OFDMA resource allocation in a network with small relay cells and macro users. *EURASIP Journal on Wireless Communications and Networking* 2012 **2012**:309.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com