Abstract—The next generation of wireless networks (e.g., Long Term Evolution (LTE) [1] and WiMAX [2]) use multiple techniques to improve channel spectral efficiencies. In this paper we focus on one such technique, namely Coordinated Beam Switching (CBS). With CBS, each sector determines a sequence of beams (possibly with repetitions) over which it continuously cycles. Each sector independently determines its beam pattern with the only constraint being that all sectors use a common cycle period. A beam can be used for the entire frame or one may use different beam patterns for each subband of the frame. Therefore the interference pattern of each subband of a sector repeats with the same period and hence the channel quality can be predicted if we assume slow moving UEs (User Equipment). Such a method may not be suitable for delay sensitive applications but, in the case of VoIP, the periodic arrivals of VoIP packets can be synchronized with the periodic service provided to the UE. Furthermore, VoIP service is limited by UEs at the edge but these are precisely the UEs that benefit from beamforming. In this paper we provide algorithms for accomplishing this synchronization and illustrate the corresponding VoIP capacity gains through simulations.

I. INTRODUCTION

During the development of the 3GPP LTE Standard, operators requested increased flexibility in order to provide more focused solutions for specific scenarios. During the initial discussions on LTE-A (Long Term Evolution Advanced), many suggestions were made on how to manage the interference levels by coordinating transmissions across cells [3]. It is well known that as we decrease cell sizes in an effort to improve spectral efficiency, the inter-sector interference problem increases and hence coordination becomes even more important.

Beamforming has been proposed as a promising technique to increase cell coverage and to improve cell edge spectral efficiencies [4], [5]. However, one major drawback of beamforming is the so called flashlight effect whereby the channel quality changes between UE measurements and eNB (enhanced Node B) data transmissions due to changes in the active beams of the neighboring sectors. Coordinated beamswitching has been suggested as a possible solution to this problem in [6], [7], [8], [9]. In this paper we assume that Coordinated Beam Switching (CBS) is used. Each eNB systematically cycles through a set of preferred beams with possible repetitions of beams within the cycle with the only constraint being that all sectors use the same cycle period. These cycling patterns can be changed on a slow basis and will typically be a function of the traffic distribution which changes on the order of seconds. This means that the received SINR of a slow moving UE will fluctuate deterministically in time as the different eNBs cycle through their beams. The reported CQI (Channel Quality Information) will therefore be valid at some known time in the future given that the UE is moving slowly relative to the cycling period. One can apply this beam-switching technique independently in each subband of each frame and hence achieve even greater flexibility in the allocation of resources.

With CBS, the eNB can predict the best frame/subband pair in which to serve a UE since the interference pattern repeats for each subband over each beam cycle period. However, this means that service to the UE is delayed until the appropriate frame. This delay is tolerable for certain traffic types (such as data) but may not be tolerable for delay sensitive traffic. However, if we consider a periodic, delay-sensitive service such as VoIP we note that as long as we synchronize the near-periodic arrivals of VoIP packets with the periodic servicing of the UE then the delay requirements can be met. Therefore with proper design, CBS can be used to support VoIP service. In this paper we investigate the resource allocation algorithms needed to achieve this combination.

In the following we will show that, with CBS, the average SINR achieved by each UE increases when compared to the non-beamforming (henceforth called baseline) case. This SINR varies because of the variation of the interference experienced in each subband/frame pair (henceforth called a beam allocation unit or BAU). The gains come from exploiting this interference diversity. On the other hand, since the feasible resource allocation space of each UE is reduced by the limited servicing opportunities then the achievable statistical multiplexing gains decreases. We must therefore find a suitable balance between these two objectives. Another issue that must be addressed is the proper balance of resources across beams. We propose a simple adaptive algorithm for choosing the number of times each beam must be allocated a BAU.

In the next section we provide further details of the CBS resource allocation algorithm as well as the model used for VoIP service. We then illustrate the interference diversity gains that we exploit. Next we use simulations to illustrate that performance gains can be achieved and show the dependence of this gain on the various design choices (subband and beam set sizes). We also illustrate the convergence of the proposed algorithm for beam resource balancing.

978-1-4244-6406-7/10/$26.00 ©2010 IEEE
II. COORDINATED BEAMSWITCHING DETAILS

Cell coordination can happen at many levels. At the highest level, a central scheduler has full knowledge of all radio conditions of all UEs in a cluster of cells and tries to find an optimal selection of UEs to be served at any instant in time. This solution would unfortunately require an excessive overhead on the backhaul, and may be simply impossible to implement if we consider the fundamental delays on the backhaul transmission. This necessitates that we find a simpler, sub-optimal solution that would be easier to implement. Simpler solutions have been proposed independently in [7] and [8].

In the approach proposed in [10] each eNB determines its own beam cycle pattern based on the geographical distribution and QoS needs of its UEs. This pattern does not need to be communicated to its UEs nor to the other eNBs. Such a pattern can be independently computed for each subband. In each frame within a cycle and each subband within the frame the corresponding beam is used to serve one or more UEs (i.e., those that lie within the corresponding beam). Note that this requires that a cell specific RS (Reference Symbol) be used in each subband for CQI measurements. The RS must be beamformed in the same manner as the data within the subband. Furthermore, all eNBs must use the same beam cycle period in order to ensure that CQI measurements repeat with this period. Each UE need only report a subset of its CQI measurements (i.e., its best BAUs (subband/frame pairs)).

In Figure 1 we provide an illustrative example of such a beam allocation pattern. For each BAU a beam is chosen and that beam is then used to serve UEs that it covers. One can also include wide beams in addition to narrow beams as long as the pattern is repeated every cycle period. Note that each beam can be repeated a different number of times based on the number of UEs within the beam.

We must determine how many times to allocate each beam and where should each beam be allocated within the beam cycle period. Since a beam may be used to serve multiple UEs and such UEs can experience different interference patterns then it will typically be difficult to allocate beams optimally to BAUs and hence we do not attempt to do this. However, the number of times that a beam is allocated within a period should be related to the number of UEs covered by the beam as well as their QoS constraints and channel conditions.

III. OVERVIEW OF VOIP SERVICE

The VoIP protocol is used for packetized voice service. Every 20ms a VoIP packet is encoded at the source and transmitted to the destination. Over the air interface the VoIP packet is encoded within one or more Resource Blocks (RBs) for transmission. The number of RBs required depends on the payload of the VoIP packet as well as the channel conditions of the target UE.

We model the VoIP packet arrivals of each UE by an ON/OFF arrival process. The ON state corresponds to voice activity while the OFF state corresponds to a silence period. During the ON period we assume full-rate voice frames arrive every 20ms with a resultant bit rate of 9.6 kbps. We therefore do not model the lower rate frames during speech activity. However since our intent is to provide a comparative analysis then our simplification is reasonable. During the OFF period (the silence period) packets arrive every 8 VoIP frame periods. These packets are used to provide background noise information at the receiver. We call the fraction of time a UE’s connection is in the active state the voice activity factor.

A VoIP connection is considered acceptable if it conforms to specified delay and packet loss rate requirements. The IP packet loss rate must not exceed a specified value (e.g., 2%). These consist of losses due to transmission errors (i.e., the maximum number of Hybrid ARQ (HARQ) transmissions for the packet was reached and the packet was still not decoded correctly) as well as packets discarded because they exceeded some specified delay and so would not have been delivered in a timely fashion. The delay requirement is that the time from source (mouth) to destination (ear) must not exceed some specified value (e.g., 250 ms). Violation of the delay requirement results in unacceptable user perceived performance.

During an active period a semi-persistent connection is allocated to the UE. In such a connection resources are reserved every 20ms for transmission of a VoIP packet. If the transmission fails then it is re-transmitted but these re-transmissions are dynamically scheduled. The size of the resource reserved for the connection depends on the Modulation and Coding Scheme (MCS) of the transmission. Therefore by suitable choice of this MCS mapping as well as providing re-transmissions with high priority one can maintain the delay and packet loss rate requirements for transmissions during an active period.

When a connection transitions from silent to active a semi-persistent connection must be allocated. However if no resources are available then the initial packet in this talk spurt must be delayed. As the system capacity is approached, such packets begin to exceed the delay threshold and at that point are dropped. Therefore by monitoring this delay one can determine when capacity is reached and also when admission control should be invoked. In our simulations we therefore assume that frame allocations are done in the following order (a) ongoing semi-persistent connections, (b) HARQ retransmissions, (c) Silence Indicator (SID) packets and (d) new semi-persistent connections.

![Fig. 1. Illustrative Example of Beam Allocations within a Cycle](image-url)
IV. AFFECT OF BEAMFORMING ON UE SINR VALUES

We first investigate the effect of periodic beamforming on the SINR values of UEs. Consider a UE in sector \( i \) and assume that it experiences interference from \( N \) other sectors. Let \( g_i \) denote the path gain of the UE from its serving sector and \( g_j \) the path gain from interfering sectors \( j = 1 \ldots N \). We assume that the same transmission power \( p \) is used in each sector over the bandwidth allocated to the UE and denote the background noise of the UE by \( n_0 \). We can therefore write the Signal to Information and Noise (SINR), \( \gamma \), for this UE as

\[
\gamma = \frac{pg_i}{n_0 + \sum_{j=1}^{N} pg_j}.
\]

Next we consider the effect of beamforming. We use a simplified model for beams as follows. We assume that the UE either falls within the beam of the interfering sector or it does not fall within the beam and experiences no interference. We use \( x_j = 1 \) to denote if the UE falls within the beam of sector \( j \) and \( x_j = 0 \) if it does not. If the UE falls within the beam then the beamforming gain is also \( \kappa \). In this case we can write the SINR as

\[
\hat{\gamma} = \frac{\kappa pg_i}{n_0 + \sum_{j=1}^{N} x_j \kappa pg_j}.
\]

Now note that

\[
\hat{\gamma} \geq \frac{\kappa pg_i}{n_0 + \sum_{j=1}^{N} \kappa pg_j} = \frac{pg_i}{n_0/\kappa + \sum_{j=1}^{N} pg_j} > \gamma
\]

since \( x_j \in \{0,1\} \) and \( \kappa > 1 \). Therefore the SINR achieved by the UE with beamforming always exceeds that achieved without beamforming. Hence as long as we serve a UE with the beam that covers it then we expect its achievable rate to be higher when beamforming is used. We can try to quantify this gain as follows. Assume \( K \) non-overlapping beams covering the entire sector. We can approximate the beamforming gain by \( \kappa \). Assuming that each beam is equally likely to occur then the probability that \( x_j = 1 \) is \( 1/K \). Therefore the expected value of the interference experienced by the UE is given by

\[
E \left[ \sum_{j=1}^{N} x_j \kappa pg_j \right] = \sum_{j=1}^{N} pg_j
\]

which is simply the interference experienced by the UE without beamforming. Hence introducing beamforming roughly increases the SINR of the UE by a factor of \( \kappa \). However this is an upper bound and in practice the gain may be significantly less but will typically be greater than unity.

The SINR of a UE can therefore be maximized if the UE is served (with the beam that covers it) when the sum interference is minimized. Therefore beams should be scheduled accordingly. Unfortunately, since several UEs are served within a beam then not all will achieve their minimum interference at the same time. Hence we do not try to achieve this gain and instead allocate beams to resources independently of the SINR values achieved by the UEs within the beam.

V. BEAM COUNT OPTIMIZATION

The objective for VoIP service is the maximization of the number of users supported given that at most 2% of UEs experience outage. A UE is deemed to be in outage if more than 2% of its packets are lost (from transmission failures or discarded because of excessive delay). Therefore in order to achieve capacity one must minimize the maximum outage over all UEs in the sector. If a UE is in outage while another was not then the capacity can potentially be increased by moving resources from the UE not in outage to the one in outage. Given this observation, capacity is achieved when all UEs achieve some common outage probability. Since outage is difficult to monitor, we will instead attempt to equalize resource utilizations (which will result in equal outage probabilities if all UEs had the same channel statistics).

Consider the CBS scenario. Here we have \( K \) beams and each UE is covered by exactly one of these beams (e.g., based on their position). Each frame is divided into \( M \) subbands and there are \( T \) frames in a cycle. Therefore there are \( MT \) BAUs that must be allocated to beams. As we increase the number of BAUs allocated to a beam the resource utilization of those BAUs decreases because there is a fixed number of UEs covered by the beam. Hence we should adjust the number of times a beam is allocated based on the utilization of its resources.

We define a utilization measurement period (e.g., every 100 frames). At the end of each such period we compute the resource utilization for each beam. This is the ratio of the number of RBs used in each BAU allocated to the beam divided by the total number of RBs in these BAUs, over all frames in the measurement period. Let \( \rho_i \) denote this utilization for beam \( i \). Let \( r_b \) denote the number of BAUs allocated to beam \( b \). Let \( r_b^* \) denote the number required to ensure that all beams have the same resource utilization. We can therefore use the present estimate of utilization values to determine the number of BAUs per beam that would be required to equalize utilization. If we do this we find that

\[
r_b^* = MT \frac{\rho_i r_b}{\sum_{i=1}^{K} \rho_i r_i}.
\]

Note that this quantity may not be integral and so must be rounded. This computation is performed periodically but the changes cannot be immediately incorporated because certain resources (and hence associated beams) may have been allocated to semi-persistent connections and hence cannot be changed. Therefore we instead determine which BAUs are empty and make changes to the beam allocated to those BAUs so as to achieve the desired beam count allocations.

In practice this approach quickly achieves the desired beam count allocations.

VI. CQI REPORTING AND SCHEDULING OF UES

In the previous section we determined how many BAUs should be allocated to each beam. In order to minimize the time required for a UE to be served, we should allocate beams so that as many beams as possible occur in each frame
(given the computed beam allocation counts). This can be accomplished by sequentially allocating beams to BAUs and removing beams from consideration once their allocated count is reached.

In this section we describe how UEs are allocated resources (RBs) within a BAU. First we need to explain the Channel Quality Information (CQI) reporting mechanism. Each BAU contains at least one Reference Symbol (RS) that is used by the UEs to measure the channel quality. This RS is beamformed using the same beamforming weights used for data transmission over the BAU. Each UE can therefore measure the channel quality of each of the MT BAUs in a cycle. It then reports, together with the associated BAU indices, the Q best SINR values within a cycle. Note that, Q should be chosen large enough so that the UE can report SINR values for all BAUs that belong to the beam that covers the UE. However, UEs do not need to know in which beam it lies or which beam is being used at any point in time.

In each frame, all RBs that were previously allocated to a semi-persistent connection must be reserved for those connections. The remaining RBs are free for assigning new persistent connections as well as for re-transmissions and SID packets. We assume that re-transmissions are given higher priority over new semi-persistent connections. Re-transmissions are sorted from largest to smallest delay and scheduled in this order. Consider any of these re-transmissions and the associated UE. Starting from largest reported SINR for the UE to the smallest SINR, the number of RBs required to serve the packet is determined and a search is made within the corresponding BAU to determine if the necessary RBs can be allocated. If not, the procedure is repeated for the next highest SINR until all reported SINR values are exhausted. If the UE cannot be allocated resources then the packet is queued and the process repeated for the next UE. The same process is repeated for SID packets.

After attempts are made at scheduling all re-transmissions and SIDs then we next allocate resources for new persistent connections. Again we sort UEs based on delay but this time we do so from smallest to largest (i.e., we use a Last in First Out (LIFO) discipline). By so doing we reduce the number of UEs that experience poor service but those that are affected have severely degraded performance (this technique is routinely used in telephony). We then allocate resources in the same manner as was done for re-transmissions. Note that as the number of UEs in the system approaches capacity, the tail of the delay distribution of new connections increases. This information can therefore be used to invoke admission controls but details are beyond the scope of this paper. The implicit assumption here is that persistent assignments are made only for the first transmission of a packet while re-transmissions and SID packets are dynamically scheduled. Note that, in our simulations we do not optimize placement of a UE (as described above) but we simply search sequentially for the first available set of resources. This is faster but less optimal.

VII. OPTIMAL SUBBAND AND BEAM SET SIZES

One parameter that must be optimized is the number of RBs per subband. As the subband size increases, the number of BAUs decreases and hence the number of times a beam is allocated within a period decreases. This means that the number of transmission opportunities of a UE decreases since it can only be scheduled within its own beam. The interference experienced by the UE varies across the BAUs within which it can be scheduled and hence the interference diversity gain also decreases. On the other hand, in most cases, a UE cannot be allocated across BAUs since its CQI will be different. Therefore UE allocations must be made completely within a BAU and hence as the BAU becomes smaller, the statistical multiplexing gains decrease.

Another parameter that must be optimized is the beam set size. If beams are non-overlapping then, as the beam set size is increased, the average SINR achieved by each UE (when served by the beam that covers it) also increases (more focused beams). However, since there are more beams then the average number of UEs per beam is smaller. This results in a reduction in the statistical multiplexing gain possible when packing these UEs together in BAUs.

In Figure 2 we illustrate how capacity enhancement is achieved. The top figure contains frames in a period and the associated subbands for the case of no beamforming. We indicate the number of RBs required for a chosen UE in each subband. Since it is the same for all BAUs then the UE can be allocated resources anywhere. The bottom figure contains the case for beamforming. In this case the first subband in each frame is used for the beam that covers the UE. Therefore the UE cannot be scheduled in any of the other BAUs (marked with an X). Hence the average number of RBs required for the UE (given that it is served in its beam) is 12/5 = 2.4. Hence if all UEs experience a similar RB reduction factor then one can potentially achieve an increased capacity of 3/2.4 which is a 25% increase.
VIII. SIMULATION RESULTS

In this section we provide simulation results based on a simplified model to illustrate the performance gain of using beamforming for VoIP service. We assume a cycle period of 20 frames with each frame being 1ms in duration. There are 24 RBs per frame and we will investigate the capacity dependence on such factors as subband size as well as beam set size. However we first investigate the beam resource balancing algorithm to show that it converges quickly to the correct beam count allocations.

Since our interest lies in relative capacity gains of performing beamforming, we use a simplified model for the physical layer transmissions. The SINR of a UE within a specific BAU depends on several factors including its location within the sector, the beam used for the BAU, the beams used in neighboring sectors and the beamforming gains. Given this SINR the eNB then determines an appropriate MCS value and then this determines the number of RBs required to deliver the data queued for the UE. Simulations for the baseline problem show that typically no more than 3 RBs are required per transmission. Hence we use a statistical model (with probabilities determined from system level simulations) to generate the number of RBs required for each UE in the sector for the baseline case. We do the same for the beamforming case.

The above model thus allows us to generate the number of RBs required for each UE in each BAU. The proposed algorithms are then used to simulate the system. Each UE transitions between the silent mode and active mode with the duration in each mode taken from an Exponential Distribution. Once a connection becomes active, resources are allocated for a semi-persistent connection. Recall that such connections are given the lowest priority (compared to already allocated semi-persistent connections as well as re-transmissions). We monitor the packet drop rate for each UE and determine the capacity based on the outage criterion.

A. Beam Resource Balancing

In this section we investigate the method used to determine the optimal number of times that each beam should be allocated. A summary of this algorithm is provided in Figure 3. For this investigation we assume 4 beams and 4 subbands. We initially allocate 20 BAUs to each beam since there are a total of 20x4=80 BAUs. However we assign 40 UEs to beam 1, 80 to beam 2, 120 to beam 3 and 160 to beam 4. If the average number of RBs per UE is the same for each beam then the optimal beam counts would be 20 for beam 1, 16 for beam 2, 24 for beam 3 and 32 for beam 4. However these may not be exactly correct because of statistical variations in the RBs required per UE. We compute the optimal beam counts every 100 frames. Once computed, we then slowly modify the beam count allocations by making changes only to those BAUs that are empty after semi-persistent connections are deallocated but before new ones are assigned or retransmission scheduled. In Figure 4 we plot the actual beam count allocation for each beam as a function of frame index. Each beam is initially allocated to 20 BAUs but we see that after 100 frames the allocations quickly approach the optimal values.

B. Capacity Gain

Next we investigate the case of evenly distributed UE drops. Here we do not perform the beam resource balancing algorithm as described in the previous section but instead we ensure that each beam contains the same number of UEs and we then evenly distribute BAUs among beams. Note that this still is not optimal because it is possible that the UEs in one beam require a larger average number of RBs than another beam and thus will also need a larger number of BAUs. A summary of the algorithm is provided in Figure 5.

For the baseline case we determined a capacity of 355 UEs. We then simulated the beamforming case for different numbers of subbands. The capacity results (normalized by the baseline capacity) are provided in Figure 6. Here we observe that, for the case of many subbands, the statistical multiplexing of UEs within subbands is not efficient. On the other hand, for few subbands the interference diversity gain is small because there
foreach BAU {
    sequentially allocate beams }
foreach UE {
    choose a serving beam
    generate number RBs required for each BAU }
for i=1:T frames
    foreach UE {
        determine if state transition }
    foreach active UE {
        allocate previously reserved resources
determine outcome of transmission }
    foreach silent UE {
        if SID is generated allocate resources }
    foreach retransmission {
        if resources can be allocated
            allocate and determine outcome
        else
            treat event as failed transmission
    }
    drop packet if max retry attempts reached }
foreach silent to active UE {
    allocate resources in serving beam
    if allocation not possible, queue packet
drop packet if delay exceeds threshold }
determine outage probability

Fig. 5. Pseudo-Code for VoIP Scheduler

Fig. 6. CBS Capacity Gain as a function of No. of Subbands

are fewer beam opportunities for each UE. For this scenario the optimal number of subbands is 4 (i.e., 6 RBs per subband). Note we achieve as much as 17% capacity gain over the non-beamforming case.

Next we investigated the dependence on the size of the beam set. In each case we assume that each UE is served by exactly one beam and that UEs are evenly divided across beams. The average number of RBs required for a UE in its serving beam decreases as the number of beams increases (i.e., more focused beams). In Figure 7 we plot the capacity results as a function of beam set size. We fixed the number of subbands at 4 although this may not necessarily be the optimal value for all cases. For small beam set sizes we obtain higher SINR values but insufficient opportunities to use them (i.e., low statistical multiplexing gains). For larger beam set sizes the performance gain is maintained but recall that we allocated equal numbers of UEs to beams and in practice this will not be the case, especially for large numbers of beams.

IX. CONCLUSIONS

In this paper we investigated the use of network-synchronized beamswitching for the case of VoIP service. Unlike the case of data, special care must be taken to allocate resources while taking into account the delay sensitivity of the VoIP traffic. We provided algorithms to perform this resource allocation. We then illustrated, through simulations, that the proposed algorithms can provide significant gains in capacity when compared to the non-beamforming case. We also proposed an algorithm for resource balancing over beams.

REFERENCES

[8] Samsung, ”Inter-Cell Interference Mitigation Through Limited Coordination”, R1-082886, Jeju, Korea, Aug 18 - 22, 2008
[9] C802.16m-08_487, Alcatel-Lucent, ”Grid-of-Beams (GoB) Based Downlink Multi-User MIMO”, IEEE802.16m, May 2008