

An Efficient Resource Allocation Strategy for Future Wireless Cellular Systems

Mohammad Abaii, Yajian Liu, and Rahim Tafazolli

Abstract—Future mobile communication systems will be designed to support a wide range of data rates with complex quality of service matrix. It is becoming more challenging to optimize the radio resource management and maximise the system capacity whilst meeting the required quality of service from user's point of view. Traditional schemes have approached this problem mainly focusing on resources within a cell and to large extent ignoring effects of multi-cell architecture. This paper addresses the influence of multi-cell interference on overall radio resource utilisation and proposes a novel approach, setting a new direction for future research on resource scheduling strategies in a multi-cell system. It proposes a concept called Load Matrix (LM) which facilitates joint management of interference within and between cells for allocation of radio resources. Simulation results show significant improvement in the resource utilization and overall network performance. Using the LM strategy, average cell throughput can be increased as much as 30% compared to a benchmark algorithm. Results also show that maintaining cell interference within a margin instead of a hard target can significantly improve resource utilization.

Index Terms—Resource allocation, scheduling, interference management, cellular radio

I. INTRODUCTION

MOBILE cellular systems are facing new challenges created by the demand for emerging services and applications. Wide range of services with diverse Quality of Service (QoS) requirements is becoming more popular and widely used. The demand of higher bandwidth and data rates has been increased substantially during recent years. This has made it so important for future mobile cellular systems to implement an efficient resource allocation scheme.

In order to achieve efficient resource utilization in all sorts of deployment scenarios and QoS requirements in the future wireless cellular systems, new resource allocation methods must be developed. In other words, resource allocation has to provide optimum or near optimum, for practical reasons, utilization of the available radio spectrum in the next generation of cellular wireless systems regardless of deployment scenarios and conditions. However, optimum point has to be found before considering the implementation practicality.

Importance of resource scheduling was appreciated with the support of high data rate services in the evolution of UMTS

Manuscript received April 12, 2006; revised December 11, 2006 and March 20, 2008; accepted April 22, 2008. The associate editor coordinating the review of this paper and approving it for publication was R. Negi. Part of this work has been presented in CCSR's Research Excellence Award Competition 2006 and received Research Excellence Award.

The authors are with the Mobile Communications Research Group, Centre for Communication Systems Research (CCSR) at the University of Surrey, GU2 7XH, United Kingdom (e-mail: M.Abaii@surrey.ac.uk).

Digital Object Identifier 10.1109/TWC.2008.060161.

standard [Rel99] to High Speed Downlink Packet Access (HSDPA) [1] and Enhanced Uplink [2].

A variety of resource allocation strategies and schemes, mainly for downlink, can be found in references [3]-[8]. In [3] a system with multiple traffic classes was considered and resource allocations were based on the specific characteristics of traffic flows resulting in minimization of power consumption or maximization of system capacity. Under mixed service traffic including both real-time and non-real time services, efficient resource allocation from a shared resource pool is a challenging task due to varied and stringent QoS requirements. In [4] authors proposed a fixed resource partitioning method in which total resource pool was partitioned between different service classes and independent resource schedulers were responsible for each resource partition whereas in [5], scheduling was more unified and partitioning was dynamic to enhance spectral efficiency. Another approach towards resource allocation, called utility based approach, tries to maximize the total network utility and thereby enhancing resource allocation. For example, pricing is a well-known utility function used in [6] for resource allocation. In [7] authors used user's QoS as utility function and then convert the resource allocation problem into a non-cooperative game where each user tries to maximize its own utility. A downlink resource allocation method based on dynamic pricing was proposed in [8] aiming to maximize the summation of users' utility.

On the link level, adaptive transmission is one of the most recent technologies being investigated for enhancing the spectral efficiency in future cellular systems [9]. Fast scheduling together with adaptive modulation-coding, facilitates exploitation of channel variations resulting in multi-user diversity gains [10]. This approach takes advantage of instantaneous channel conditions of different users where the channel fading are relatively independent. Adaptive transmissions are more effective for low mobility users compared with fast moving users' channel.

Uplink resource allocation methods can be categorized as centralized or decentralized in terms of the network location/node in which scheduling takes place. In Universal Mobile Telecommunications System (UMTS) for example, if the scheduler resides in Radio Network Controller (RNC), it is called centralized and if it resides in base station it is called decentralized.

In an interference-limited system such as UMTS, the uplink cell capacity is basically limited by the total received uplink power at the base station due to the transmit power limitation of user terminals [18]. In decentralized scheduling, each base

station assigns radio resources to its users on a priority basis until the estimated Rise over Thermal noise (RoT) level reaches a pre-defined target. Recent studies in Enhanced Uplink UTRA, also called High Speed Uplink Packet Access (HSUPA), show that the decentralized scheduling has better performance compared with centralized one [2]. The subject of centralized versus decentralized scheduling has been studied extensively in recent years both in 3rd Generation Project Partnership (3GPP) standard body for HSUPA and in the literature [2][26]. In [25] the performance of centralized packet scheduler of the UMTS system is evaluated while in [26] the performance of a decentralized scheduling is evaluated and compared with the centralized one in [25].

The basic advantage of decentralized over centralized approach is due to its fast response to dynamic and fast varying environment of mobile systems for resource allocation. However, the decentralized scheduling algorithms have an inherent shortcoming, due to their vulnerability to intercell interference, which has not been addressed yet. In other words, considerable proportion of RoT at the base station is made up from multiple access intercell interference which the base station has little knowledge about or control upon. This in turn may lead the system to interference outage and poor resource utilization particularly when interfering cells have similar traffic load variations.

Inadequate (intercell) interference management strategy particularly in highly loaded systems is an inherent problem of decentralized scheduling, regardless of the algorithm being used. Several interference mitigation techniques such as Multi User Detection (MUD) [20], Interference cancellation (IC) [21], antenna beamforming [22], and their combinations have been studied extensively and proved to be effective in mitigating interference to some extent and thereby increase system capacity. However, in a highly loaded system, the problem of intercell interference remains an important issue. For instance, MUD with Minimum Mean Square Error (MMSE) detection MMSE-MUD is recognised as an effective interference suppression technique for increasing the system capacity [23]. Yet it has been demonstrated in [24] that MMSE-MUD performs approaches that of a single user band in a fully loaded system.

From the scheduling perspective, although intercell interference problem is more severe in decentralized scheduling, it is also present in centralized scheduling due to the fact that the intercell interference impact of a scheduled user is not known and therefore has not been considered by the central scheduler.

In this paper, we address intercell interference problem of scheduling process by introducing a new and efficient resource allocation strategy called Load Matrix (LM). In order to prove the concept, we have used HSUPA system as a case study. Section II describes the destructing effect of intercell interference on HSUPA decentralized scheduling performance in terms of interference outage. In section III we will prove that optimum uplink resource allocation is in fact a Non deterministic Polynomial time (NP)-hard problem in both a single-cell and multi-cell cases. The Load matrix approach is detailed in section IV. Extensive simulation results on interference outage, throughput and packet delay performance of a reference decentralized scheduling [2][11] together with the proposed Load Matrix approach are provided and compared

in section V and finally section VI provides conclusion for this paper.

It should be noted that although HSUPA system is used to demonstrate the performance of the Load Matrix, the concept is generic for single-carrier spread spectrum based systems where cell RoT is widely used as a good load indication directly linked to cell load. In multi-carrier systems however, the load on subcarriers can differ significantly and therefore RoT (averaged) is no longer a good measure for load over all subcarriers. One can think of an effective RoT instead, encouraged by the introduction of effective SINR in multi-carrier systems to provide a better and more accurate link-system mapping [27]. Finding an effective RoT, however, is out of the scope of this paper and Load Matrix approach in multicarrier systems remains for further study. Applicability of Load Matrix in a system with adaptive techniques and multiuser scheduling is further explained in section IV.

II. IMPORTANCE OF INTERCELL INTERFERENCE CONTROL

Uplink cell capacity in interference-limited systems is basically limited by the total received power at the base station. As the uplink load increases, user terminals have to increase their transmit power substantially to overcome the increased interference level at the base station [18]. Due to the fact that the transmit power of user terminals is limited, total received power at the base station actually limits the uplink capacity.

In decentralized scheduling, each base station assigns radio resources (i.e. rate and time) to its users until the estimated RoT reaches a predefined target value, $\text{RoT}_{\text{target}}$. We assume $\text{RoT}_{\text{target}}$ is a fixed target value set by the network controller to maintain the uplink interference level [2]. The main shortcoming for decentralized scheduling in general becomes more visible in a multi-cell scenario where a considerable proportion of RoT is intercell interference and base station has little knowledge about and control upon. By intercell interference here, we mean intercell multiple-access interference in general i.e. any signal received by a base station coming from those users which belong to other cells. The relation between cell RoT and intercell interference is given in (1). The RoT of cell j (RoT_j) is defined as the total in-band received power at base station j (BS_j) over thermal noise. Let be the received noise power in BS_j , bs_j be the set of BS_j users, P_i be the transmission power of user i and $G_{i,j}$ be the channel gain from user i to BS_j . For M active users in the network, RoT_j can then be written as:

$$\text{RoT}_j = \left(\sum_{i=1, i \in \text{bs}_j}^M P_i G_{i,j} + \sum_{i=1, i \notin \text{bs}_j}^M P_i G_{i,j} + N' \right) / N' \quad (1)$$

For simplicity, we have not considered soft hand over in this paper meaning the user is connected only to one base station at a time.

To highlight the importance of intercell interference problem and show the impact of it in the overall interference outage performance and resource utilization, we set up a specific simulation scenario. In this scenario, all the cells have same traffic distribution and interference condition (i.e. identical user distribution per cell). In each cell 10 users with full buffer

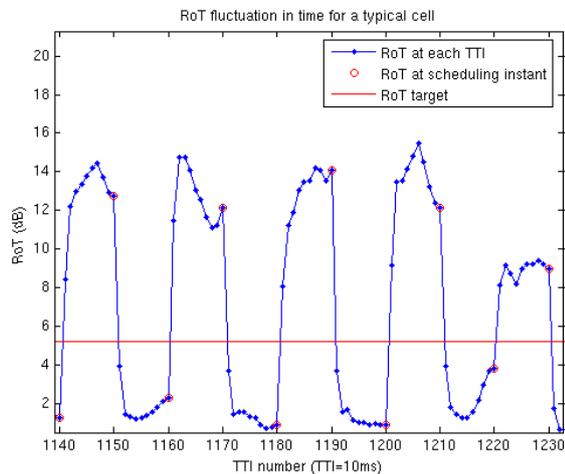


Fig. 1. RoT fluctuation in a multi-cell scenario.

are waiting for transmission. Same user distribution makes the scheduling order and therefore generated interference exactly the same in all cells. We have chosen this extreme scenario to highlight the effect of intercell interference in its extreme situation as the worst case scenario.

We also use wrap-around technique, which makes all the cells identical. Users are randomly and uniformly distributed over one cell and then repeated with identical pattern over other cells. Other simulation parameters are same as in table II. We use the decentralized scheduling algorithm in [11] which allocates resources individually in each cell. Assumption of full buffer occupancy for users helps to model and study the network behaviour under heavy traffic load.

Figure 1 shows the RoT fluctuation of a typical cell in this scenario for a period of 1 second. Transmission Time Interval (TTI) is considered 10ms. At TTI= 1140, RoT level is way below the RoT_{target} so scheduler decides to allocate resource to more users unaware that all other cells will do the same thing. The consequence of this decision is much higher RoT than expected in the next scheduling interval. At TTI= 1150, the opposite happens; RoT level prior to scheduling instant is way above the RoT target so scheduler severely decreases the amount of allocated resources unaware that all other cells will do the same thing. This phenomenon continues and RoT takes a pulsy shape far from RoT_{target} as shown in figure 1.

As mentioned earlier, this is the worst case scenario in respect to the intercell interference problem. Nevertheless, some degree of fluctuation (less than the case shown in figure 1) has been observed in RoT regardless of the scheduling type or algorithm being used. RoT fluctuation obviously gets worse when the traffic load increases. We have assumed an extreme scenario where the load in neighbouring cells varies in a synchronised way which is not realistic. In reality however, in peak hour it is expected that neighbouring cells are all facing very high traffic (i.e. users with full buffers waiting for transmission) at the same time. Also it is likely to have similar traffic behaviour in neighbouring cells since they are geographically close and therefore may have users with similar social environments and activities.

This case study clearly shows that intercell interference is a crucial factor which can not be ignored in the scheduling process for future wireless cellular systems.

III. RESOURCE ALLOCATION PROBLEM

The aim of resource allocation in wireless cellular system is to assign radio resources to individual users in a way to achieve maximum system capacity whilst meeting the required quality of service. In this section, we formulate the resource allocation problem in wireless cellular system and show that it is an NP-hard problem [13]. We consider a basic scenario where resource allocation is down to assigning transmission rate and time to individual users with the objective of throughput maximization. To analyze the problem, we begin with the single cell scenario and then extend the conclusion to the multi-cell case. Without loss of generality, we assume that transmission rates are chosen from a limited set of rates.

Let $S_{i,1}$ denote Candidate Rate Set (CRS) of user i , which includes all the allowed transmission rates for the user to choose from. Rate "0" is always included in $S_{i,1}$, and will be chosen if the user is not scheduled to transmit in the current scheduling instant. We treat transmission rates in different CRSs as different items even if they have the same rate value:

$$S_{i,1} \cap S_{j,1} = \Phi \quad \forall i \neq j \quad (2)$$

Let S_1 denote the union of all the CRSs from $S_{1,1}$ to $S_{M_1,1}$, and M_1 is the total number of users in the cell sharing the radio resource pool. Choosing an element t from set S_1 is an assignment action, which means allocating a specific transmission rate to a particular user. Apparently, each assignment action generates a certain amount of throughput while consumes some amount of the cell capacity. We use binary variable x_t to indicate whether element t is chosen or not (1 for 'Yes' and 0 for 'No'). p_t and $c_{t,1}$ denote the generated throughput and consumed cell capacity respectively if element t is chosen. p_t is equal to the transmission rate itself, whereas $c_{t,1}$ can be interpreted differently, e.g. as consumed BS transmit power or generated load factor [12], depending on the system type.

Using above terms and definitions, the Single Cell Radio Allocation Problem (SCRAP) can be described as follows: given a particular system snapshot (cell capacity, user location, propagation and traffic status etc.), how to choose elements from set S_1 in each scheduling instant so as to achieve maximum system throughput, subject to the following two constraints: C1 : The aggregated cell capacity consumption of all the chosen elements from S_1 should be less than the total available cell capacity Cap_1 . C2 : for each CRS ($S_{1,1}$, ..., $S_{M_1,1}$), only one element is chosen. Mathematically, SCRAP can be formulated as follows:

$$\text{maximize : } p = \sum_{t \in S_1} p_t x_t \quad (3)$$

subject to:

$$\sum_{t \in S_1} c_{t,1} x_t \leq Cap_1 \quad (4)$$

$$\sum_{t \in S_1} x_t = 1 \quad i = 1, \dots, M_1 \quad (5)$$

$$x_t = \begin{cases} 1 & \text{if } t \text{ is chosen} \\ 0 & \text{if } t \text{ is not chosen} \end{cases}, t \in S_1 \quad (6)$$

Theorem: The SCRAP is NP-hard.

Proof: We show SCRAP is in fact the Multi-Choice Knapsack Problem (MCKP) which has been proven to be NP-hard [14]. MCKP can be expressed as follows:

Given a knapsack, an item set, and a partition of the item set into a number of subsets, how we choose items so as to maximize the total profit from all the chosen items, while the aggregate weight of all the chosen items is less than the allowed weight bearing of the knapsack. There is also a condition that only one item is chosen from each item subset. SCRAP can be mapped to MCKP by regarding the knapsack capacity as the available cell capacity, the item set as the set S_1 , and the item subsets as CRSs $(S_{1,1}, \dots, S_{M,1})$. Also p_t and $c_{t,1}$ can be regarded as the profit and weight of element t respectively.

In MCKP all the components of the problem including knapsack capacity, items, profits, etc. are known and the question is which items to put in the knapsack. Similarly in the SCRAP, the only unknown in equations (3)-(6) is x_t . p_t is known because it is equal to the transmission rate itself. In CDMA uplink systems, $c_{t,1}$ and Cap_1 can be interpreted as the load factor of the given user (e.g. 10%) and the maximum load threshold of the cell (e.g. 70%) respectively. Then the following way can be used to calculate $c_{t,1}$ for each element t before the rate allocation actually takes place: 1) find out the target SINR at the receiver based on the given data rate and Block Error Rate (BLER). These target SINRs can be obtained from the BLER versus SINR performance curves in the physical layer (also known as link to system mapping); 2) This target SINR can then be transformed into the load factor, i.e. c_t , with the equation: load factor = $\text{SINR}/(1+\text{SINR})$ [12]. Moreover, in this single-cell scenario considered, no intercell interference exists, therefore the available cell capacity Cap_1 can be regarded exactly the same as the given cell capacity, which is assumed to be a fixed and pre-known value in our work. Furthermore, CRSs $(S_{1,1}$ to $S_{M,1})$ are pre-known sets based on the system restrictions, power headroom and queue status of individual users. Therefore, SCRAP is a type of Multi-Choice Knapsack Problem and NP-hardness of it follows by a trivial transformation from the MCKP.

We now consider the multi-cell case of the problem (MCRAP), where the interested area is covered by N cells, with total number of M users (i.e. $\sum_{j=1}^N M_j = M$ with M_j representing the number of users in cell j). In this case, $c_{t,1}$ and Cap_1 are upgraded to $c_{t,j}$ and Cap_j where the subscript j represents the cell index:

$$\sum_{t \in S_j} c_{t,j} x_t \leq \text{Cap}_j \quad \forall j = 1, \dots, N \quad (7)$$

In which $c_{t,j}$ is the consumed capacity from cell j if element t is chosen. The expression of MCRAP is very similar to SCRAP except that constraint C1 should be replaced by the following C1' in order to take all the cells into account: C1': For any cell j ($j = 1, \dots, N$), the total cell capacity consumption by all the chosen elements from S_j should be less than the total available cell capacity Cap_j . In fact, SCRAP was the simplest case of MCRAP where $N=1$. Consequently,

if SCRAP is not solvable by a polynomial time algorithm, neither is MCRAP and therefore MCRAP is also NP-hard.

IV. LOAD MATRIX CONCEPT

One of the main challenges in resource allocation in a multicell system is the control of intercell interference. In uplink scheduling, the basic problem is to assign appropriate transmission rate and time to all active users in such a way that result in maximum radio resource utilization across the network whilst satisfying the QoS requirements of all the users. Amongst other constraints, another important factor in the resource allocation is the user's transmit power. For a network of M users and N cells the constraints to be satisfied are Cnst1 : For each active user i in the network, its transmit power P_i must be maintained in an acceptable region defined as

$$0 \leq P_i \leq P_{-,max} \quad i \in \{1, \dots, M\} \quad (8)$$

$P_{i,max}$ is the maximum transmission power available to the user due to hardware limitation and/or any other restrictions. Cnst2 : The total received power at base station should be kept below a certain threshold for all N base stations in the network (as a load control measure in a single-carrier Spread Spectrum system). We use Rise over Thermal noise (RoT) as defined in HSUPA [2] to represent the interference constraints. RoT_j is the total in-band received power at the base station j (BS_j) over thermal noise. Let N' be the receiver noise power in a BS, P_i be the transmission power of user i and $G_{i,j}$ be the channel gain from user i to BS j . For M active users in the network, RoT_j can be written as

$$\text{RoT}_j = \left(N' + \sum_{i=1}^M P_i G_{i,j} \right) / N' \quad (9)$$

In this case Cnst2 can be formulated as

$$\text{RoT}_j \leq \text{RoT}_{\text{target}} \quad j \in \{1, \dots, N\} \quad (10)$$

where $\text{RoT}_{\text{target}}$ is assumed to be a fixed target value set by the network operator to maintain the uplink interference level.

Cnst3 : For each user, depending on its channel type (e.g. pedestrian, vehicular) and speed, each rate k has a minimum required SINR called $\text{SINR}_{\text{target},k}$. $\text{SINR}_{\text{target},k}$ is the signal to noise plus interference ratio required at the serving base station j if rate k is being assigned to the user in order to achieve a given block/frame error rate. Rate $k \in \{1, \dots, K\}$ is the highest rate acceptable (and therefore is the preferred rate) for user i with serving base station j if $\text{SINR}_{\text{target},k}$ is the highest that can be achieved under both Cnst1 and Cnst2

$$\text{SINR}_{i,j} \geq \text{SINR}_{\text{target},k} \quad i \in \{1, \dots, M\}, k \in \{1, \dots, K\} \quad (11)$$

Load Matrix (LM) can be regarded as a database containing the load factors of all active users in the network. LM scheduling can be implemented in both centralized and decentralized strategies. In a decentralized LM scheduling, each base station should implement identical LM database. Here for simplicity, we only present the centralized LM scheduling where a central scheduler entity assigns radio resources to all the users in the network. Figure 2 illustrates an example of LM scheduling implementation based on the proposed system architecture for

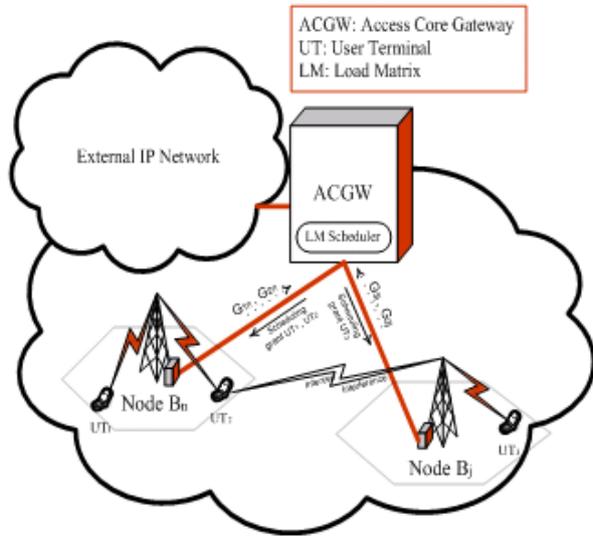


Fig. 2. Centralized LM scheduling in a 3G LTE system.

the 3rd Generation Long-Term Evolution (3G LTE) [17]. We assume the averaged channel gain (over the scheduling period) from users to base stations is known to scheduler prior to rate assignment. In a network of M users and N cells, $LM_{i,j}$ is the load factor [12] contributed by user i at BS j

$$LM_{i,j} = \frac{P_i G_{ij}}{N' + \sum_{m=1}^M P_m G_{mj}} \quad (12)$$

From the $LM_{i,j}$ values stored in column j of LM database, RoT of cell j can be written as

$$RoT_j = \frac{1}{1 - \sum_{i=1}^M LM_{i,j}} \quad (13)$$

Note that RoT_j obtained from (13) is identical to RoT_j definition given in (9). Let BS j be the serving base station for user i which controls user's transmission power and $G_{i,j}$ be the total channel gain from user i to BS j averaged over scheduling period. $SINR_{i,j}$ can be written as

$$SINR_{i,j} = \frac{P_i G_{i,j}}{N' RoT_j - P_i G_{i,j}} \quad (14)$$

Let $p_{i,k}$ be the required transmit power for user i should it be assigned the rate k . Starting with the highest applicable rate from the set (i.e. $k = K$), for rate k to be assigned to user i , (11) must be satisfied. Therefore the minimum required $SINR_{i,j}$ is $SINR_{target,k}$. Also recall the aim of LM rate assignment to keep RoT_j always as close as possible (ideally equal) to RoT_{target} . By rearrangement of (14), replacing $SINR_{i,j}$ with $SINR_{target,k}$, P_i with $p_{i,k}$, and RoT_j with RoT_{target} , required transmit power for user i (should it be assigned rate k) can be found as

$$p_{i,k} = \frac{N' RoT_{target}}{G_{i,j}} \frac{SINR_{target,k}}{1 + SINR_{target,k}} \quad (15)$$

However, rate k and consequently $p_{i,k}$ is acceptable if and only if all three constraints are satisfied. First of all, $p_{i,k}$ obtained from (15) must satisfy Cnst1 which states the maximum user's transmit power. Cnst3 constraint is already satisfied

by considering $SINR_{target,k}$ as $SINR_{i,j}$. Additionally, $p_{i,k}$ must satisfy the Cnst2 which takes into account the impact of assigning rate k to user i on the intercell interference. This ensures the intercell interference caused by user i in other cell does not increase other cells' RoT above RoT_{target} . In order to check this, next step is to update $LM_{i,n}$ for all the elements in row i of Load Matrix. $LM_{i,n}$ is the load factor imposed by user i in cell n using rate k defined as

$$LM_{i,n} = \frac{P_{i,k} G_{in}}{N' + \sum_{m=1}^M P_m G_{mn}} \quad (16)$$

From (13) one can estimate the new RoT for all other cells and check if the Cnst2 constraint has been satisfied. If so the rate k is the highest acceptable rate for user i and will be assigned, otherwise the same process is repeated for the rate $k-1$ and so on. If at the end, none of the rates $k \in \{1, \dots, K\}$ can satisfy the three constraints, user i will not be scheduled for transmission at this scheduling instant and will be given higher priority for the following scheduling instant. After the first round of rate assignment to all users, LM elements are updated and new RoT is calculated for each cell using (13). This is necessary because (15) and (16) are valid only if RoT is close to the RoT_{target} . Since the rate assignment is an NP-complete problem (see section III), it is not possible to exactly achieve RoT_{target} in all cells in the first round of rate assignment. This requires additional rounds (which we refer to it as iterations) of rate/power adjustments in order to minimize the difference between a cell RoT and its RoT_{target} . In other words, at each scheduling instant, $p_{i,k}$ is iteratively adjusted in (15) and then (16) by replacing RoT_{target} with updated RoT from LM in (13) after each round of rate assignment. This check is an important step ensuring low probability of interference outage by keeping RoT below RoT_{target} and at the same time increasing resource utilization with highest cell RoT possible (i.e. RoT close to RoT_{target}). Obviously the number of iterations depends on the difference between RoT and RoT_{target} at the end of each scheduling process. However, the simulation results presented in section V are with no iteration and yet the difference between RoT and RoT_{target} was found to be negligible. It should be noted that if a user is not in the full buffer status, the maximum rate index K in (11) is limited to a rate that would result in emptying the buffer at the next scheduling instant. Cell or network throughput per bandwidth (bps/Hz/cell) is often taken as resource utilization measure. However, there is a trade off between maximum cell throughput and fairness amongst users [15]. Priority functions are used to rank users in the scheduling process and make a balance between cell throughput and fairness. Commonly used priority functions are Round Robin, DL SINR (also called Max C/I), Proportional Fair (see e.g. [1][2][16]) and also most recently introduced Score-Based [19]. The Round Robin tries to maximize fairness amongst users regardless of a user channel condition and therefore results in poor throughput performance. Max C/I on the other hand, ranks users in terms of their channel quality and aiming for maximizing cell throughput at the expense of fairness. Both Proportional Fair and Score-based functions performance are better than Round Robin in terms of throughput and better than Max C/I in terms of fairness. Load Matrix concept,

Scheduling procedure
Initialize $LM[i,j]$ to zero;
for cell $j = 1 : N$
for user i in cell j
set priority $[i,j]$ as in (18);
end
sort all users in Cell j according to priority $[i,j]$
end
for assignment round = 1: Number of users per cell
for cell $j = 1 : N$
for the highest priority user i in cell j if $LM[i,j]=0$
for $Rate_{min} = K_{i-1}$
get $P_{i,j}$ for Rate $_{min}$ as in (14);
if condition 3 is already satisfied
next rate if condition 1 is not satisfied;
capacity check $RoT[i,j]$ with "intra-cell-margin";
next rate if condition 2 is not satisfied;
capacity check for all other cell $RoT[i,j]$
with "inter-cell-margin";
next rate if condition 2 is not satisfied;
update $LM[i, j]$;
end
end
remove user i from sorted users of cell j
end
end

Table 1. Load Matrix Algorithm.

provides a generic solution for resource scheduling that does not preclude any priority function and can be combined with any of them. However, priority function has major impact on overall system performance for any scheduling algorithm including the LM. Here a priority function is introduced based on a user's load vector that includes intra and intercell impact on the network. It is evident that giving priority to a user with better channel condition increases the cell throughput but in a multicell network could have severe impact on the throughput of other cell's. Here it is considered by defining Global Proportional Priority function as

$$\text{priority}_i = \frac{G_{i,j}}{\sum_{n=1, n \neq j}^N G_{i,n}} \quad \forall i \in \{1, \dots, M\} \quad (17)$$

where $G_{i,j}$ is the total channel gain from user i to BS j averaged over the scheduling period. The LM approach tries to maximize network capacity through inter and intracell interference management. Table I summarises LM algorithm used for rate assignment in a multicell wireless system. The first step is initialization where all the LM elements are set to zero and also users in each cell are sorted according to the priority function in (17). The LM allocation process simultaneously increases allocated resources in each cell to avoid interference imbalance amongst the cells. The process consists of a number of assignment rounds equal to the maximum number of users per cell (e.g. 10 rounds for 10 users per cell as stated in Table II). In each round, the LM assigns rates to the highest priority user in each cell, updates LM elements and performs capacity checking. Capacity Check (CC) function calculates RoT as in (13) and compares with RoT_{target} making sure (10) is always satisfied. Passing this "check" means assigned rates are valid and will not cause interference outage. If CC fails, scheduler attempts the next available rate and continues until CC is satisfied. Then the user is considered scheduled and will be removed from the user priority list of its serving cell. The scheduling process continues until all the users are processed. In the LM scheduling process, the CC function is especially important and plays a major role in the overall system performance. In particular, a margin concept rather than a fixed threshold for RoT_{target} is introduced. The CC operates on this small margin around RoT_{target} instead of a fixed RoT_{target} threshold. Two independent margin variables

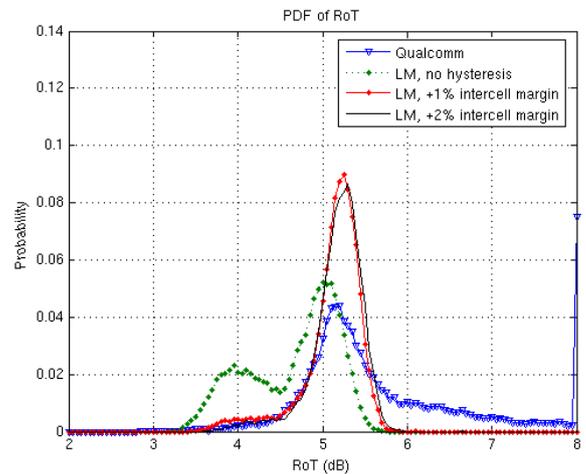


Fig. 3. PDF of RoT (inter-cell margin effect).

called intercell margin and intracell margin are specified for handling intercell and intracell interference respectively and assist better decision during the CC process. The intracell margin is a region set around the RoT_{target} where a serving cell's user loading should not exceed. Similarly, intercell margin is another region specified around RoT_{target} which limits variations of overall RoT caused by a user from other cells. The intercell and intracell margins can be equal or different resulting in different performances. It is shown in section V that maintaining RoT in a small margin around RoT_{target} instead of using a single RoT_{target} threshold will result generally in a much improved interference outage performance and higher resource utilization. The concept of the Load Matrix can be used in conjunction with other adaptive techniques and priority functions. In single carrier Spread Spectrum systems, RoT_{target} defines the maximum cell capacity even though the instantaneous capacity in a cell is not fixed. In other words, the CC function explained above regards a cell as fully loaded only when its estimated RoT reaches RoT_{target} . That means cell capacity is not a direct function of the transmission rates being assigned. The LM concept operates on RoT , channel gains and the specified constraints such as user power and available rates. These are common parameters in CDMA based mobile cellular systems although their calculations are different for different air interface standards, and dependant on the network architecture. In this paper, to demonstrate the performance of the Load Matrix, the HSUPA standard was chosen for this purpose. HSUPA and HSDPA were the first standards which seriously address the importance of the scheduling problem.

V. SIMULATION RESULTS

To evaluate the performance of the LM concept, extensive system level simulations have been carried out using HSUPA [2]. The results are compared with the scheduler in [11] that is also used as a benchmark for comparison in [2]. Simulation parameters used for both the benchmark and the LM algorithms are shown in Table II. For simplicity, Hybrid Automatic Repeat reQuest (HARQ) is not considered here in

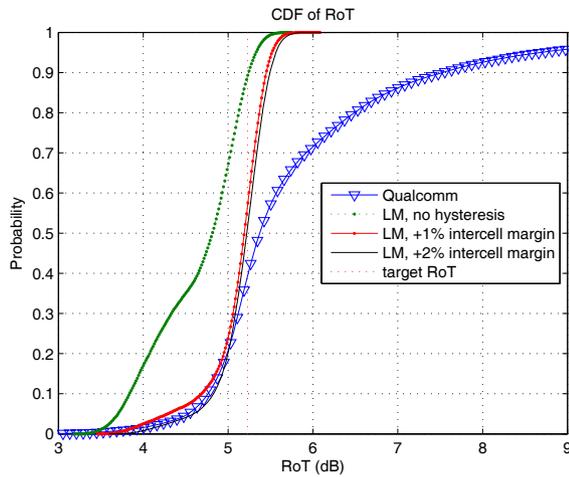


Fig. 4. CDF of RoT (inter-cell margin effect).

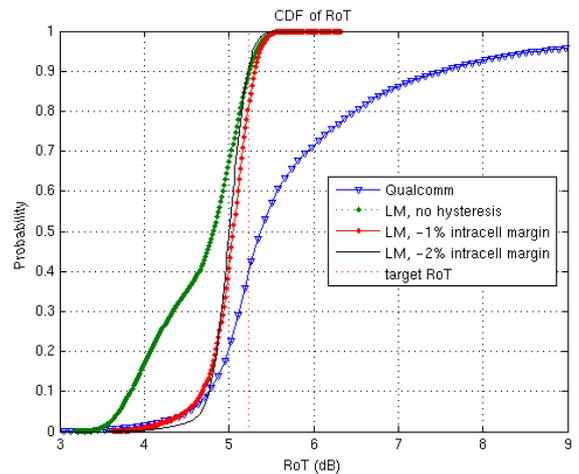


Fig. 6. CDF of RoT (intra-cell margin effect).

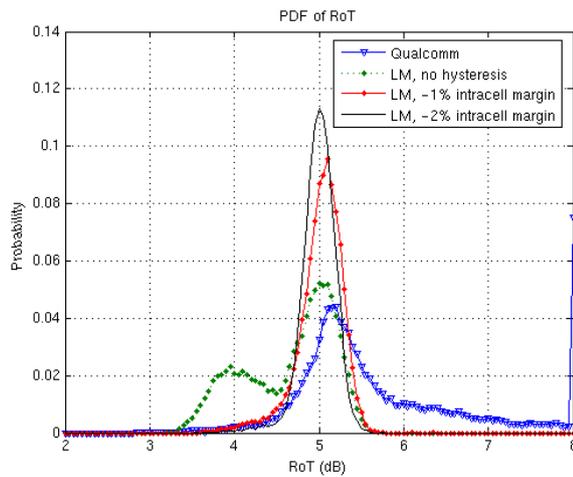


Fig. 5. CDF of RoT (intra-cell margin effect).

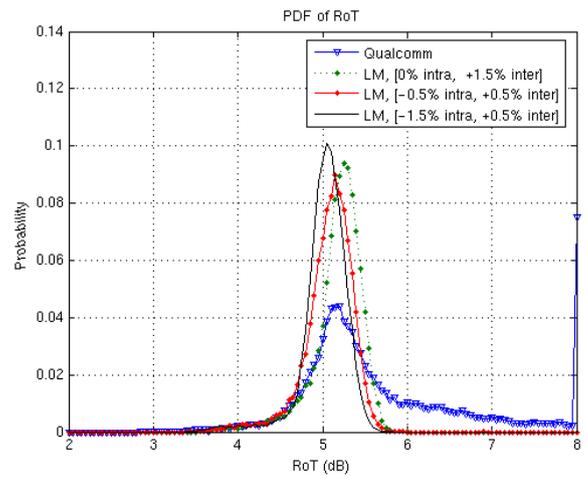


Fig. 7. PDF of RoT (best three combinations)

neither algorithms. It should be emphasised that the main objective in this paper is to highlight the impact of other cell interference existing in both centralized and decentralized scheduling algorithms. Another important objective is to show the performance of the scheduling algorithms compared with the upper-bound limit rather than comparison between different algorithms. Comparison with the upper-bound limit is a better indication of scheduling algorithm efficiency. Here the upper-bound limit on the interference outage performance is defined as a “step function” in CDF of RoT (as represented by target RoT in figures 4,6 and 8). Interference outage performance directly affects all other performance measures like throughput and packet delay. The comparison with the benchmark algorithm is provided here as an example to show the effectiveness of the LM scheduling compared with a typical scheduling algorithm used in [2]. General comparison between centralized and decentralized scheduling is already available in literature e.g. in [26]. The system level simulator models 19 omni-directional cell structure with 10 users per cell randomly and uniformly distributed. The resource allocation performance is carried out in terms of interference outage

probability, averaged cell throughput and packet delay. The simulation results provided here have two different objectives. The first objective is to show the impact of the margin concept (both intercell and intracell) on the interference outage performance. These are shown in figures 3-6. The second objective is to illustrate the performance of the LM (based on the best margin setup) compared with the benchmark algorithm and the upper-bound limit in terms of interference outage. Figures 7 and 8 show the performance in terms of interference outage, whereas service throughput and packet delay performances are depicted in figure 9 and figures 10-11 respectively. The figures 3 and 4 where no intracell margin was considered, show the Probability Density Function (PDF) and Cumulative Density Function (CDF) of RoT respectively for benchmark algorithm and three LM cases. The LM case one represents no intercell margin i.e. “no hysteresis” while the other two cases represent intercell margins of +1% and +2%. These figures demonstrate considerable performance in interference outage compared with the benchmark algorithm and a close performance to the upper-bound limit indicating the sensitivity of performance to intercell margin. Same behaviour can be observed in figures

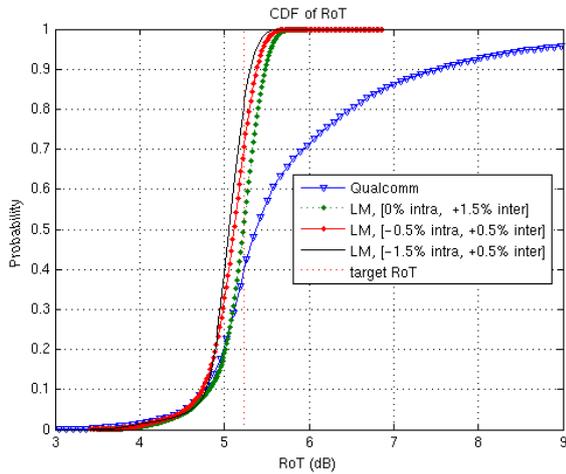


Fig. 8. CDF of RoT (best three combinations).

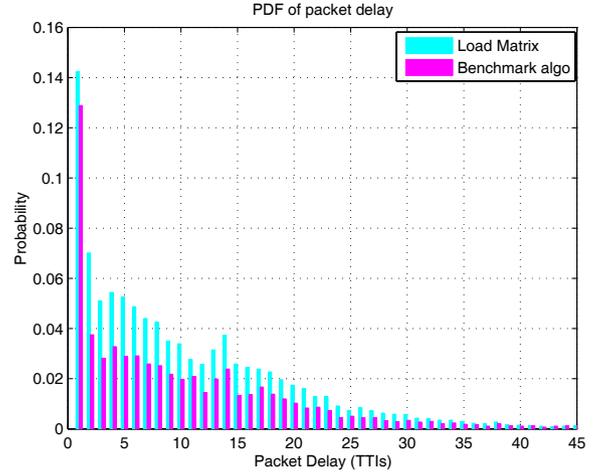


Fig. 10. PDF (Histogram) of Packet delay.

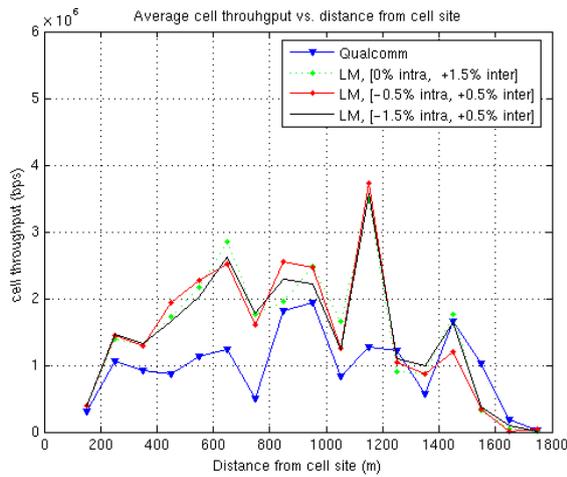


Fig. 9. Average service throughput versus distance.

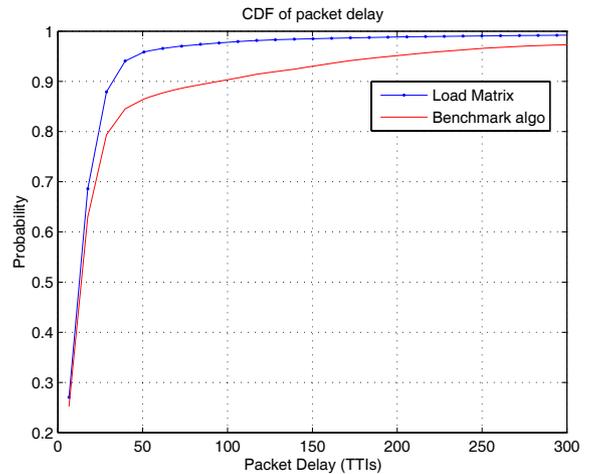


Fig. 11. CDF of Packet delay.

5 and 6 where intracell margin is used instead of intercell. It should be noted that intercell and intracell margins has positive and negative values with respect to RoT_{target} . This is due to the fact that under all the conditions, the intracell interference is more dominant than the intercell interference. Therefore the LM capacity check has to be more strict with own cell users contribution to cell loading by not permitting RoT to exceed the RoT_{target} . However, it can show more flexibility in accommodating the other cell interference allowing their contribution to RoT exceed the RoT_{target} but within the specified intercell margin. In order to find an appropriate set of intercell and intracell margin, we have carried out simulations for all 25 combinations of intercell margin = [0, 0.5%, 1%, 1.5%, 2%] and intracell margin = [0, 0.5%, 1%, 1.5%, 2%]. Figures 7 and 8 represent the interference outage performance of the best three combinations. The LM approach outperforms the benchmark algorithm, in all cases, with a very close performance to the upper-bound limit. The figure 9 shows the service throughput versus distance from a base station for the LM and the benchmark algorithms. It can be observed that in terms of cell throughput performance, the LM outperforms

the benchmark algorithm on average by more than 30%. The spikes in the cell throughput are due to the limited number of users generating transmit data. The figures 10 and 11 show the packet delay performance for the LM and the benchmark algorithms. This again demonstrates another performance improvement. Figure 11 shows that with the LM approach, 95% of the packets experience delay of less than 40 TTI compared with 200 TTI experienced by the benchmark algorithm. It is worth noting that all the LM results presented in this paper are achieved with no iteration, as described in section IV, resulting in negligible interference outage as can be observed when comparing with the upper-bound limit.

VI. CONCLUSION

A novel approach towards efficient resource allocation for future wireless cellular systems was presented. The vulnerability of traditional resource allocation and scheduling schemes to intercell interference resulting in interference fluctuations was demonstrated. Such interference fluctuation results in capacity wastage and excessive packet delay performance. The Load Matrix concept presented addresses this problem

Parameter	Explanation
system layout	Hexagonal grid, omni sites, 3 tiers (19 base stations) wrap around
Number of users	190 (10 users per cell)
Cell radius R	1.8 km
Path loss model	$L = 128.1 + 37.6 \text{Log}_{10}(R)$
Channel	AWGN + shadowing
Std. deviation of slow fading	8.0 dB
Correlation distance of slow fading	50 m
BS antenna gain plus Cable Loss	14 dBi
Carrier frequency	2 GHz
Rx antenna	1
Tx antenna	1
User antenna gain	0 dBi
Maximum User EIRP	21 dBm
Maximum BS EIRP	24 dBm
CL Power Control	1 dB step size
Transmission rates (kbit/s)	8,16,32,64,128,256,384
TTI	10 ms
Scheduling period	Every 10 TTI
Traffic model	Full buffer, video streaming (3GPP H.263)
Number of UEs per cell	10
Simulation time	20 s
RoT target	5.23 dB (= %70 load factor)
Number of iterations (in Load Matrix approach)	No iteration

Table 2. Simulation Parameters.

specifically and provides an efficient resource allocation by jointly considering intercell and intracell interference before making decision on allocating radio resources. Extensive simulation results for the LM demonstrated its capability through considerable performance improvements over the benchmark scheduling algorithm in terms of both averaged packet delay and cell throughput. The performance was also shown to be very close to the upper-bound limit of interference outage. By incorporating a new concept of separate margins for intercell and intracell interferences into the LM, it was shown that better control over such interferences can be provided resulting in high overall network performance. As an example, in the LM approach interference can be always kept close to the specified target whilst average cell throughput can be increased by more than 30% compared with a well known benchmark scheduling algorithm. Although HSUPA system level simulator is used in this paper to produce the performance results, the concept is generic for single-carrier spread spectrum based systems. In case of multi-carrier systems, the load on subcarriers can differ significantly and therefore RoT (averaged) is no longer a good measure for load over all subcarriers. It will be interesting to investigate the possibility of finding an effective RoT instead, to provide a better and more accurate measure for cell load in those systems. Effective RoT and application of Load Matrix in multicarrier systems remains for further study.

REFERENCES

- [1] 3GPP TR 25.848, "Physical layer aspects of UTRA High Speed Downlink Packet Access V4.0.0," www.3gpp.org, Mar. 2001.
- [2] 3GPP TR 25.896, "Feasibility Study of Enhanced Uplink for UTRA FDD V6.0.0," www.3gpp.org, Mar. 2004.
- [3] I. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Trans. Veh. Technol.*, vol. 45, no. 6, pp. 506-519, May 2000.
- [4] M. Ono, Y. Matsunaga, M. Momona, K. Okanou, "A proposal of all-IP mobile wireless network architecture QoS packet scheduler for base stations," Tech Rep. IEICE, MoMuC2002-3, 2002-05.
- [5] S. Abedi, "Efficient radio resource management for wireless multimedia communications: a multidimensional QoS-based packet scheduler," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2811-2822, Nov. 2005.
- [6] E. P. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecommun.*, vol. 8, pp. 33-37, Jan. 1997.

- [7] C. U. Saraydar, N. B. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1883-1892, Oct. 2001.
- [8] M. Shabany, K. Navaie, E. S. S. Soma, "Downlink resource allocation for data traffic in heterogenous cellular CDMA networks," in *Proc. ISCC 2004*, vol. 1, pp. 436-441, July 2004.
- [9] M. Sternad, S. Falahati, T. Svensson, D. Aronsson, "Adaptive TDMA/OFDMA for wide-area coverage and vehicular velocities," in *Proc. 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.
- [10] M. Sternad, T. Ottosson, A. Ahl'en, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDMA downlinks," in *Proc. IEEE Vehic. Tech. Conf. (VTC)- 2003 Fall*, vol. 4, pp. 2486-2490, Orlando, FL, Oct. 2003.
- [11] 3GPP R1-031246, "Reference Node-B Scheduler for EUL," Qualcomm Europe, www.3gpp.org, Nov. 2003.
- [12] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, second edition. John Wiley & Sons, 2002..
- [13] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: Freeman, 1979.
- [14] S. Martello and P. Toth, *Knapsack Problems: Algorithm and Computer Implementation*. John Wiley & Sons, 1990.
- [15] A. Pandey, S. Emeott, J. Pautler, and K. Rohani, "Application of MIMO and proportional fair scheduling to CDMA downlink packet data channels," in *Proc. IEEE Vehic. Tech. Conf. (VTC)- 2002 Fall*, vol. 2, pp. 1046-1050, Vancouver, Canada, Sept. 2003.
- [16] 3GPP R1-031004, "Example of Rel-99 TFC control algorithm," Qualcomm Europe, www.3gpp.org, Oct. 2003.
- [17] Hannes Ekström *et al.*, "Technical solutions for the 3G long-term evolution," *IEEE Commun. Mag.*, pp. 38-45, Mar. 2006.
- [18] H. Holma and J. Laakso, "Uplink admission control and soft capacity with MUD in CDMA," in *Proc. IEEE Vehic. Tech. Conf. (VTC)-1999 Fall*, pp. 19-22, Amsterdam, Netherlands, Sept. 1999.
- [19] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," in *Proc. 5th European Wireless Conference*, Feb. 2004.
- [20] S. Moshavi, "Multi-user detection for DS-SS communications," *IEEE Commun. Mag.*, vol. 34, pp. 124-136, Oct. 1993.
- [21] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code division multiple access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509-519, Apr. 1990.
- [22] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Mag.*, pp. 67-94, July 1996.
- [23] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683-690, May 1990.
- [24] J.-F. Hlard, J.-Y. Baudais, and J. Citerne, "Linear MMSE detection technique for MC-CDMA," *IEE Electron. Lett.*, pp. 665-666, vol. 36, no. 7, Mar. 2000.
- [25] K. Dimou, C. Rosa, T. B. Sorensen, J. Wigard, and P. E. Mogensen, "Performance of uplink packet services in WCDMA," in *Proc. IEEE Vehic. Tech. Conf. (VTC)- Spring 2003*, vol. 3, pp. 2071-2075, Apr. 2003.
- [26] C. Rosa, J. Outes, K. Dimou, T. B. Sorensen, J. Wigard, F. Frederiksen, and P. E. Mogensen, "Performance of fast node B scheduling and L1 HARQ schemes in WCDMA uplink packet access," in *Proc. IEEE Vehic. Tech. Conf. (VTC)- Spring 2004*, vol.3, pp. 1635-1639, May 2004.
- [27] E. Tuomaala and H. Wang, "Effective SINR approach of link to system mapping in OFDM/multi-carrier mobile network," in *Proc. 2nd Int. Conf. on Mobile Technology, Applications and Systems*, Nov. 2005.



Mohammad Abaii received his B.Sc. in Telecommunications from Tehran University in 1992 and M.Sc. in mobile communications from IUST, in 1995. From 1998 to 2000 he was senior systems engineer developing software for physical layer algorithms in Wifi and WCDMA platforms as well as network dimensioning in system level. He joined Genista Research Pte Ltd in Singapore in 2000 as senior system engineer developing novel algorithms in perceptual QoS for mobile communications in which he participated in several patents and publications.

From 2002 to 2004 he was with BBS Access, Singapore, as Network Planning expert and later as Network planning manager. From May 2004 to December 2007, he was a research fellow and project coordinator for IST European project WINNER at CCSR, University of Surrey, UK, developing new scheduling algorithms for B3G systems including HSUPA and OFDMA-based systems. He is currently project/technical manager for several projects at Mycom U.K Ltd mainly responsible for system Performance Management for international mobile operators such as Vodafone and O2. His research interests include radio propagation, radio resource management, and packet scheduling for 3G and B3G mobile cellular systems.



Yajian Liu received the B.Eng. and M. Eng. degrees from the School of Electrical and Information Engineering, Hunan University, China, in 1998 and 2001 respectively, and the Ph.D. degree from Centre for Communication Systems Research, University of Surrey, U.K., in 2008. He is currently working for ip.access Ltd., Cambridge, U.K., mainly responsible for system specifications and radio resource management algorithm designs for 3G home base station systems.

From 2001 to 2003, he was a full-time researcher on mobile communication systems in Shanghai R&D Center of Huawei Technologies Co. Ltd., China. From 2003 to March 2007, he was actively involved in IST European projects MUMOR, WINNER, and FIREWORKS. From April 2007 to January 2008, he was working for Aeroflex Cambridge Ltd., U.K., mainly responsible for layer 1 and layer 2 algorithm designs for test mobile systems for 3GPP Long Term Evolution. His research interests include radio resource management for multi-hop cellular networks, cooperative relaying for OFDM/OFDMA systems, packet scheduling for 3G and beyond systems, home base station systems.



Rahim Tafazolli is a Professor of Mobile/Personal communication and Head of Mobile Communications Research Group and is also the Director of Research in the School of EITM, the University of Surrey, UK. He has been active in research for more than 20 years and has published more than 500 research papers in refereed journals, international conferences and as invited speaker. He currently has more than 15 patents in the field of mobile communications. He is the editor of two books on "Technologies for Wireless Future" published by

Wiley's Vol.1 in 2004 and Vol.2 2006. Professor Tafazolli has been consultants to many mobile companies, lectured and chaired a number of IEE Summer schools and IEEE workshops and conferences. He has carried out many consultancies for mobile companies, Home Office and European Union all in the field of mobile communications. He is the Founder and Chairman of IEE International Conference on 3rd Generation Mobile Communications. He is currently the chairman of e.Mobility Technology Platform Expert Group tasked with defining strategic research agenda for EU FP7 programme of research on mobile/wireless communications.