ROUND-ROBIN SCHEDULING FOR TIME-VARYING CHANNELS WITH LIMITED FEEDBACK

Claude Simon and Geert Leus

TU Delft, Fac. EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands

ABSTRACT

We investigate round-robin scheduling for time-varying channels with a strict SINR constraint and limited feedback. The presented algorithm aims at scheduling the users at identical slots over different blocks. The algorithm allocates the users using orthogonal beamforming based on the quantized feedback provided by the users. The feedback consists of the estimated energy that is necessary to fulfill the SINR constraint. In order to fulfill the SINR constraint, the available transmit energy is distributed between the users based on their feedback. The performance of the algorithm is demonstrated through simulations.

1. INTRODUCTION

One of the challenging problems of multi-user schemes is scheduling. The scheduling algorithm should have a low complexity, but the transmissions to the users must still fulfill strict Quality-of-Service (QoS) constraints. An important QoS constraint is the minimum Signal-to-Interference-plus-Noise Ratio (SINR). The minimum SINR constraint requires that every scheduled user in the cell has an SINR larger than a predefined threshold. We further try to schedule the users in a round-robin fashion, i.e., the delay between two transmissions to the same user should remain constant. Especially modern voice and video communication systems have strict delay requirements. Another advantage of round-robin scheduling is the reduced overhead caused by the scheduling. The base station does not need to sacrifice transmission time to inform the users in every block about their allocated slot positions. However, due to the stochastic nature of the wireless channel, it is not possible to provide hard QoS guarantees.

An algorithm that takes the time-varying nature of the channel into account, but still provides strong bounds on the maximum delay, is the Channel Aware Round Robin (CARR) scheduling algorithm [1]. It schedules every user once inside each block, but it does not implement true round-robin scheduling, since the position of the users inside the block might change. The CARR algorithm chooses the positions depending on the channel state of the different users in the different slots. A similar scheduling algorithm for SDMA is the Best Fit algorithm [2], i.e., the Best Fit algorithm also tries to assign all the users in every block. However, it uses SDMA to dynamically assign multiple users to the same slot depending on the resulting SINR. It further considers a SINR constraint. A low-complexity variant is the partial Best Fit (PBF) algorithm [3]. It just adds new users according to the Best Fit strategy and removes the expired users, i.e., the users that have no more packets to transmit. An overview of other algorithms that consider scheduling under the exploitation of the spatial diversity can be found in [4].

Our proposed algorithm tries to schedule all the users in the cell in a round-robin fashion as long as possible. The application of orthogonal beamforming allows to reduce the necessary feedback to the base station. We propose a corresponding feedback metric, and we consider the necessary quantization due to the data-rate limited feedback link. The feedback is used by the base station to dynamically divide the available transmit power among the users. This allows the weakest users, i.e., the users with the worst channel conditions, to fulfill the SINR constraint longer than with equal power distribution. A user is rescheduled if he is no longer able to fulfill the SINR constraint despite receiving additional power, i.e., the user is scheduled at a different slot and with a different beamformer in the next block. The performance of the algorithm is depicted through simulations for a time-varying channel.

Notation: We use capital boldface letters to denote matrices, e.g., $\mathbf{A}$, and small boldface letters to denote vectors, e.g., $\mathbf{a}$. $E(\cdot)$ denotes expectation, and $P(\cdot)$ probability. $\lceil x \rceil$ rounds x up the next largest integer.

2. SYSTEM MODEL

We assume a narrowband single-cell scenario where a base station with M antennas transmits data to N single-antenna users. At a given time user i receives the symbol

$$y_i = \sum_{j \in \mathcal{S}} \mathbf{h}_i \mathbf{w}_{g(j)} \sqrt{E_j} s_j + n_i \qquad (1)$$

where $\mathcal{S}$ contains the indices of the users scheduled at that time instant, $\mathbf{h}_i \in \mathbb{C}^{1 \times M}$ is the channel of user i, and $\mathbf{w}_{g(j)} \in$

This research was supported in part by NWO-STW under the VIDI program (DTC.6577).

$\mathbb{C}^{M\times 1}$ is the beamforming vector assigned to user j. The mapping $g(j)$ maps a beamforming vector from the beamformer codebook $\mathcal{W}$ to every user. The energy assigned to user j is denoted E_j, and the data symbol s_j, that is transmitted to user j, is selected from a constellation with average unit energy. The noise n_i is complex Gaussian distributed with zero mean and variance N_0, i.e., $n_i \sim \mathcal{CN}(0, N_0)$. The SNR for user i is $\rho_i = \frac{E_i}{N_0}$, and the total allocated transmit energy is limited to $E_T = \sum_{i\in\mathcal{S}} E_i$.

The users have the possibility to feed back information to the base station at the start of every block. The feedback link itself is instantaneous, error-free, and data-rate limited to B bits. All the users have to fulfill a strict SINR constraint denoted SINR_{cut}.

We assume that the individual users acquire perfect channel state information (CSI) at the start of each block through training. Every block consists of K slots. We assume that the channel is block-fading, i.e., the channel is constant throughout the K slots of a block. The block index k starts at $k = 0$, and the slot index l restarts at the beginning of each new block at $l = 0$. Thus, the relation between the current time instant t and the current block/slot index is $t = kK + l$.

We are using a set of orthogonal beamforming vectors from a codebook $\mathcal{W}$ to simultaneously transmit to maximally M users [5]. The codebook $\mathcal{W}$ contains M orthogonal beamforming vectors $\mathbf{w}_m$. The M beamformers in the codebook all have unit norm, i.e., $\|\mathbf{w}_m\|_2 = 1, m \in \mathcal{M} = \{1, \ldots, M\}$. The codebook $\mathcal{W}$ is known to the users and to the base station.

3. PROBLEM DESCRIPTION

The main objective is to schedule the users in a round-robin fashion. If a user i has been scheduled at time instant $t = (k-1)K + l$ using the beamformer $\mathbf{w}_m$, then we want to schedule him also at time instant $t = kK+l$ using the beamformer $\mathbf{w}_m$. Further, all the scheduled users have to fulfill a strict SINR constraint, i.e., they need to have a SINR higher than SINR_{cut}.

Serving the users in a round-robin fashion should result in a packet delay variation of zero. However, due to the time-varying nature of the wireless channel, there is a non-zero probability that the channel is in a deep fade, i.e., reliable communication is not possible. Thus, it is not possible to guarantee the QoS constraints, i.e., to have hard QoS guarantees.

The problem is now to exploit the available feedback link to schedule the users as long as possible in a round-robin fashion while still fulfilling the SINR constraint.

4. ALGORITHM OVERVIEW

We assume that the different users are able to acquire perfect CSI at the beginning of each block, i.e., $t = kK, \forall k$. Due to the block-fading nature of the channel the individual user thus has perfect channel knowledge for every slot in the block. Using this channel knowledge the user then calculates how much energy he needs in order to reach the SINR constraint. Next, this minimum energy is quantized and fed back to the base station. Once the base station receives all the feedback from the users, it checks for every time slot if the sum of the fed back quantized minimum energies exceeds the maximally allocatable transmit energy E_T at the base station. If the sum is lower, then every user is assigned its fed back minimum energy. The remaining transmit energy is equally distributed over all the scheduled users. However, if the sum is higher then it is not possible to schedule all the users. The users with the highest energy demands are dropped and the transmit energy is distributed over the remaining users. The dropped users are assigned to new time slots and are assigned new beamforming vectors according to Section 4.4. The case where a new user enters the cell is treated identical.

4.1. Feeding Back the Required Energy

The SINR for user i is calculated as

$$\text{SINR}_i = \frac{|\mathbf{h}_i \mathbf{w}_{g(i)}|^2 E_i}{\sum_{j\in\mathcal{S}\setminus\{i\}} |\mathbf{h}_i \mathbf{w}_{g(j)}|^2 E_j + N_0}. \quad (2)$$

We see that the SINR of a user i depends on the individual transmit energies of the users in the set $\mathcal{S}$.

In order to determine the minimum amount of energy that is required by a user i to reach SINR_{cut}, it is necessary to know the amount of energy that is assigned to the other users scheduled in the same slot. However, the individual energy levels assigned to the other users in the set are not known to the individual users. A solution is to feedback the full CSI to the base station and to balance the SINR between the different users using the algorithm in [6]. The drawback is that it requires full channel knowledge or at least knowledge of the composite channel energies $|\mathbf{h}_i \mathbf{w}_{g(j)}|^2, j \in \mathcal{S}$ at the base station and thus incorporates a lot of feedback.

In this paper, we try to find an estimate of the energy that has to be assigned to a user i so that the SINR constraint is fulfilled, and that does not depend on the energy levels assigned to the other users in the set $\mathcal{S}$. This minimum energy will be denoted $\hat{E}_i$. We start by defining an estimate of the true SINR, denoted $\hat{\text{SINR}}_i$, that does not depend on how the total transmit energy is distributed over the users in $\mathcal{S}$, but is guaranteed to be smaller than the true SINR

$$\hat{\text{SINR}}_i \leq \text{SINR}_i. \quad (3)$$

Due to (3), it is certain that if the estimated SINR fulfills the SINR constraint, so does the true SINR, i.e., if $\text{SINR}_{\text{cut}} \leq \hat{\text{SINR}}_i \Rightarrow \text{SINR}_{\text{cut}} \leq \text{SINR}_i$. We propose to use

$$\hat{\text{SINR}}_i = \frac{|\mathbf{h}_i \mathbf{w}_{g(i)}|^2 E_i}{\max_{j\in\mathcal{M}\setminus\{g(i)\}} |\mathbf{h}_i \mathbf{w}_j|^2 (E_T - E_i) + N_0}. \quad (4)$$

The proposed estimate $\hat{\text{SINR}}_i$ is lower than or equal to the real SINR since inserting (4) and (2) into (3) results in

$$\sum_{j \in \mathcal{S} \setminus \{i\}} |\mathbf{h}_i \mathbf{w}_{g(j)}|^2 E_j \leq \max_{j \in \mathcal{M} \setminus \{g(i)\}} |\mathbf{h}_i \mathbf{w}_j|^2 (E_T - E_i) \quad (5)$$

which is always true. The minimum energy assigned to a user i that is sufficient to fulfill the SINR constraint can thus be calculated as

$$\hat{E}_i = \frac{\max_{j \in \mathcal{M} \setminus \{g(i)\}} |\mathbf{h}_i \mathbf{w}_j|^2 E_T + N_0}{\frac{1}{\hat{\text{SINR}}_{i,\text{cut}}} |\mathbf{h}_i \mathbf{w}_{g(i)}|^2 + \max_{j \in \mathcal{M} \setminus \{g(i)\}} |\mathbf{h}_i \mathbf{w}_j|^2}. \quad (6)$$

Finally, the energy $\hat{E}_i$ is quantized and fed back to the base station.

4.2. Quantizing the Feedback

The data rate limitation on the feedback link makes a quantization of the minimum energy necessary before it can be fed back to the base station. The quantization Q maps the minimum energy $\hat{E}_i$ to an element c_q of a predefined codebook $\mathcal{C} = \{c_1, \ldots, c_b\}$, i.e., $Q : \mathbb{R} \rightarrow \mathcal{C}$. We assume that the codebook size is limited to $b = 2^B$ entries in order to fulfill the data-rate limitation of the feedback link. It is important that the quantized minimum energy is as close as possible to the unquantized minimum energy in order to minimize the quantization error. However, the quantized minimum energy also has to be larger than the unquantized minimum energy. If the quantized minimum energy would be smaller than the unquantized minimum energy, then the base station would wrongly underestimate the required energy of that user. Thus, the codebook must contain an element that is higher than all possible unquantized minimum energy levels, i.e., $c_b = +\infty$. The minimum energy is quantized using

$$Q(\hat{E}_i) = \arg\min_{c_q \in \mathcal{C}} \; c_q - \hat{E}_i \quad \text{s.t.} \quad \hat{E}_i \leq c_q \quad (7)$$

and the index q of the element $c_q = Q(\hat{E}_i)$ of the codebook $\mathcal{C}$ is fed back to the base station. The quantized minimum energy of user i is denoted $\hat{E}_{Q,i} = Q(\hat{E}_i)$.

The necessary feedback can be reduced by using entropy coding [7]. In entropy coding, short codewords are used to feed back the highly probable indices, in order to reduce the required average feedback.

We can also exploit the time-correlation of the channel to better use the data-rate limited feedback link. A simple approach is predictive quantization [7] where we simply use the last quantized minimum energy at time instant $t - K$ as the predicted current minimum energy at time instant t

$$Q(\hat{E}_i[t]) = \arg\min_{c_q \in \mathcal{C}} \; c_q - \hat{E}_i[t] + \hat{E}_{Q,i}[t-K] \quad (8)$$

$$\text{s.t.} \quad \hat{E}_i[t] \leq \hat{E}_{Q,i}[t-K] + c_q. \quad (9)$$

Predictive quantization requires the use of two quantization codebooks. The first codebook $\mathcal{C}_{\text{init}}$ is used to quantize the initial minimum energy when no prediction is possible, i.e., the user is scheduled for the first time or the user is rescheduled. The second codebook $\mathcal{C}$ is used afterwards to quantize the prediction difference (8).

The codebooks for both approaches can be designed using simple Monte-Carlo codebook generation [7].

4.3. Scheduling

We assume that the base station receives the feedback from all the users instantaneously. For every slot in the block the base station calculates the sum of the minimum energies. If the sum is negative then it is not possible to schedule the users so that they all fulfill the SINR constraint while still using the same slot and the same beamformer as in the last block. The straightforward approach is to remove the user from the set $\mathcal{S}$ that has the highest minimum energy, i.e., the highest energy demand. This is repeated until the sum of the minimum energies of the remaining users is smaller than the available transmit energy E_T. Every user is assigned its required minimum energy, and the remaining transmit energy is uniformly distributed over the remaining users. This scheduling algorithm is denoted Algorithm 1. Note that this approach tries to balance the SINRs of the different users in the same slot. However, due to the quantization, and due to the unknown interference between the users in the set $\mathcal{S}$, it is not possible to truly balance the SINRs as it is possible with full CSI at the base station [6].

Algorithm 1 Scheduling algorithm for a random block k

```
for l = 0 to K − 1 do
    t := kK + l
    S[t] := S[t − K]
    while Σ_{i∈S[t]} Ê_{Q,i}[t] > E_T do
        i := arg max_{i∈S[t]} Ê_{Q,i}[t]
        S[t] := S[t] \ {i}
        U_resched := U_resched ∪ {i}
    end while
    E_i[t] := Ê_{Q,i}[t] + (E_T − Σ_{i∈S[t]} Ê_{Q,i}[t]) / |S[t]|
end for
```

4.4. Rescheduling

The users in the set $\mathcal{U}_{\text{resched}}$ have not yet been scheduled. They are either new users, i.e., users who just entered the cell, or users that have not been able to reach the SINR constraint in the previous block and are now getting rescheduled. Those users have to feed back their preferred beamforming vector to the base station besides their minimum energy. The base station then tries to successively schedule all the users in $\mathcal{U}_{\text{resched}}$ according to their fed back minimum energy, i.e., the users

with a large energy requirement are scheduled first. The base station then looks for a slot that has not yet a user assigned for the preferred beamformer of the considered user. For every one of these free slots the base station calculates the sum of the minimum energy levels of the users in the slot, assuming the considered user is added, and finally chooses the slot that results in the lowest sum. If the base station does not find a valid slot, i.e., the resulting sum of the minimum energies is smaller than the overall allocatable energy E_T, then the user is skipped for the current block. However, if the base station finds a slot then the user is scheduled for transmission.

5. SIMULATIONS

The simulation depicted in Fig. 1 shows how long $M = 3$ users can be scheduled together in a slot until one of them violates the SINR constraint as a function of the product of the Doppler frequency f_D and the block length T_f. We assume a homogenous cell where all the users experience SNR $=$ 15 dB. The SINR constraint is fixed to $\mathrm{SINR}_\mathrm{cut} = 5$ dB.

The time-correlation between the blocks is modeled according to Jakes' model [8]. At time instant t the pth element from the channel $\mathbf{h}_i[t]$ is modelled as

$$[\mathbf{h}_i[t]]_p = \frac{1}{\sqrt{Q}} \sum_{q=1}^{Q} a_{p,q} \exp[j\, 2\pi f_\mathrm{D} T_\mathrm{f} \left\lceil \frac{t}{K} \right\rceil \cos \alpha_{p,q}] \quad (10)$$

where Q is the number of scatterers, $a_{p,q}$ is i.i.d. complex Gaussian distributed with zero mean and variance 1, and $\alpha_{p,q}$ is uniformly distributed over $[0, 2\pi]$.

The influence of quantizing the fed back minimum energy is depicted for a codebook with 2 entries, for a codebook with 2^4 elements, and for no quantization. If the feedback link is limited to 1 bit, then the users can just feed back whether they can fulfill the SINR constraint or not, i.e., the codebook consists of the elements $\mathcal{C} = \{\frac{E_T}{M}, +\infty\}$. Power adaptation is not possible in this case. The 4-bit codebook is created using Monte-Carlo codebook design. We see that for slowly changing channels, i.e., channels with a low product of Doppler frequency f_D and block length duration T_f, the average number of consecutive blocks increases. Assuming no limit, the number of blocks would increase as $f_\mathrm{D} T_\mathrm{f}$ decreases. When $f_\mathrm{D} T_\mathrm{f} = 0$, then the channel is constant over time. However, if $f_\mathrm{D} T_\mathrm{f}$ increases, then the channel becomes more volatile. When $f_\mathrm{D} T_\mathrm{f} \geq 10^{-3}$, the channel becomes highly uncorrelated between different blocks. This increases $P(\sum_{i \in \mathcal{S}[t-K]} \hat{E}_i[t] > E_T \mid \sum_{i \in \mathcal{S}[t-K]} \hat{E}_i[t-K] \leq E_T)$, i.e., the probability that the users from the set $\mathcal{S}[t-K]$ cannot fulfill the SINR constraint in a slot of the current block, assuming they were able to fulfill the SINR constraint in the same slot of the previous block.

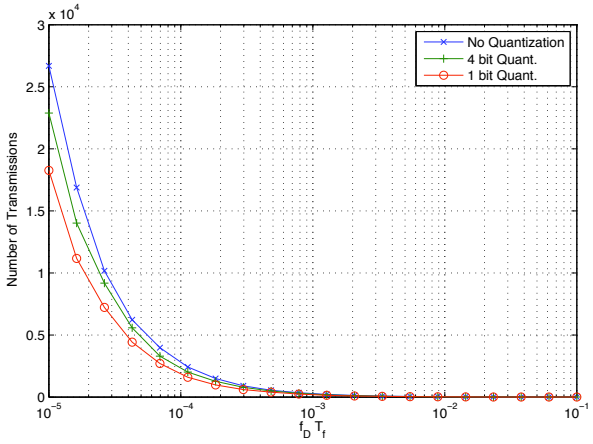

Fig. 1. Average number of transmissions that a slot is occupied by the same set of users. ($M = 3$, $Q = 30$, SNR $= 15$ dB, $\mathrm{SINR}_\mathrm{cut} = 5$ dB)

6. CONCLUSIONS

We presented a scheme to implement round-robin scheduling using orthogonal beamforming and data-rate limited feedback. The scheme uses scalar feedback from the users to divide the transmit energy amongst the users so that they all fulfill a given SINR constraint. The simulations show that the presented algorithm is attractive to implement round-robin scheduling for time-varying channels.

7. REFERENCES

[1] H.-Y. Wei and R. Izmailov, "Channel-aware soft bandwidth guarantee scheduling for wireless packet access," in *Proc. IEEE WCNC*, Atlanta, GA, USA, Mar. 2004, pp. 1276–1281.

[2] F. Shad, T. D. Tod, V. Kezys, and J. Litva, "Dynamic slot allocation (DSA) in indoor SDMA/TDMA using a smart antenna basestation," *IEEE/ACM Trans. Networking*, vol. 9, no. 1, pp. 69–81, Feb. 2001.

[3] H. Yin and H. Liu, "Performance of space-division multiple-access (SDMA) with scheduling," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 611–618, Oct. 2002.

[4] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems," *IEEE Network*, vol. 19, no. 5, pp. 43–48, Sept. 2005.

[5] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[6] W. Yang and G. Xu, "Optimal downlink power assignment for smart antenna systems," in *Proc. IEEE ICASSP*, vol. 6, Seattle, WA, USA, Apr. 1998, pp. 3337–3340.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compressing*. Kluwer Academic Publishers, 1995.

[8] W. C. Jakes, Jr., *Microwave Mobile Communications*. John Wiley & Sons, 1974.