Cooperative Scheduling of Downlink Beam Transmissions in a Cellular Network

Patrick Hosein
Huawei Technologies Co., Ltd.
10180 Telesis Court, Suite 365, San Diego, CA 92121, US
Tel: 858.882.0332, Fax: 858.882.0350, Email: phosein@huawei.com

Abstract—The next generation of wireless networks (e.g., [1]) will use multiple techniques to improve channel spectral efficiencies. In this paper we focus on one such technique, beamforming, in the downlink channel. In each sector a number of fixed beams are used to serve users. For each frame the scheduler chooses a beam and then schedules one or more mobiles that lie within that beam. Since the basestation’s power is focused on a narrow beam the served users can achieve higher rates and the sector coverage is also increased. However, this beam causes more interference in adjacent sectors. If a neighboring sector schedules a user that lies within the coverage of the beam then the achievable rate for that user will be low. The neighboring sectors must therefore schedule users outside of the beam. This requires coordination among the sectors. In this paper we assume that each sector coordinates with the sector that lies directly opposite. We provide a framework and show what information needs to be exchanged and how the schedulers in each base station uses the exchanged information to schedule users so that beam collisions are avoided. This must be done with the objectives of high spectral efficiencies and also user fairness. We compare our approach with a simple, non-coordinated approach to illustrate the advantages that coordination provides.

I. INTRODUCTION

The next generation of wireless networks will be based on the OFDMA (Orthogonal Frequency Division Multiple Access) radio transmission technology and will use MIMO (Multi-Input Multi-output) antenna arrays [2] to achieve increased spectral efficiency. One significant cost of increasing the number of antennas at the base station is the additional radio frequency (RF) chains needed. A low cost alternative is to install more antennas than RF chains and then, using a Butler matrix, dynamically switch which antennas are used for transmissions. This allows the use of one of a small set of narrow beams for each transmission and the most appropriate beam for each frame can be determined by the scheduler. However these beams also introduce more focused interference and hence beam coordination among neighboring sectors is needed. In this paper we address this problem and propose a semi-distributed algorithm for its solution.

We assume that multiple narrow beams (possibly together with a sector-wide beam) are used to serve mobiles in the downlink channel. The achievable SINR of each mobile depends on the signal strength of the serving beam as well as the interference due to beam transmissions in neighboring sectors. The interference experienced by a mobile may change dramatically from frame to frame because of the highly directional beams also being used in neighboring sectors [3]. Each mobile determines the Signal to Interference and Noise Ratio (SINR) for each frame transmission and reports this information in a Channel Quality Indicator (CQI).

Therefore, in order to ensure that the SINR achieved by a mobile when it is actually served is approximately equal to that it experienced when it made its CQI report then all beam allocations must be identical as at the point in time when the measurement was taken. Note that this only holds if each beam is served at the same power level. This implies that the set of beam allocations must be cyclic and the measurement and servicing instants must be separated by an integer number of periods of this cycle.

A simple round robin beam servicing approach can be used with beams in adjacent sectors chosen to avoid collisions. Such an approach has been used, for example, in [4]. However, this will provide uniform coverage of all beams within each sector and this may not always be desirable. For example, if there exists a cluster of users within a sector then the corresponding beam should be serviced more often. However such a change will require global re-optimization of the beam allocation cycles. Hence a centralized beam allocation algorithm is needed to ensure low interference as well as high channel efficiency. Even if users are uniformly distributed some may have higher QoS demands and so a uniform beam allocation strategy may not be optimal.

Centralized scheduling of beams is both computationally expensive and also requires significant backhaul resources for exchanging information. Therefore any distributed algorithm that depends on local information will perform poorly because of intercell interference while any centralized algorithm will either require too many resources or will be too simplistic resulting in poor spectral efficiency (e.g., if round robin beam allocation is used).

We therefore propose a semi-distributed algorithm that coordinates transmissions between sectors that directly face each other. Since a facing sector is the major cause of interference (assuming a reuse factor of 1) then this coordination can reduce most of the interference. By sharing their beam information a coordinated schedule can be determined that better provides each base station the ability to serve their users and this coordinated schedule can be changed as the distribution of the users in each sector changes. In Figure 1 we illustrate the model that we use for scheduling beams.
II. COOPERATIVE SCHEDULING

In this section we provide a simple algorithm for achieving the goals of low intercell interference together with high channel utilization. We first consider an idealized version of the problem. We will consider the case of two facing sectors although in practice one typically finds that three sectors face each other. Our approach can be extended to the latter case but with increased computational and communication complexity.

Consider any sector (called sector A) and assume that it can use one of $M$ beams while the sector facing it (called sector B) can use one of $N$ beams. We assume that the interference caused by all other sectors is small and is taken into account by a fixed interference component. Therefore, for each of the $MN$ possible beam combinations there is a corresponding SINR for each mobile. Assume that each mobile can determine and report the corresponding SINR for each beam pair. Let $S_A$ denote the set of users within sector A’s active set and similarly define $S_B$.

For each user $k \in S_A$ let $\gamma_{mn}(k)$ denote the SINR experienced if sector A uses beam $m$ and sector B uses beam $n$. Similarly define the SINR for users in sector B. We assume that each mobile determines these SINR values and reports them to its serving sector. For the reported SINR let $r_{mn}(k)$ denote the corresponding achievable rate. We assume that a utility function $U_k(r)$ is allocated to user $k$ and that this represents the benefit of the connection provided to the user as a function of the throughput provided. A similar approach can be used if a delay dependent function is more appropriate.

The objective is to maximize the sum utility over all users in both sectors. Note that here we are addressing a joint resource optimization problem. We assume that in each frame a single beam is used in each sector. A similar approach can be used for the case where a frame is partitioned into zones (in the time dimension) and each zone uses a different beam. Hence we must determine the beams to be used in each sector as well as the user resource allocations for each frame.

The minimum resource unit will be called a slot. This is a collection of a specified number of subcarriers together with a specified number of symbols. Note that, for a given pair of beams, the optimal allocation of resources in sector A is independent of allocations made in sector B. Hence allocations for each sector can be computed independently.

This utility based optimization problem can be solved as, for example, using the algorithm in [5]. For the optimal solution, let $F_{mn}(A)$ denote the increase in total utility for the frame allocations in sector A for beams $m$ and $n$. We can similarly define $F_{nm}(B)$ as the increase in utility for sector B. We can now determine the optimal beam pair as

$$\left(m^*, n^*\right) = \operatorname{arg}\max_{(m,n)} \left( F_{mn}(A) + F_{nm}(B) \right)$$  \hspace{1cm} (1)

Therefore if the values of $F_{mn}(A)$, which are computed by sector A, are passed to sector B (over the backhaul) and the values $F_{nm}(B)$ are passed from sector B to sector A then each sector can independently determine the optimal pair $(m^*, n^*)$. For example, if $m = n = 4$ and we use one byte per value then 16 bytes of information must be exchanged per frame period. Given the optimal beam pair each sector can then schedule the corresponding users using that pair.

III. PRACTICAL ISSUES

A. Channel Reporting

In order to determine the utility gain for each beam pair we must first determine the SINR achievable with each of the possible pairs. Each mobile determines the SINR for a beam pair only when that beam pair is being used. Using pilots (or the preamble) from its serving base station frames it determines the resulting SINR and reports these values. Therefore if a particular beam pair is rarely used then SINR measurements for the pair cannot be determined.

Instead of measuring the SINR for all beam pairs, one can instead determine the channel gain for each beam in the serving sector together with the interference and noise for each beam in the interfering sector. This requires $M + N$ measurements instead of the $MN$ measurements required for all beam pairs. However, the present standards only allow reporting of the SINR.

In the case of a TDD network, uplink sounding can be used to estimate channel gains for beams in the serving sector since channel reciprocity can be assumed. Given these channel gains and the reported SINR the base station can then determine the interference and noise component. Therefore if each of the $n$ interfering beams are served over $n$ frames then the base station can determine the SINRs of the complete set of beam pairs. In a heavily loaded system all beams in the interfering sector are very likely to be served.

If however, a particular beam in the interfering sector is not served for a long period of time then the SINRs for the beam pairs that include this particular beam cannot be determined. If a beam is not served then this means that the coverage area of the beam has no users. If the scheduler were to schedule transmission of this beam then only power for the pilots would be allocated and hence it is safe for the base station to assume that the interference from that beam, if scheduled, is zero and that only the background noise should be taken into account. If a new user enters the coverage area and must be served, the interference generated will still be small because a single user is being served. The gradual increase in loading within
the sector as more users enter will ensure that the interference generated can be measured and taken into account.

B. User Scheduling for a given Beam Pair

In the previous subsection we demonstrated that the base station can determine reasonably accurate estimates of the SINR for each beam pair for each user in its sector. Given this information it must then determine which beam should be switched on and which users within the beam should be allocated resources.

For a given beam pair this scheduling problem is identical to the single sector scheduler whereby each mobile reports its channel conditions and the scheduler makes resource allocations. Therefore as mentioned earlier we can find the optimal solution for each beam pair and determine the sum utility increase for each pair.

Note however, that this approach requires solving \( MN \) single sector scheduling problems. The intent of solving these scheduling problems is to obtain the utility gain for each beam pair and then to use this information together with similar information from the interfering sector to determine the best beam pair. Therefore the precise solution of each sub-problem is not necessary. Once the best beam pair is determined then we can obtain a more precise solution for the chosen beam. Therefore we outline a simple approximation that can be used.

For a serving beam \( m \) and interfering beam \( n \) let \( \gamma_i \) denote the SINR for each user \( i \) that lies within the coverage of beam \( m \). Our focus is solely on this beam pair and hence we drop the dependence on \( m \) and \( n \). We will use a throughput dependent utility function \( U(r) \) where \( r \) represents the present filtered throughput of the user. The decision variable is the number of slots \( x_i \), allocated to the user. Now note that the throughput of a user depends on the number of slots allocated to the user, \( r(x) \). Hence we can write the utility as a function of the number of allocated slots, \( U_i(x_i) \). A utility function must be a concave and non-decreasing function of the throughput. The throughput is a linear function of the number of allocated slots. Hence the function \( U_i(x_i) \) is also concave and non-decreasing. The optimal decision variables can be obtained by iteratively allocating slots one at a time to the user with the highest utility gradient \( U_i'(x_i) \) where \( x_i \) is the number of slots already allocated to the user (see [5]).

If we instead made the simplification that fractional slots can be allocated to each user then we can explicitly solve the problem. In this case the optimal allocations, \( x_i^* \), are such that \( U_i'(x_i^*) = c \) for some constant \( c \). Note that this solution implicitly assumes that user \( i \) has sufficient data to use \( x_i^* \) slots. If this is not the case then, for those users we must subtract the maximum number of slots they can use from the total number of slots and then repeat (with a new constant \( c \)) to determine the optimal allocations for the other users.

In order to better illustrate this approach let us consider a simple example. We will assume that \( U(r) = \ln(r) \) which will provide proportionally fair throughput allocations. In this case (see [5]) one can show that the optimal allocation is such that

\[
x_i^* + \kappa_i = c
\]

where \( \kappa_i \) is an allocation independent, user specific parameter given by

\[
\kappa_i = \frac{\alpha r_i(0)}{(1-\alpha)\mu_i}.
\]

\( r_i(0) \) denotes the throughput of the user before any slot allocations are made and \( \mu_i \) represents the increase in rate for each additional slot allocated (spectral efficiency). The parameter \( \alpha \) is the filter constant used in updating the user throughput given the number of allocated slots

\[
r_i(x_i) = \alpha r_i(0) + (1-\alpha)x_i\mu_i.
\]

Under the assumption that all users have sufficient data to fill their allocated slots and that slots can be allocated in fractions then each user in the beam will be allocated resources. If we denote the number of users by \( K \) and the total number of slots by \( S \) then, using 2 together with the fact that all slots are used we have

\[
S = Kc - \sum_{i=1}^{K} \kappa_i.
\]

Hence we can explicitly determine the optimal number of slots allocated to user \( i \) as

\[
x_i^* = \frac{1}{K} \left( S + \sum_{i=1}^{K} \kappa_i \right) - \kappa_i.
\]

Given these optimal decision variables we can use 4 to determine the updated throughput, \( r_i(x_i^*) \) for each user and finally we can determine the sum utility of sector \( A \) when using beam pair \((m,n)\) as

\[
F_{mn}(A) = \sum_{i=1}^{K} \ln(r_i(x_i^*)).
\]

Note that, as pointed out earlier, if a particular user does not have sufficient data to fill the slots allocated to it then this user must be removed from the algorithm, the total number of slots must be reduced by the number of slots that the user can in fact use and the above algorithm repeated. Using this algorithm one can explicitly and easily determine the sum utility that can be achieved for each beam pair.

C. Backhaul Signaling Overhead

Using the algorithm in the previous section, each sector can determine the achievable sum utility for each beam pair. However this information is insufficient to determine the optimal pair. The sector also needs the utility information computed by the sector facing it (the interfering sector). If we assume that this information \( F_{nm}(B) \) is passed from sector \( B \) to sector \( A \) for all beam pairs then sector \( A \) can then determine the optimal beam pair by using 1. Sector \( B \) performs the same computations and will hence use the same pair as determined by sector \( A \) and so both sectors are coordinated.
With a fast enough backhaul this information exchange can be done on a frame basis. However if such a fast exchange rate is not feasible then performance can be traded for reduced signaling overhead as follows. Suppose that sectors exchange utility information every \( p < \min\{m, n\} \) frames. From the received information each sector computes beam pairs for the subsequent \( p \) frames as follows. As before, it first determines the best beam pair. This will be used for the first frame. Next it removes all beam pairs that included either of the beams used for the first transmission and again determines the optimal beam pair. This pair is used for the second frame. The process is repeated until all beam pairs for the next \( p \) frames are chosen. The reason for removing already chosen beams from the selection process is as follows. The utility information that is passed in based on the information available for the particular instant in time. However, once the first frame is transmitted the utility information, especially for users in those beams that were served, is no longer accurate. Hence it is best to remove the served beams from the selection list.

\( D. \text{ Beam Set Determination} \)

In the previous sections we assumed that the beam set for each sector was given. However, the set of beams used to serve users can also be varied. This should be done on a much larger time scale compared to the frame duration since it requires changes in the channel quality reporting messages.

Suppose we have a fixed set of narrow beams. The beam set can be changed by forming new beams consisting of collections of two or more of these narrow beams. For example two adjacent narrow beams can be used simultaneously to obtain a wider beam with better coverage. However, note that the power allocated to each of the beams has to be halved and hence the reach of each beam is diminished. Hence such a beam can be used to cover users that lie between the coverage areas of the component narrow beams but this user must be sufficiently close to the base station because of the reduced reach of the beams.

The optimal beam set will also depend on those being used in neighboring cells. However network-wide coordination will be impractical. We therefore focus on single sector beam set determination. We again consider a TDD system. The uplink can be used to determine the location of each user within the sector. The uplink pilot strength provides distance information while joint processing across the receive antennas can provide directional information. Given the location of the mobile, the base station can then determine, for each potential beam, the forward link channel gain for the user if served with the beam. Therefore for each beam set we can determine the average forward link spectral efficiency. The beam set with the largest average value is chosen for the next major cycle.

\( IV. \text{ Simulation Results} \)

In this section we provide simulation results to illustrate the improvement of the proposed coordinated scheduling algorithm compared to the fixed beam cycle allocation approach. Since our focus is on the algorithm rather on physical layer aspects, we will use a simplified model for the transmissions but capture the important aspects for the addressed problem.

We consider two 120 degree sectors, \( A \) and \( B \) that face each other (see Figure 1). \( M \) beams are used in sector \( A \) and \( N \) are used in sector \( B \). We assume that the beams in sector \( A \) provide flashlight coverage from the base station to the edge of the sector and the angle of the beam is exactly \( 120\degree/M \). The beams in sector \( B \) are similarly defined. In sector \( A \) the transmission power density for a sector-wide beam is denoted by \( P \) and hence for \( m \) beams per sector the density for a single beam is \( mP \). We randomly drop users within the two sectors. Each user therefore lies within a single serving beam as well as a single interfering beam. We assume the path gains from each sector varies inversely with the distance taken to the power of 3.5. If the beam in the interfering sector does not cover the mobile then only background noise is used to determine the mobile’s SINR. Otherwise the interference is determined based on the path loss as well as the transmission power used for the beam (which depends on the number of beams used in that sector). Hence for each user and each beam pair we can determine the corresponding SINR and use this information in our proposed algorithm to determine the optimal pair.

We use the following algorithm (which needs no coordination) for comparison purposes. In each sector the beams are chosen in a round robin fashion. The beam allocations are chosen so that whenever a mobile is served it experiences little interference from the opposite sector. Note that this is just one of the many possible allocations that can be used in our proposed algorithm and hence the solution for the proposed algorithm will always be as good as the round robin algorithm. We must show that the resulting gain of the proposed algorithm is worth the additional complexity of exchanging utility information over the backhaul.

We consider the case of \( M = N = 4 \) beams. We simulate the beam round robin algorithm as well as our coordinated scheduling approach. The metrics of concern are sector throughput and fairness. We use the Jain fairness index [6] to measure fairness. This index provides an indication of the throughput fairness. If we denote the Coefficient of Variation (ratio of standard deviation and mean) of the mobile throughputs by \( c_v \), then this fairness index is given by \( 1/(1 + c_v^2) \). This index varies from 0 to 1 with a value of 1 indicating perfect fairness. We also monitor the sum utility of the resulting solution which in this case is the sum of the natural logarithm of the user throughputs. The utility metric combines both fairness and throughput properties and hence is useful for those cases where one metric is better while the other is worse for a particular scenario.

In Figure 2 we plot the ratio of the metric value for the coordinated scheduler to that of the round robin scheduler for each of the metrics, throughput, fairness and utility. Note that a value greater than unity indicates that the metric performance is better for the coordinated algorithm. These ratios are plotted as a function of the number of users per sector. Note that the coordinated scheduler performs better with respect to all three metrics even for a large number of users.
Next we keep the number of users fixed at $K = 20$ per sector and instead vary the number of beams per sector. The performance metric ratios are plotted as a function of the number of beams in Figure 3. Here we find that as the number of beams increases, the relative performance of the coordinated scheduler increases. This is because the probability of having no user in a beam increases and when this occurs the round robin scheduler cannot schedule a user when that beam is chosen and this reduces the throughput. Furthermore, since users within a beam share that beam’s transmission then users within a beam achieve proportionally fair rates. However, if we consider two beams with differing numbers of users then the users within the more populated beam will be provided with lower throughputs than those in the less populated beam leading to cross-beam unfairness.

Note that, even with the round robin scheduler, users within a beam are allocated resources to achieve proportionally fair rates. However this degree of fairness cannot be achieved across beams. This is illustrated in Figure 4 in which we provide the Cumulative Distribution Function of the normalized throughputs for both algorithms. We also provide the fairness constraint that has been used in 3GPP and 3GPP2 evaluation methodologies which is the line from $(0,0)$ to $(0.5,0.5)$. Note that the round robin scheduling algorithm results in throughputs that violate this constraint while the coordinated scheduler provides throughputs that does not.

V. SUMMARY AND CONCLUSIONS

In this paper we addressed the problem of scheduling users within fixed beams in the downlink of a cellular network. In the coordinated algorithm, neighboring sectors exchange information to avoid beam collisions. We propose a utility based scheduling algorithm for jointly scheduling beams in each of the two neighboring sectors as well as the users within each beam. Simulation results show that the coordinated approach provides significantly better performance with respect to both throughput and fairness than that obtained with a static beam allocation approach. This performance advantage holds for a wide range of user populations as well as number of beams per sector.

REFERENCES