Abstract

We introduce here a new class of packet scheduling techniques for TDMA-based (pure or hybrid) cellular packet radio networks with full frequency reuse. These algorithms, which are suited for networks with both centralized and distributed control, assign the available resources according to the SIR predicted at each slot of the frame. Partial prediction of the interference is made possible by the static preassignment in each cell of a maximum transmitted power level to each slot of the frame. This technique, also named power shaping, exploits a set of suitably reused power profiles to partially organize the intercell and intracell interference in the available slots. Unlike traditional channel state dependent scheduling techniques, which are not able to deal with quasi-stationary location-dependent radio channel conditions, the proposed method is able to provide channel resources with different levels of SIR inside the frame, even in stationary environments, and to assign them to the packets waiting for service. We show that these techniques are able to increase the capacity of systems with and without centralized resource management control, while maintaining the capability of providing acceptable quality of service to heterogeneous classes of users.

1 Introduction

In cellular wireless packet networks resource assignment and scheduling techniques play a key role to ensure an efficient use of the radio spectrum and to provide quality of service guarantee [1, 2]. This is of particular relevance for next generation wireless systems where, to support data and multimedia traffic, data units from different users at the radio interface have to be suitably scheduled in order to share the available radio channels and the transmitted power should be suitably controlled to limit cochannel interference. In order to achieve efficient use of the radio resources, it is needed that scheduling and resource allocation algorithms are channel and interference aware [1, 2].

In the past, a large number of scheduling algorithms have been proposed for wireline networks. Later, it was realized that such algorithm could not be efficient in wireless networks, due to the presence of time and location dependent signal attenuation and interference leading to burst errors and unfair channel capacity. Therefore, different techniques considering channel state as a basic information for the scheduler and dealing with time and location varying channel capacity were proposed [3, 4]. A useful mechanism of these techniques is to defer transmission when the channel is in bad state; however, when the possible interference can not be easily predicted due to the bursty nature of packet transmission, or when fading changes are very slow, i.e. bad state are very long, such techniques are not able to guarantee quality of service to a significant part of users.

Here, we focus our attention on wideband cellular networks equipped with a pure or hybrid TDMA structure in the radio interface to support packet-switching communication for multimedia and data services. Examples of these systems can be found in third generation TD/CDMA proposals, in broadband wireless access networks; both are characterized by the adoption of a full-reuse plan. In the new approach that we want to introduce here, a maximum power level to each slot of the frame is first assigned, in advance (static allocation) with a predefined scheme, thus creating resources with different power and interference conditions, then SIR-(signal to interference ratio-) dependent packet scheduling is applied to distribute network resources. This is done by exploiting a set of suitably reused power profiles that limit (or shape) the maximum power transmitted in each slot/code of the frame. In this way it is possible to partially organize the intercell and intracell interference in the network and to schedule the packets of each user on the basis of the power required to fulfill a predefined carrier-to-interference ratio, leading to a new class of scheduling algorithms. With this technique, even when scheduling is running independently at the different base stations or access points or when the channel is very slow time-varying, slots with different partially predictable capacities are available inside the frame.

The proposed class of algorithms is introduced here by considering a mechanism based on the Earliest Due Date (EDD) [5] service discipline which tries to keep the delay bounded especially for real time service classes. The promising results obtained here suggest for future work a further extension to joint Sir dependent and Wireless Fair Queuing [4] mechanisms.

The use of static power preallocation, also named power-shaping, was recently proposed by the Authors to improve resource assignment in fixed broadband wireless access systems [6] where interference is mitigated by the use of highly directional antennas, and in [10], whereas the idea of using an a-priori static power control scheme before scheduling in packet-switched wireless systems first appeared in [7] and was investigated for a generic linear cellular system. Here, we will apply the new concept to systems with and without centralized resource management control and investigate its impact on system capacity.
2 System model

We investigate here the downlink of a cellular system with hexagonal cells. Each cell has three 120 degrees sectors covered by three antennas with radiation pattern $f(\phi)$ centered in the middle of the sector at $\phi = 0$. A 3 dB-beamwidth equal to 120 degrees, and an infinite front to back ratio would correspond to an ideal antenna pattern (zero intercell interference). However, in our study we consider overlapping antenna pattern with 3 dB-beamwidth equal to 130 degrees and front to back ratio 25 dB, in order to have intracell interference. Each sector is assigned a label (A or B or C), which is suitably reused according to the reuse plan illustrated in Fig. 1. $U$ user terminals, equipped with omnidirectional antennas are randomly placed in the service area with uniform spatial distribution: all users share the common channel by using a TDMA radio interface. The reuse factor is one: each transmitting antenna can use the whole spectrum and all resources of the system. We consider stationary (fixed terminals) or quasi stationary users (e.g. pedestrian) and for this reason we will assume fixed user positions during each simulation run. The propagation model includes path-loss, lognormal shadowing and Rayleigh fading. The latter is assumed to be quasi-stationary in time due to the low degree of mobility of the users and is described by means of unitary mean random variables independently chosen for each user. Lognormal shadowing is taken as constant in time. Each user is served by the sector and the cell for which the received power is maximum.

The delivery of the packets to users is performed on a TDM frame composed of $N$ slots, numbered from 0 to $N - 1$. All base stations are perfectly synchronized on each slot and try to allocate one packet per slot.

The traffic sources are modeled as in [8] to reproduce in a qualitative fashion the burstiness of packet arrival process. Each user can be assigned either a real(R) or non-real (NR) traffic source. Each traffic source evolves, frame by frame, from the ON state to the OFF state and vice versa, according to a two state Markov chain with the transition probabilities $p_{on-off}$ and $p_{off-on}$ which are set in our results to $1/20$ and $1/50$ and $1/350$ for NR sources, respectively. In a $ON$ frame, a source generates $N_B$ burst and the burst size is $S_B$ packets. For R sources $N_B$ is set to 1 and $S_B$ is a Poisson random variable with mean equal to 2, whereas for NR sources, both $N_B$ and $S_B$ are Poisson distributed with mean equal to 2.

A packet does not reach successfully the user if there are at least $B$ packets in the slot, $B$ being the number of slots in each branch of the star, the allowed values of the frame length is $N = 3R$. Simple star impose a constraint ($N = 3R$) on the number of the slots: for $N \neq 3R$, others geometries can be used (e.g triangle [10]) and hence this is not a problem.

Moreover, from Table 1 we also note that $K = 2R$ power levels $p_j, j = 0..K - 1$ are set starting from an edge of the pattern and ending to the other two edges, in a symmetric fashion: $j$ decreases as the "distance" from the starting edge increases. The "distance" indicates the grade of protection of-

<table>
<thead>
<tr>
<th>Slot</th>
<th>A-edge</th>
<th>B-edge</th>
<th>C-edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$p_0$</td>
<td>$p_0$</td>
<td>$p_0$</td>
</tr>
<tr>
<td>1</td>
<td>$p_1$</td>
<td>$p_1$</td>
<td>$p_1$</td>
</tr>
<tr>
<td>2</td>
<td>$p_2$</td>
<td>$p_2$</td>
<td>$p_2$</td>
</tr>
<tr>
<td>3</td>
<td>$p_3$</td>
<td>$p_3$</td>
<td>$p_3$</td>
</tr>
<tr>
<td>4</td>
<td>$p_4$</td>
<td>$p_4$</td>
<td>$p_4$</td>
</tr>
<tr>
<td>5</td>
<td>$p_5$</td>
<td>$p_5$</td>
<td>$p_5$</td>
</tr>
<tr>
<td>6</td>
<td>$p_6$</td>
<td>$p_6$</td>
<td>$p_6$</td>
</tr>
<tr>
<td>7</td>
<td>$p_7$</td>
<td>$p_7$</td>
<td>$p_7$</td>
</tr>
<tr>
<td>8</td>
<td>$p_8$</td>
<td>$p_8$</td>
<td>$p_8$</td>
</tr>
</tbody>
</table>

Table 1: 3-labels star patterns
fered by the serving edge against the two possible interfering edges. In fact, in the proposed planning of Fig. 1, sectors with label A, as an example, see with the same “weight” the interference coming from the sectors with the other two labels, B and C, (i.e., the main interferer for an A-labeled sector can be a B-labeled or a C-labeled sector with the same probability). This also justifies the axial symmetry of the proposed pattern.

One of the simplest choices for the power levels $p_j$, is the straight line (linear shaping)

$$p_j = P_{\min} + j \frac{P_{\max} - P_{\min}}{K-1}$$

(1)

where $P_{\max}$ and $P_{\min}$ are parameters that should be suitably set.

In the case depicted in Fig.1 the three profiles are suitably reused according to the geometrical constraints of a hexagonal cellular scenario. Specifically, the labeling is done in such a way that for a user in a sector labeled with A, the largest interference comes from sectors labeled with B and C. In this way, slots that allow a large (small) transmission power, are mainly interfered by users with a small (large) power leading to slots with different levels of protection. Hence, terminals in unfavorable position can take advantage from protected slots, whereas interference resistant terminals can be scheduled in slots where serving power is small.

### 4 SIR dependent packet scheduling

To introduce this class of algorithms we consider here a mechanism based on the Earliest Due Date (EDD) [5] service discipline which tries to keep the delay bounded especially for real time service classes. It runs once a frame and tries to schedule one packet per slot.

We also consider here different types and grades of centralization. Let us define the cluster as a set of sectors coordinated by the same controller: each controller independently runs the scheduling algorithm for its users. In the case of ideal full centralization we have a unique controller coordinating all sectors in the system, whereas, when a distributed scheme is adopted, each controller works only for one sector of one single cell (scheme 1-1). Besides these schemes (see fig.2), we also consider the case of partial centralization: possible solutions are a controller which coordinates all the 9 sectors of groups of 3 cells (scheme 9-3), the 3 sectors of each cell (scheme 3-1) and each triple of opposing sectors (scheme 3-3). In this paper 1-1 and 3-3 schemes will be considered in numerical results.

Each controller, due to the a-priori static power allocation driven by the use of power-shaping, knows for each user and for each slot of the frame the worst case interference coming from other clusters (extra-cluster interference). This could be obtained, for example, through feedbacks, and supposing slow varying channel conditions. It is also assumed that each controller knows the channel attenuation for each user of the cluster.

In the following, we will refer to resource to indicate the channel to be allocated. In the case of TDMA the resource is the couple (serving sector, time slot). The same could be applied to other techniques such as OFDMA. The goal is to assign packets to the resources available in the frame. In this paper we consider the case of fixed user-sector assignment (i.e a users is always served by the sector with the smaller short-time averaged attenuation). However, the algorithm explained in this section deals with the general case which allows a user to be dynamically served by any sector of its cluster in order to have more flexibility against different traffic-interference conditions.

The algorithm runs at the beginning of each frame: let us define, $P_{cons}(i), P_{act}(i), P_{min}(i)$ as the power levels that would be needed by resource $i$ to serve one packet of user $u$ (meeting SIR_{thr}) by considering, respectively,

- for $P_{cons}(i)$: the set of all extra-cluster and intra-cluster resources $k \neq i$ used at the maximum power $P(k,l_k)$, which is the power profile with label $l_k$ referred to resource $k$ (this represent the worst case interference)
- For $P_{act}(i)$: the set of all extra-cluster resources used at the maximum power $P(k,l_k)$ and the set of the actually allocated intra-cluster resources used at the worst case power $\min(P_{cons}(k), P(k,l_k))$. (this does not necessarily implies that resource $k$ is actually allocated at the worst case power)
- For $P_{min}(i)$: only the set of the actually allocated intra-cluster resources used at the worst case power $\min(P_{cons}(k), P(k,l_k))$.

A resource $i$, if assigned to a packet of user $u$, is secure (S) if $P_{cons}(i) \leq P(i,l_i)$, conditionally secure (CS) if $P_{act}(i) \leq P(i,l_i) \leq P_{cons}(i)$, insecure (NS) if $P_{min}(i) \leq P(i,l_i) \leq P_{act}(i)$ and, finally, useless (U) if $P_{min}(i) > P(i,l_i)$, (since
we propose to avoid its use). A resource is marked as useless also if its allocation to a packet of user \( u \) causes a degradation (i.e., from \( CS \) to \( NS \) or from \( NS \) to \( U \)) of already allocated resources. Obviously, a resource is busy (\( B \)) if it has already been allocated.

For each traffic source there is a first-in-first-out queue located at the controller to which the user is affiliated. The scheduling algorithm, frame by frame, assigns the available resources to the packets in the queues according to a priority level \( P \), evaluated for each packet as a function of the Channel State \( CS \) and the Time to Deadline \( TD \).

The SIR dependent channel state is defined for each packet as

\[
CS = \begin{cases} 
R_S & \text{if } R_S > 0 \\
R_{CS} & \text{if } R_S = 0, R_{CS} > 0 \\
R_{NS} & \text{if } R_S = 0, R_{CS} = 0
\end{cases}
\]

where \( R_S \) is the number of not yet assigned \( S \) resources, \( R_{CS} \) is the number of not yet assigned \( CS \) resources, \( R_{NS} \) is the number of not yet assigned \( NS \) allowed resources. All \( S \) and \( CS \) are allowed resources; a \( NS \) resource is allowed if it guarantees the 'admission criterion' (defined later) and \( TD \leq TD_0 \) (i.e., the packet is approaching the deadline). \( TD_0 \) is set to 3 in the results. The time to deadline is defined for each packet as the difference between the due time and the actual time. The due time is evaluated by adding to the time of arrival the maximum delay which is desired to be guaranteed. This maximum delay is fixed to 5 frames for \( R \) services and to 50 for \( NR \) services.

The priority level is defined for each packet as

\[
P = CS \cdot TD
\]

The packet with the smallest \( P \) has to be served first. If \( CS = 0 \), \( P \) is set to infinity. Therefore, the meaning of the proposed strategy can be explained as follows: a packet has to be served urgently when the deadline approaches or when the number of available resources approaches zero. Parameter \( TD \) is lower bounded to 0 and upper bounded to 5 for \( NR \) services to avoid unfair treatment of \( NR \) users. The efficient allocation of channel resources is taken into account in the definition of \( CS \) which first consider secure slots for service; when secure slots are no more available conditionally secure slots (having a greater cost for the network) are considered for service, and so on. Additional resources of type \( NS \) are made available when the deadline is approaching to help users in bad location dependent channel conditions.

When a packet is served it chooses its preferred resource among free (not \( B \)) and not useless (not \( U \)) resources: \( S \) has priority over \( CS \), and \( CS \) over \( NS \). Among \( S \) (or \( CS, NS \)) resources, priority is given to the ones with the smallest number of unallocated users having such resource classified as \( S \) (or \( CS, NS \)). If this is not enough to decide, the resource with smaller \( P_{cons} \) is preferred, and, if two or more resources have the same \( P_{cons} \), the choice is made randomly.

If a packet is not successfully transmitted in a given frame, it remains in the queue to be served in the next frame. If a packet of a \( R \) service cannot be successfully transmitted before deadline, it is ‘dropped’. Packets of \( NR \) services are not dropped, even if the due time is exceeded. The algorithm allocates one (not yet allocated) packet at a time, until no more packets or resources can be allocated. Each time a user is chosen, its preferred resource is assigned to it.

In order to have a more efficient use of network resources and a better quality guarantee, an admission control policy is also introduced. For a system with static channel conditions a user is admitted for service if at least one resource \( i \) exists such that

\[
P_{act}(i) \leq P(i, l_i) + \Delta(\text{load}) \ (dB)
\]

where \( P_{act}(i) \) is evaluated in the absence of allocated slots and \( \Delta(\text{load}) \) is a non negative threshold which depends on the traffic load of the network. Note that this condition is always verified when at least one \( S \) or \( CS \) slot exists.

## 5 Simulation results

The simulated environment consists of 36 hexagonal cells with 3 sectors each, wrapped around onto themselves in order to avoid border effects. We have defined load as the number of packets generated in the system, normalized to the total number of resources available in the system itself. \( SIR_{thr} \) is chosen equal to 6.8 dB. Propagation exponent \( \beta \) is set to 4 and a dB spread of 6 dB is taken for lognormal shadowing. Linear shaping with \( P_{max}/P_{min} = 100 \) on 9 slots is chosen, when not specified otherwise. Obviously, \( P_{max}/P_{min} = 1 \) corresponds to a flat profile (i.e., no shaping used on slots).

The number of real time users has been set to 25 per cent of the total users. Two simulation runs of 1500 frames have been performed for each point of the numerical results. The user throughput is defined for each user as the fraction of correctly received packets normalized to the number of generated packets. When used, the admission control policy marks as useless all the \( NS \) slots whose actual power \( P_{act} \) before allocation exceeds 6 dB the power profile (for the system load relative to Figs. 7, 8 and 9) i.e., \( \Delta(\text{load}) = 6 \) dB. When not specified we refer to the scheduling strategy (3), which jointly consider channel and queues (QCS). We also refer in the results to a case (CS) where only channel state is considered, i.e. \( P = CS \) and \( TD_0 \rightarrow \infty \), for comparison purposes.

Figures 3 and 4 show the impact of power shaping on the different centralization schemes for \( R \) and \( NR \) users. No admission control scheme is used here. The proposed scheduling algorithm significantly improves the system capacity with respect to a simple scheme without power shaping, which is based on the EDD policy only, and the performance improvement due to an increasing centralization decreases when power shaping is used (i.e., the distance between totally distributed and partially centralized curves is smaller with power shaping). With a high propagation exponent that limits the effect of intercell interference the partially centralized allocation algorithm is able to cope with the interference also in absence of power shaping, which achieves a gain only on \( NR \) users if QCS is used with scheme 3-3 (see Fig.4). If scheduling is based only on CS, the difference between power shaping and non power shaping increases. For \( NR \) users (Fig.4), the effect of power shaping is more remarkable with respect to \( R \). Figures 5 and 6 display a situation when the load is much more heavy (about 0.61). Obviously performance get worst but our technique achieves a gain in the 1-1 scheme,
over no PS. In the 3-3 scheme a gain is achieved only for NR users. The figures also show the results obtained by using analytically derived power profile (the details of the derivation are not reported here): we can note that the behavior is similar to that of linear shaping, also due to the fact that the linear profile well approximates the analytic solution. Figures 7, 8 and 9 compare the performance of QCS and CS priority algorithm in terms of per user throughput and packet delay, in presence of admission control.

Fig.7 shows that the joint consideration of channel and queues state lead to a considerable gain in terms of throughput for RT users. This is very important because RT services are very sensitive to packet losses. This is done at the expense of NR users (Fig.9) whose performance results degrades in terms of throughput and delay.

From Fig.9 it can be noted that almost all users suffer low delays and QCS does not introduce a significant degradation over CS. The fraction of non admitted users is 0.11, whereas without power shaping would be about 0.39, which is unacceptable.

6 Conclusions

We investigated a new class of packet scheduling algorithms for TDMA-based cellular radio networks with full frequency reuse which assign the available resources according to the SIR predicted at each slot of the frame. This can be done by statically assigning in advance a maximum power level to each channel according to a set of power-profiles.

We showed that this technique is able to increase the capacity of systems with and without partially centralized resource management control and it is also able to fill the efficiency gap between partially centralized and fully distributed strategies reducing the need of coordination among cells. This result is obtained while maintaining the capability of providing acceptable quality of service to heterogeneous classes of users.

References

Figure 7: Comparison between the CS policy and the joint channel-queues QCS for real time users.

Figure 8: Comparison between the CS policy and the joint channel-queues QCS for non real time users.

Figure 9: Comparison between the simple CS policy and the joint channel-queues QCS for non real time users.


