Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2007, Article ID 76193, 11 pages
doi:10.1155/2007/76193

Research Article

A Utility-Based Downlink Radio Resource Allocation for Multiservice Cellular DS-CDMA Networks

Mahdi Shabany,[1] Keivan Navaie,[2,3] and Elvino S. Sousa[1]

[1] *Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G4*
[2] *Department of Electrical and Computer Engineering, Faculty of Engineering, Tarbiat Modares University, P.O. Box 14155-4838, Tehran, Iran*
[3] *The Broadband Communications and Wireless Systems (BCWS) Center, Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada K1S 5B6*

Received 30 May 2006; Revised 1 December 2006; Accepted 8 January 2007

Recommended by Wei Li

A novel framework is proposed to model downlink resource allocation problem in multiservice direct-sequence code division multiple-access (DS-CDMA) cellular networks. This framework is based on a defined utility function, which leads to utilizing the network resources in a more efficient way. This utility function quantifies the degree of utilization of resources. As a matter of fact, using the defined utility function, users' channel fluctuations and their delay constraints along with the load conditions of all BSs are all taken into consideration. Unlike previous works, we solve the problem with the general objective of maximizing the total network utility instead of maximizing the achieved utility of each base station (BS). It is shown that this problem is equivalent to finding the optimum BS assignment throughout the network, which is mapped to a multidimensional multiple-choice knapsack problem (MMKP). Since MMKP is NP-hard, a polynomial-time suboptimal algorithm is then proposed to develop an efficient base-station assignment. Simulation results indicate a significant performance improvement in terms of achieved utility and packet drop ratio.

1. INTRODUCTION

Third generation wireless cellular networks provide a variety of services ranging from multimedia to Internet access. In order to enable these services cellular networks are required to support multiple classes of traffic with diverse quality-of-service (QoS) requirements. Due to the limited availability of radio resources, designing a resource control mechanism to utilize the network resources efficiently is a crucial task for the next generation cellular communication systems. However, designing an optimal resource allocation scheme in CDMA cellular networks is a challenging problem especially when different parameters are involved in the system such as the rate, QoS, and delay requirements of various services.

The optimization can be performed either in the network level or in the cell level. Conventional methods for resource allocation in wireless networks are based on the characterization of traffic flows. In these methods the objective is either to minimize base-station power consumption or to maximize the system capacity [1–4]. There are two major limitations in these approaches: they require the traffic characteristics of each flow, which may be difficult to obtain unless standard assumptions such as Poisson traffic are made. Furthermore, admission and access control must be considered in conjunction with the resource allocation mechanism. Moreover, these classical approaches fail to address the throughput-delay tradeoff efficiently [5].

For the multirate delay-constrained services, as in 3G, the conventional approaches are not effective enough in terms of the optimization of the network resources. Therefore, an alternative approach that avoids the above limitations is required. An efficient approach, which surmounts this challenge, is to assign a utility function to each user based on its QoS requirements and channel status. This utility function represents the benefit that the network can earn by serving that user. In other words, by introducing the utility function, no matter how many various services are involved in the network, each service is specified and integrated in the system modeling via a utility function. This implies that the system treats multiclass services in a unified way. The utility function then can be used as a tool to design an optimal resource allocation scheme. The objective of the allocation scheme is to optimize the total network utility, which is defined as the summation of all the users' utility functions.

There is no clear way to define the utility function for multirate delay-constrained services. It is a complicated task because a comprehensive and yet meaningful utility function requires to take all the various aspects of the network and service types into account. Some of these aspects include the channel status, required data rates, and delay constraints of the services.

In this paper, we define a novel utility function for each user that is a function of its channel status, its required service as well as the load condition of the corresponding serving base station. This new definition of the utility function incorporates the information of both the network side (channel) and the user side (rate and delay) in a unified way for radio resource allocation. We focus our attention on the downlink resources (i.e., power and bandwidth), which is considered to be the bottleneck in multiservice systems [6]. To design such a scheme, we take into account the system variations in the physical layer as well as the traffic load of the base stations.

In other words, we propose a utility-based base station assignment and resource scheduling scheme for the downlink in multiservice cellular DS-CDMA networks. Unlike previous works, we solve the problem with the general objective of maximizing the total network utility (multiple base) instead of maximizing the utility of each base station (BS) individually. The scheme can be considered as a scheduler determining the set of users that should be served within each time slot. For the special case of having only packet traffic the work in this paper is a general case of the work in [7, 8].

2. LITERATURE REVIEW

Radio resource allocation for the downlink in a DS-CDMA cellular network is considered in [9, 10] based on the joint power allocation and base-station assignment. A pricing framework based on the utility concept has been introduced in [11]. Using this concept, the uplink resource allocation for power and spreading gain control for one type of non-real-time service is studied in [12]. Utility-based modeling is also utilized for uplink power control in a single service multicell data network in [13]. In the proposed method in [13], QoS for data users is modeled through a utility function that indicates the value of information per assigned power level (bits per Joule). Using the utility function the problem is solved by modeling it as a noncooperative game where each user tries to maximize its own utility.

For multiservice cellular networks with a mixture of symmetric and asymmetric services, it has been shown that in most cases the downlink performance is more critical than that of the uplink [6]. For the downlink, the power allocation problem for multiservice DS-CDMA wireless networks is studied in [14], where the downlink power control problem for multicell wireless networks is formulated as a noncooperative game, although they do not consider downlink power limitation. In practice, transmission power limitation in DS-CDMA cellular systems is a major concern. Therefore, it is necessary to develop algorithms for the power-constrained case as it is presented in this paper.

The pricing framework is also used in [15] to develop a distributed joint power allocation and base-station assignment with the objective of maximization of the total network utility. However, in the strategy adopted in [15], each base station tries to maximize its total utility without considering the status of others. Therefore, the proposed scheme does not necessarily result in maximum total network utility. Furthermore, other QoS parameters such as delay constraint is not discussed. An opportunistic transmission scheduling with resource-sharing constraints has been proposed in [16], which exploits time-varying channel conditions in a single cell. However, the user's delay constraint is not taken into account in [16]. Moreover, their proposed utility function only depends on the channel status in the time slot that the user is being served.

Downlink resource allocation problem for multicell multiservice DS-CDMA system is also studied in our previous works [17, 18]. Both papers, besides per-user throughput, take into account delay requirements of data services as well. The optimum power allocation scheme in a multiservice environment, which supports both data and real-time services, is then modeled using the multiple-choice multidimensional knapsack problem(MMKP); however, the detailed analysis of the problem as well as corresponding heuristic algorithm for MMKP has not been presented in [17, 18].

In our later work [7], we show that optimal packet scheduling in a packet-oriented cellular CDMA/TDMA network can also be modeled as an MMKP. Exploiting delay tolerance of data traffic, we then introduced the notion of multiaccess-point diversity, which is a potential form of diversity in cellular networks, where a signal can be transmitted to the corresponding mobile user via multiple base stations. In [8] we derived analytical performance gain bound on multiaccess-point diversity.

3. SYSTEM MODEL

We consider a time hierarchy for wireless cellular systems where there are three main types of temporal variations in the system.

(1) *Small-scale variation* that is mainly due to the fast fading effect of wireless channel. Fast fading is a consequence of multipath propagation due to reflections of the signal by physical obstacles. We consider T_f second as the time-scale of small-scale variations, that is, fading is assumed to be constant during each T_f seconds.

(2) *Medium-scale variations* that is because of the shadowing effect. Shadowing is the result of the existence of some obstacles between the transmitter and the receiver, usually modeled by a log-normal distribution. Here T_w indicates the time-scale of the medium-scale variations.

(3) *Large-scale variations* that is due to the mobility of users in the network, which results in variations in the system

Table 1: Notations.

Symbol	Definition
M	Number of base stations in the network coverage area
$\mathbf{B}$	Set of base stations in the network coverage area, which are controlled by the RNC
N	Number of total users in the network coverage area
N_i^R	Number of real-time users assigned to base-station i
N_i^D	Number of non-real-time users assigned to base-station i
τ_j	Maximum tolerable delay for user j
$d_j(n)$	The remaining tolerable delay of user j at time n
α	Orthogonality factor
$g_{i,sj}$	The channel gain from base-station i to the user j of service s
$P_{i,sj}$	The transmitted power from BS i, to user j of service s
RT	Set of real-time users
NRT	Set of non-real-time users
$\mathbf{AS}_j$	Active set of user j
$\mathbf{A}_i$	Set of users assigned to base-station i
Ω_m	A feasible base-station assignment
$\overline{R_j}$	Average required data rate for user j
P_{Ti}	Total available transmit power for BS i
P_{Ri}	Total remaining transmit power for BS i to be allocated to nonreal-time users

connectivity. In this paper, T_p indicates the time scale of such variations.

In each time scale, appropriate mechanisms should be utilized to manage the above variations. In this paper, a multiservice DS-CDMA cellular network is considered. Base-stations and users are *nonuniformly* distributed in the network coverage area. This system supports both real-time and nonreal-time (data) services. Real-time services include voice and multimedia. In this paper, we utilize the method presented in our previous work, [18], in T_p time-scale to adaptively adjust coverage areas of base stations based on their traffic loads. Based on this adjustment, in a smaller time scale, each T_w seconds, the more detailed decisions about assigned base stations and data rate of each individual user are made. The typical values for T_f, T_w, and T_p are 1 millisecond, 10 milliseconds, and 100 milliseconds, respectively. For the easy reference, we present the notations used in the rest of the paper in Table 1.

A nested-loop power control is used. A central radio network controller (RNC) performs outer-loop power control every T_w seconds. T_w is assumed to be less than the maximum tolerable delay of user j, τ_j. RNC also performs base-station pilot power adjustments with a time scale of T_p seconds; the coverage area of base stations are adjusted to tackle the large-scale mobility of users. For nonreal-time users, QoS is defined as a maximum delay constraint and a required average bit rate. Data traffic is packetized into equal size packets and served by the DS-CDMA air interface.

Note that our proposed scheme for joint base-station assignment and time scheduling (JBSATS), which will be described in Sections 4 and 5, is performed every T_w seconds. The scheme can be considered as a scheduler determining the set of users that should be served within each time slot. Adaptive pilot power adjustment schemes for base stations, [18], can be performed every T_p seconds. In other words, every T_p seconds, the pilot powers of BSs and consequently their coverage areas are adjusted. Based on these determined coverage areas, the active set of all users are determined. Using these active sets, within each T_w seconds, the base-station assignment scheme is performed to determine the actual assignment of users to the network.

4. BASE-STATION ASSIGNMENT

The system is time slotted and at any time slot each base station first allocates power to the real-time users.

4.1. Real-time users

We consider a system with hexagonal cells including a central cell and the cells in its first and second tier. The received bit-energy-to-interference-plus-noise-spectral-density ratio of user j served by service s while being in the coverage of base-station i, $\Gamma_{i,sj}$, can be written as

$$\Gamma_{i,sj} = \frac{W}{\overline{R_j}} \frac{g_{i,sj}p_{i,sj}}{\sum_{k=1,k\neq i}^{M} P_{Tk}g_{k,sj} + (1-\alpha)(P_{Ti} - p_{i,sj})g_{i,sj} + \eta} \tag{1}$$

for all i in B, s in RT, and j in N_i, where W is the chip rate, r_s is the data rate of user j, and η is the spectral density of the additive white Gaussian noise. The term in the numerator represents the desired received power at the location of the user j, where $P_{i,sj}$ is the transmitted power of the base-station i, and $g_{i,sj}$ is the gain between the base-station i and user j of the class s, which accounts for the effect of path loss, as well as the large scale fading (shadowing). A fast power control is assumed to be running with a separate mechanism, and the outer loop power control is performed within each T_p seconds.

The first term in denominator represents the total received interference from the other base stations, inter cell interference, while the second term shows the intra cell interference, resulted from the portion of the power of base-station i that is allocated to the other users within the coverage area of the base-station i, $P_{Ti} - P_{i,sj}$. The parameter α is the orthogonality factor that is due to the effect of the multipath fading.

Based on (1), the achieved rate of each user, r_j, depends directly on the amount of allocated power to that user by its base station, $P_{i,sj}$, as well as its received interference. Basically these are the two main factors that enable us to manage the total capacity of the system. Using the above definitions, the problem of optimal power allocation to real-time users is formulated as the following classic downlink power control problem:

$$\min\left\{\sum_{i=1}^{M}\sum_{j=1}^{N_i^R} P_{i,sj}\right\}, \quad s \in \mathbf{RT}, \tag{2}$$

$$\text{s.t.} \quad 0 \le \sum_{j=1}^{N_i^R} P_{i,sj} \le P_{Ti}, \tag{3}$$

$$\Gamma_{i,sj} \ge \gamma_s, \quad \forall i \in B,\ \forall s \in RT,\ \forall j \in N_i^R, \tag{4}$$

where (4) denotes the constraint for the maximum allowable BS transmit power that can be assigned based on an upper layer mechanism (i.e., managed by RNC). Constraint (4) indicates the air interface QoS satisfaction of the real-time users. The allocated power based on the downlink power control is the solution of (3), (e.g., see [19–21]).

4.2. Nonreal-time traffic

After power allocation to the real-time users, the available power for allocation to the nonreal-time data users is *upper bounded* by the remaining power of each base-station, which comes from the hardware limitation. We denote this available power of BS i at time slot n by $P_{Ri}(n)$ as

$$P_{Ri}(n) = P_{Ti}(n) - \sum_{s\in RT}\sum_{j=1}^{N_{s,i}} P_{i,sj}(n). \tag{5}$$

The solution of (3) results in maximum available power. Note that all of the remaining power is not necessarily the remaining resource of the system because of the more interference generated in the system by admitting more and more nonreal-time users. Therefore, to prevent real-time users' call degradation after power allocation to nonreal-time users, someone may allocate powers to the real-time users based on the worst-case interference. Worst-case interference is when all base-stations transmit with their maximum transmit power. In this case, the received E_b/I_0 of the real-time users are higher than the threshold value and after some degradations due to the assignment of the nonreal-time users; they will still get their minimum required E_b/I_0. Therefore, at the end all real-time users will experience an acceptable level of QoS. The bit energy to the interference spectral density ratio for user j of the base-station i served by the service s is

$$\Gamma_{ij} = \frac{W p_{i,sj} g_{i,sj}}{\overline{R}_j (I_{ij} + \eta_j)} \ge \Gamma_s, \tag{6}$$

where Γ_s is the minimum required E_b/I_0 of the service s, W is the chip rate, η_j is the additive white Gaussian noise at the user j's receiver, and I_{ij} is the total received interference at the location of user j served by the base-station i calculated by RNC as follows:

$$I_{ij}(n) = \sum_{k=1,k\ne i}^{M} P_{Tk} g_{kj} + (1-\alpha) P_{Ti} g_{ij}. \tag{7}$$

Based on (6), data rate of each user depends on its allocated power, $p_{i,sj}$, channel gain, $g_{i,sj}$, and received interference, I_{ij}. Hereafter, we simply refer to $g_{i,sj}(n)/I_{ij}(n)$ as the channel status and drop subscript s for the brevity of discussion.

Providing service to a user with poor channel status would require more air interface resources such as transmission power, p_{ij}, or longer transmission time due to a lower data rate. As a result, providing the service to a user with better channel status leads to an efficient system resource utilization. On the other hand, among users with the same channel status, providing service to users with less remaining tolerable delay leads to QoS satisfaction of these users while does not degrade the service level of the others. Therefore, utility-based resource allocation is the technique of choice, where the user's service and channel quality is jointly integrated and considered by a utility function, which is used as a tool to optimize the resource allocation scheme.

4.3. Utility-based resource allocation

Considering the delay tolerance of a nonreal-time data user, the network can wait for a good channel status and then send to that user. This idea has been used in recently proposed methods based on utility-based resource control [13, 15]. In these methods, the total network throughput is maximized subject to a set of QoS and resource constraints. For each user, a utility function is defined as an indicator of user's achieved throughput.

In the case where each user has a finite delay constraint, the user's throughput can only indicate the user's satisfaction if it is served in its predetermined tolerable delay period. Taking a network side insight, for a data user with a given maximum delay tolerance, serving that user can be done during its maximum delay tolerance period. This is an opportunity for the network to postpone serving that user and serve other users with better channel status, which corresponds to the less air interface resource to be allocated, and/or a worse delay condition. In this paper, we define a novel utility function that shows the network's benefit due to the above mentioned opportunity.

For user j being served by the BS i in time-slot n, we propose the utility function as

$$u_{ij}(n) = \begin{cases} \Phi(d_j(n))\Psi(\Gamma_{ij}(n)), & i \in \mathbf{AS}_j, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $d_j(n)$ is the remaining tolerable delay of user j, $\Phi(\cdot)$ is an increasing function of $1/d_j(n)$, and $\Psi(\cdot)$ is the probability of success in packet transmission that is assumed to be an increasing function of $\Gamma_{ij}(n)$, defined in (6). The function $\Phi(d_j(n))$ manages the delay-throughout tradeoff by increasing the priority of the users with a given minimum delay tolerance, while $\Psi(\Gamma_{ij}(n))$ characterizes multiaccess-point and multiuser diversity gains. For instance, from two users with the same channel status, the one with less $d_j(n)$ has the higher priority to be served by the network, while between two users with the same delay constraint, the one with a better channel status is served first. In brief, the utility function defined in (8) is a decreasing function of $d_j(n)$, which has its maximum value at $d_j(n) = 0$.

Total network utility, $Q : \vec{u}$, is defined as a function of the individual utilities of the users that are assigned to the BSs, where $\vec{u} \triangleq (u_{1b_1}, u_{2b_2}, \dots, u_{Nb_N})$ is the utility vector, index b_i shows the assigned BS to the user j, and $Q(\cdot)$ is a casual policy defined based on the network performance perspective.

The mathematical definition of $Q(\cdot)$ is related to the service provider's resource management strategy and generally is as follow:

$$Q(\vec{u}) \triangleq \sum_{j=1}^{N} \sum_{i=1}^{M} u_{ij}(n) x_{ij}(n), \tag{9}$$

where $x_{ij}(n)$ is the assignment indicator in time-slot n, that is, $x_{ij}(n) = 1$ if BS i is assigned to user j and $x_{ij}(n) = 0$, otherwise. If a specific user is not assigned to the network at time-slot n, this means that a BS that is out of its active set is selected for serving. Therefore, by the definition in (8), its corresponding utility would be zero. The total network utility represents the total benefit that network earns by serving the users while their delay requirements are also being satisfied.

In this paper, the total network utility is defined as the sum of all individual user's utility. In other words, the higher network utility shows the more resource control efficiency in terms of providing service to the users with the maximum achievable utility.

5. BASE-STATION ASSIGNMENT ALGORITHM

In this paper, our objective is to maximize the total network utility. Such optimization leads to maximizing the total allocated data rate in the network while considering the channel status, and the delay constraints of all users. In other words, maximizing the total network utility shows that the network waits intelligently for a better accessible channel status for each user while considering its maximum tolerable delay. Based on (8), the utility function of a user depends on its assigned base station. Therefore, for a given set of available powers for nonreal-time users, the problem of maximizing the total utility of the network leads to the problem of finding the optimum base-station assignment, which is implemented by RNC.

In DS-CDMA networks, for each user, the base-station assignment is performed based on the selection of a base-station whose corresponding received E_c/I_0, the bit energy of pilot channel to the total received interference spectral density, is the maximum. In other words, each user has an active set of base stations from which it chooses its best server. This active set is defined as a set of base stations whose corresponding received E_c/I_0 are greater than a performance threshold, that is,

$$\mathbf{AS}_j = \{i \mid i \in B,\ (E_c/I_0)_{ij} \geq \gamma_{\min}\}, \tag{10}$$

where $\gamma_{\min}$ is the minimum required E_c/I_0.

In this case, in selecting the best server for each user, the traffic profile of the network and the target base station is not taken into account while in our scheme it is possible for a specific user, whose best server is overloaded, to be served by another base station in its active set with better load condition. Therefore, the total utility of the network can be improved.

Here, we propose a base-station assignment mechanism, which selects the best server of each user to maximize the total network utility. The input of the algorithm consists of the values of the utility functions of all users, which can be defined in an arbitrary but meaningful way. Therefore, our proposed modeling can be applied in a more general case by *any* definition of utility. Let $\mathbf{P}_R = [P_{R1}, \dots, P_{RM}]$ be the vector of base-stations' remaining powers. Therefore, the optimal base-station assignment in the time-slot n is a solution of the following optimization problem:

$$\max_{x_{ij}} \left(\sum_{i=1}^{M} \sum_{j=1}^{N} u_{ij}(n) x_{ij}(n) \right), \tag{11}$$

$$\text{s.t.} \quad \sum_{j \in A_i} p_{ij}(n) x_{ij}(n) \leq P_{Ri}(n), \quad \forall i \in B, \tag{12}$$

$$\sum_{i=1}^{M} x_{ij}(n) = 1, \quad x_{ij}(n) \in \{0, 1\} \quad \forall j = 1, \dots, N, \tag{13}$$

where $x_{ij}(n)$ is one if the user j is assigned to the base-station i at the time-slot n, and zero, otherwise. For the brevity of discussion in the following we drop the time index n.

Let $MS_i = \{j \mid i \in AS_j\}$ be the set of nonreal-time users that base station i is in their active sets. The total required power to serve a valid subset of MS_i should be smaller than or equal to P_{Ri}. Each user is assumed to be served by only one base-station. Therefore, a *feasible base-station assignment*, Ω_m, is the combination of nonintersect *valid* subsets of MS_i, $i = 1, \dots, M$. A valid subset means a subset whose sum of required powers of its individual users is less than or equal to the total remaining power of its corresponding base-station. Our objective is to find Ω_{m^*} as its corresponding total utility, $U(\Omega_{m^*})$, such that

$$m^* = \underset{m}{\operatorname{argmax}}\, U(\Omega_{m^*}). \tag{14}$$

The base-station assignment scheme is summarized in Algorithm 1.

In the following, we map the downlink resource allocation problem in (12) to a multidimensional multiple-choice knapsack problems (MMKP).

(1) For each $j \in \mathbf{NRT}$, RNC obtains u_{ij} for all $\mathrm{BS}_i \in \mathbf{AS}_j$,
(2) RNC obtains valid subsets for all base stations,
(3) RNC searches different feasible base-station assignments, Ω_m, and the optimal assignment is determined based on (14).

Algorithm 1: Proposed base-station assignment scheme.

Definition 1 (MMKP). An MMKP is the problem where there is an M-dimensional knapsack with M total allowable volumes of $W_1, W_2, \dots, W_M$ and there are N groups of items. Group j has n_j items. Each item has a value and M volumes corresponding to the knapsack's M dimensions. The objective of the MMKP is to pick up exactly one item from each group for the maximum total value of the selected items, subject to the volume constraints of the knapsack's dimensions. In mathematical representation, let v_{kj} be the value of the kth item of the jth group, let $\vec{w}_{kj} = (w_{kj1}, \dots, w_{kjM})$ be the required volume of the kth item of the jth group corresponding to M dimensions, and let $\vec{W} = (W_1, \dots, W_M)$ be the volume constraints of different knapsack's dimensions. Then the problem can be written as

$$\begin{aligned} &\max_{x_{kj}} \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}u_{kj}, \\ &\text{s.t.} \quad \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}w_{ikj} \le W_i \quad \forall i \in \{1,\dots,M\}, \\ &\sum_{k=1}^{n_j} x_{kj} = 1 \quad \forall j \in \{1,\dots,N\},\ x_{kj} \in \{1,0\}. \end{aligned} \tag{15}$$

5.1. Algorithm for optimal base-station assignment

Problem (12) is mapped to a multidimensional multiple-choice knapsack problem (MMKP) as follows. We consider M base stations as a knapsack with M dimensions and each user as a group. Each group has n_j (here M) items equal to the number of base stations. Item k of the user j has a value u_{kj} defined in (8), that is, the utility of user j when it is assigned to the base-station k, and M volumes $\vec{\mathbf{p}}_{kj} = (p_{1jk}, \dots, p_{Mjk})$, which is defined as

$$p_{ijk}(n) = \begin{cases} p_{ij}(n), & k \in AS_j,\ i = k, \\ 0, & \text{otherwise}, \end{cases} \tag{16}$$

which ensures that item k of any group (user), that corresponds to base-station k, can only be assigned to base-station k, which is meaningful.

Therefore, if item k of group j is selected in the optimal solution, it means that the user j has been assigned to the base-station k, its corresponding achieved utility is u_{kj}, and the amount of power it takes from the base-station k is p_{kj}. We have to choose *exactly* one item from each group, meaning that each user can be assigned to *at most* one base station. It is worth mentioning that by the definition of MMKP we have to choose exactly one item from each group. However, the selection of all users is not feasible in many cases. Therefore, if user j does not exist in the optimal solution it means that one of its items whose corresponding value and volumes are zero has been selected. This indirectly implies that user j has not been assigned to the network.

Using above mapping, problem (12) can be rewritten as

$$\max_{x_{ij}} \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}u_{kj}, \tag{17}$$

$$\text{s.t.} \quad \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}p_{ijk} \le P_{Ri} \quad \forall i \in B, \tag{18}$$

$$\sum_{k=1}^{n_j} x_{kj} = 1 \quad \forall j \in \{1,\dots,N\},\ x_{kj} \in \{0,1\}, \tag{19}$$

where x_{kj} is one when the item k of user j is selected.

Since the problem was formulated as an MMKP, any technique available to solve MMKP can be used. Generally, there are two approaches to solve an MMKP; exact and heuristic. The exact solution is based on the branch-and-bound algorithm [22]. The computational complexity of these algorithms is $O(2^{M^2N})$. Therefore, branch-and-bound linear programming approach (BBLP) is often too slow to be useful for radio resource allocation. The alternative is to use a heuristic approach. There are some heuristic algorithms in the literature like the ones in [23, 24]. We use the modified version of [24] to solve our MMKP. Here, we briefly outline some of the known theory on Lagrange multipliers and the algorithm for solving our MMKP to simplify the understanding of our approach.

Theorem 1 (see [25]). *Let $\lambda_1, \dots, \lambda_M$, be M nonnegative Lagrange multipliers, and let $x^*_{kj} \in \{0,1\}$ be the solution of*

$$\max\left\{\left(\sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}u_{kj}\right) - \sum_{i=1}^{M}\lambda_i\sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}p_{ijk}\right\}, \tag{20}$$

then the binary variables x^*_{kj} are also the solution to

$$\max_{x_{ij}} \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}u_{kj}, \tag{21}$$

$$\sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}p_{ijk} \le \sum_{j=1}^{N}\sum_{k=1}^{n_j} x^*_{kj}p_{ijk}. \tag{22}$$

Theorem 1 is the fundamental result that makes Lagrange multipliers applicable to discrete optimization problems such as the MMKP. According to this theorem, the solution to the unconstrained optimization problem (20) is also the solution to the constraint optimization problem (22), which is our MMKP with the constraint values P_{Ri} replaced by $\sum_{j=1}^{N}\sum_{k=1}^{n_j} x^*_{kj}p_{ijk}$. Therefore, if the multipliers λ_i are known, the optimization problem is easily solved, because by a simple manipulation equation (20) can be written as

$$\max\left\{\sum_{j=1}^{N}\sum_{k=1}^{n_j}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right)x_{kj}\right\}, \tag{23}$$

which in turn implies that the solutions are

$$x^*_{kj} = \begin{cases} 1, & \text{if } u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk} > 0, \\ 0, & \text{otherwise}. \end{cases} \tag{24}$$

I. INITIALIZATION PHASE
$\lambda_i \longleftarrow 0 \quad \forall i = 1,\dots,M;$
$p_{ijk} \longleftarrow p_{ijk}/P_{Ti} \quad \forall j = 1,\dots,N;\ \forall k = 1,\dots,n_j;$
$\widehat{K}_j = \operatorname{argmax}_k (u_{kj})$ and $x_{\widehat{K}_j j} \longleftarrow 1 \quad \forall j = 1,\dots,N;$
$T_i \longleftarrow \sum_{j=1}^{N} p_{ij\widehat{K}_j} \quad \forall i = 1,\dots,M;$
II. DROP PHASE
While ($T_i > 1$ for any i) do
$\widehat{I} = \operatorname{argmax}_i \{T_i\}$
For $\{j \mid \widehat{K}_j = \widehat{I}\}$
For $k = 1 : M$
$\Delta_{kj} \longleftarrow (u_{\widehat{I}j} - u_{kj} - \lambda_{\widehat{I}}(p_{\widehat{I}j\widehat{I}} - p_{kjk}))/p_{\widehat{I}j\widehat{I}}$
end
end
$K^*J^* = \operatorname{argmin}_{kj} \{\Delta_{kj}\} \quad \forall j,k$
$\lambda_{\widehat{I}} \longleftarrow \lambda_{\widehat{I}} + \Delta_{K^*J^*}$
$x_{\widehat{K}_{J^*}J^*} \longleftarrow 0$
$x_{K^*J^*} \longleftarrow 1$ (i.e., $\widehat{K}_{J^*} \longleftarrow K^*$)
$T_{\widehat{I}} \longleftarrow T_{\widehat{I}} - p_{\widehat{I}J^*\widehat{I}}$
$T_{K^*} \longleftarrow T_{K^*} + p_{K^*J^*K^*}$
end
III. ADD PHASE
While more items can be exchanged
For $j = 1 : N$
For $k = 1 : M$

$$\mu_{kj} = \begin{cases} u_{kj} - u_{\widehat{K}_j j} & \text{if } (u_{kj} - u_{\widehat{K}_j j} > 0,\ T_k + p_{kjk} \le 1) \\ 0 & \text{otherwise} \end{cases}$$

end
end
$K'J' = \operatorname{argmax}_{kj} \{\mu_{kj}\} \quad \forall j,k$
$T_{\widehat{K}_{J'}} \longleftarrow T_{\widehat{K}_{J'}} - p_{\widehat{K}_{J'}J'\widehat{K}_{J'}}$
$T_{K'} \longleftarrow T_{K'} + p_{K'J'K'}$
$x_{\widehat{K}_{J'}J'} \longleftarrow 0$
$x_{K'J'} \longleftarrow 1$ (i.e., $\widehat{K}_{J'} \longleftarrow K'$)
end

ALGORITHM 2: Heuristic algorithm for base-station assignment.

Since we have another constraint in (19), among the solutions in (24), we have to look for the one which satisfies (19) and is optimal at the same time.

Therefore, the only step to do so is to compute the Lagrange multipliers λ_i. It is worth noting that if these multipliers are computed such that the terms $P_{Ri} - \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}^* p_{ijk}$ are nonnegative, the solution is feasible. The solution is optimal, if the following condition holds:

$$\sum_{i=1}^{M} \lambda_i \left(P_{Ri} - \sum_{j=1}^{N} \sum_{k=1}^{n_j} x_{kj}^* p_{ijk} \right) = 0 \tag{25}$$

(i.e., the case whereby error is zero). The MMKP algorithm is given in Algorithm 2.

5.2. Heuristic algorithm

The algorithm starts with the most valuable item of each user j as the selected item ($\widehat{K}_j$), and the Lagrange multipliers initialized to zero such that the constraints in (19) and (24) are satisfied, Initialization Phase. In general, however, the volume constraints will now be violated. The initial choice of selected items is adapted to obey the volume constraints by repeatedly improving on the most violated constraint, $\widehat{I}$. This step is done in DROP phase.

Consider the users whose selected items correspond to base-station $\widehat{I}$ (i.e., $\{j \mid \widehat{K}_j = \widehat{I}\}$). For each item k of these users, the increase Δ_{kj} of multiplier $\lambda_{\widehat{I}}$, that results from exchanging the selected item of group j, is computed. Eventually, the item K^* of user J^* causing the least increase of multiplier $\lambda_{\widehat{I}}$ is chosen for exchange. This choice minimizes the widening of the gap between the optimal solution characterized by (25) and the solution returned by MMKP algorithm. The process is repeated until for each user an item has been selected such that the volume constraints are satisfied. Since each user has always an item whose value and M-dimension volume are zero corresponding to the base station that is not in its active set, the solution is always feasible.

After completion of *Drop* Phase, there may be some space left in the knapsack. This space may be utilized to improve the solution by replacing some selected items with more valuable ones. Therefore, in the *Add* Phase of the algorithm, each item k of every user j is checked against the selected item of that user ($\widehat{K}_j$). It is tested whether item k is more valuable than the selected item, and if k can replace the selected item without violating the volume constraints. Among all exchangeable items, the item K' of user J' causing the largest increase of the knapsack value is exchanged with the selected item of that user ($\widehat{K}_{J'}$). This process is repeated until no more exchanges are possible. The resulting solution comprised of the selected items is feasible, and even optimal, if (25) is satisfied.

Proposition 1. *The maximum difference between the total achieved throughput using above suboptimal algorithm and globally optimal solution is*

$$\sum_{i=1}^{M} \lambda_i \left(P_{Ri} - \sum_{j=1}^{N} \sum_{k=1}^{n_j} x_{kj}^* p_{ijk} \right), \tag{26}$$

where x_{kj}^* are the outputs of the heuristic algorithm.

Proof. See the appendix. □

5.3. Computational complexity

Step I is just the initialization whose effect on the time complexity of the algorithm is negligible $O(M + 3NM + M^2N)$. Drop phase is the determining factor in the complexity of the algorithm. Basically this step can be repeated at most NM times until no infeasible knapsack ($T_i > 1$) remains. At each iteration, there are $NM^2 + NM + 2M$ additions and/or comparisons, which means that the complexity of this phase is at most $O(MN(NM^2 + NM + 2M))$. Therefore, ignoring the negligible terms, we end up to the total complexity of $O(N^2M^3)$, which is polynomial time. For detailed complexity analysis, see [17].

6. SIMULATION RESULTS

We consider a two-tier hexagonal cell configuration with a wrap-around technique [26]. A universal mobile telecommunication system (UMTS), with a fast power controller running at 1500 updates per second, is simulated. Cross-correlation between the codes in a cell at the mobile receiver is assumed to be equal to 0.3. We simulate a mixture of voice and data users; voice services with 12.2 kbps, activity factor of 0.67 and minimum required $E_b/I_0 = 5$ dB, while data services have minimum required E_b/I_0 of 3 dB. Packet arrival is modeled by a Poisson process.

In this paper, we define

$$\Phi(d_j(n)) = \begin{cases} \exp\left(\dfrac{1}{T_w + d_j(n)}\right), & 0 \leq d_j(n) \leq \tau_j, \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

In fact, any function that is a decreasing function of $d_j(n)$ will result in the same performance result. It is seen that if $d_j(n)$ of a user approaches zero, its corresponding $\Phi(\cdot)$ becomes very high, and overrides channel considerations in (8). Note that when all services have no delay constraint, the problem is simply reduced to the conventional SIR-based base-station assignment.

Channel fading is based on the Gudmundson model with fading standard deviation equal to 6.5 dB. A distance-dependent channel loss with path exponent of −4 is considered. We focus on the central cell and use the delay constraint and channel status of users to determine the utility function for each user relative to the base stations in its active set.

We now compare the gain of our proposed base-station assignment to the conventional SIR-based assignment. Initially, N_{uni} users were distributed uniformly throughout all the cells. After that, N_{nonuni} users were added to the boundary of the central cell. All users have the same delay constraint. The ratio of total achieved utility of our scheme to that of SIR-based scheme versus the number of added nonuniform users in an 8-set cell corresponding to the central cell and seven cells in its first tier is shown in Figure 1.

It is seen that our proposed scheme performs better for small values of N_{uni}, which means more total utility is gained when neighboring cells are lightly loaded or have users with more relaxed delay constraints. Therefore, the rate of increase in total utility is maximum for $N_{\text{uni}} = 2$. This idea is seen more clearly in Figure 2, where the rate of increase in achieved utility for different cases is shown.

It is seen by increasing the number of added nonuniform users in the boundary of the central cell, the performance is better when the number of uniform users is smaller. This is because adjacent cells can serve more users of the central cell when they have a smaller number of users. Moreover, by increasing the number of nonuniform users, N_{nonuni}, the total achieved gain approaches a steady-state value, which is the maximum capacity that can be obtained using our scheme.

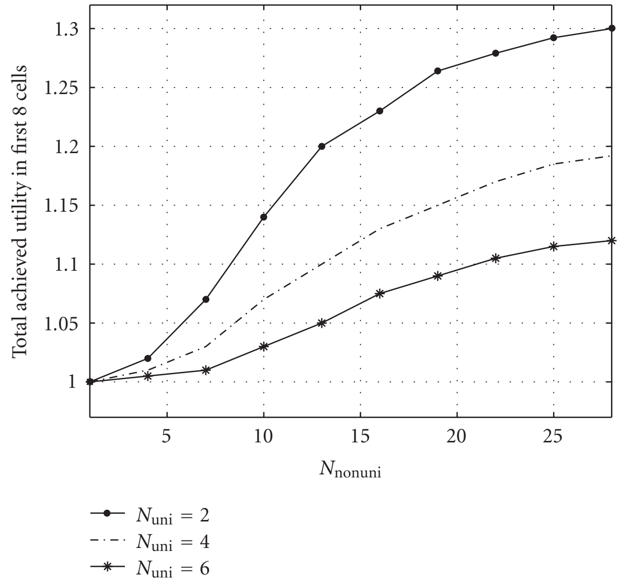

Figure 1: The ratio of total achieved utility of our scheme to that of SIR-based scheme in first eight cell versus different number of added nonuniform users in the central cell.

In another scenario, we distributed 5 users in all cells like before, but limited the number of base stations in the active set of each user. Moreover, we considered the results for the two different patterns of nonuniform users' distributions. In the first case (pattern A), we distributed more users throughout the central cell randomly, while in the second one (pattern B) the users were grouped in subcells located at the cell boundary in the corner of three adjacent cells. The result is shown in Figure 3. It is seen that by increasing the number of allowable BSs in the active set of each user the performance is improved slightly. Moreover, if all nonuniform users are located in the cell boundary for large values of N_{nonuni}, the total achieved utility is improved while for small values of N_{nonuni} the results are almost the same.

We also consider the total network utility as in (12) and compare the system performance for three distinct resource control schemes: SIR-based (SIR-BSA), the individual BS utility maximization (IU-BSA) [15], and the proposed JBATS. Nonuniform user distribution in the network coverage area is expressed by the nonuniformity factor μ_D, which is the ratio of the users that are distributed nonuniformly to the total number of users. The result is shown in Figure 4.

In order to study the run-time performance of the algorithm, we implemented it along with the optimal algorithm based on branch and bound search using linear programming for upper bound computation. Although branch-and-bound is infeasible in practical application for larger data sets, we run this algorithm to determine the optimality of the heuristics by finding an upper bound using the linear programming approach. We have performed experiments on an extensive set of problem sets where we used randomly generated MMKP instances for our tests. For each set of parameters N and M, we run the algorithm ten times and tabulated

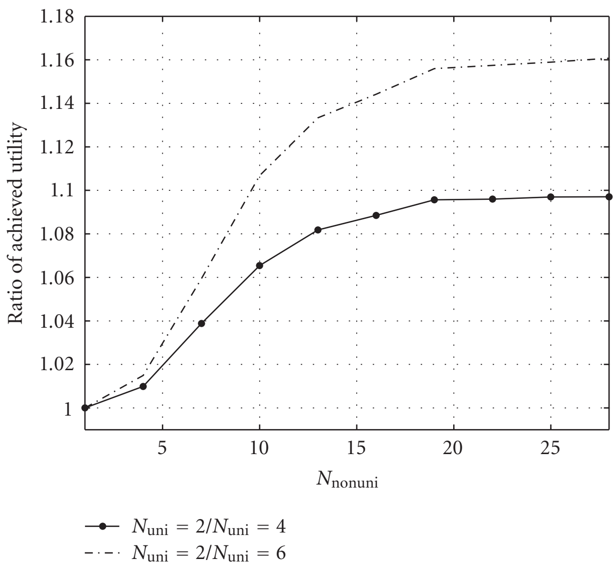

FIGURE 2: The ratio of total achieved utility of the case, where $N_{\text{uni}} = 2$, to the other two cases ($N_{\text{uni}} = 4$ and $N_{\text{uni}} = 6$).

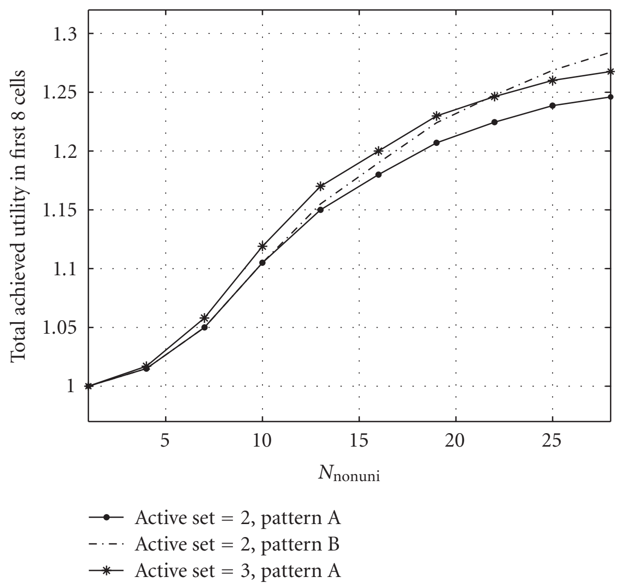

FIGURE 3: The ratio of total achieved utility of our scheme to that of SIR-based scheme in first eight cell versus different pattern of nonuniform users and number of active sets.

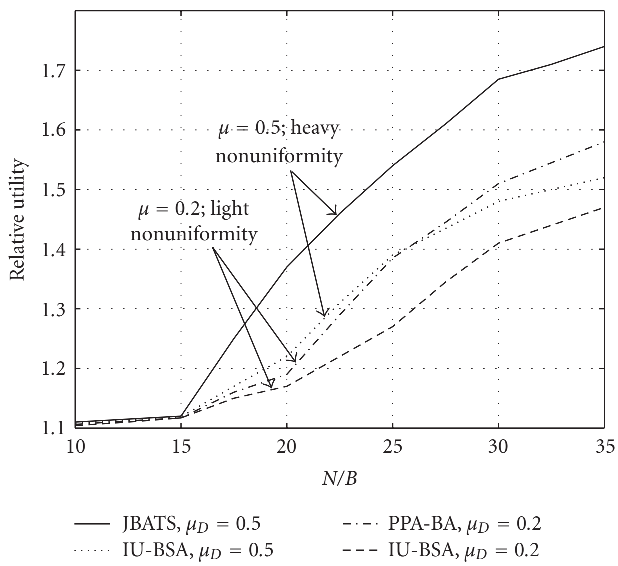

FIGURE 4: The average achieved total network utility for IU-BSA and JBATS normalized by the average achieved total network utility of SIR-BSA versus average number of users per BS (N/B). Two nonuniformity cases: $\mu_D = 0.2$ and $\mu_D = 0.5$.

TABLE 2: Performance comparison of branch-and-bound and a heuristic algorithm in terms of total achieved throughput and execution time.

N	Value %	Time %
40	92.5	15.3
70	95.6	4.2
100	97.3	3.9
130	98.1	2.7
160	97.7	2.7
190	98.1	2.9
220	98.5	3.1
250	98.7	3.1
280	97.5	3.9
310	97.4	3.0
340	98.3	2.4
370	99.3	1.9
400	99.2	2.6

the averages of achieved throughput and execution time. Table 2 shows the percentage of the achieved throughput using our heuristic method compared to the value achieved in the optimal case. Moreover, the third column of the table shows the required execution time in the heuristic method compared to that of branch-and-bound method. It shows that the performance is really good for large sets (greater than 95% most of the time), while the execution time is just a few percent of the time required for optimal solution (less than 5%).

7. CONCLUSION

In this paper, we propose a novel comprehensive scheme, which leads to utilizing the network resources more efficiently. To design such a scheme we take a multi time scale approach. Then in large time scales, we adaptively adjust base-station coverage area based on the corresponding traffic profile of the users in the coverage area. Then in medium time-scales we utilize a utility-based platform to formulate downlink resource allocation based on a novel defined utility function. This utility function quantifies the degree of utilization of network resources. Unlike previous works, we solve the problem with the general objective of maximizing the total network utility instead of achieved utility of each base station. We then map this problem to multidimensional multiple-choice knapsack Problems (MMKP). Since MMKP is NP-hard, a polynomial-time suboptimal algorithm was then modified to develop an efficient base-station assignment. Simulation results indicate significant performance improvement using the proposed scheme.

APPENDIX

Proof of Proposition 1. Assume $X^* = \{x_{kj}^*\}$ is the output of the algorithm, and $Y^* = \{y_{kj}^*\}$ is the result of the globally optimum solution. Lets denote $T_i^* = \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}^* p_{ijk}$. Therefore, the total achieved throughput using the heuristic algorithm can be written as (A.1)-(A.2). For the optimal solution, Y^*, we can rewrite the same expression as in (A.2) as

$$\sum_{j=1}^{N}\sum_{k=1}^{M} x_{kj}^* u_{kj} = \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{n_j} \lambda_i x_{kj}^* p_{ijk} + \sum_{j=1}^{N}\sum_{k=1}^{M} x_{kj}^* u_{kj} - \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{n_j} \lambda_i x_{kj}^* p_{ijk} \tag{A.1}$$

$$= \sum_{k=1}^{M} \lambda_i T_i^* + \sum_{j=1}^{N}\sum_{k=1}^{M}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right) x_{kj}^*, \tag{A.2}$$

$$\sum_{j=1}^{N}\sum_{k=1}^{M} y_{kj}^* u_{kj} = \sum_{k=1}^{M} \lambda_i T_i^{'*} + \sum_{j=1}^{N}\sum_{k=1}^{M}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right) y_{kj}^*, \tag{A.3}$$

where $T_i^{'*} = \sum_{j=1}^{N}\sum_{k=1}^{n_j} y_{kj}^* p_{ijk}$. By definition, we know that all $T_i' \le P_{Ri}$. Therefore, the upper limit for (27) can be written as

$$\sum_{j=1}^{N}\sum_{k=1}^{M} y_{kj}^* u_{kj} \le \sum_{k=1}^{M} \lambda_i P_{Ri} + \sum_{j=1}^{N}\sum_{k=1}^{M}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right) y_{kj}^*. \tag{A.4}$$

Using (A.3) and (A.4), the difference between total achieved throughput using the sub-optimal algorithm and the global optimal solution is

$$\sum_{j=1}^{N}\sum_{k=1}^{M} u_{kj}(y_{kj}^* - x_{kj}^*) \le \sum_{k=1}^{M} \lambda_i (P_{Ri} - T_i^*) + \left\{ \sum_{j=1}^{N}\sum_{k=1}^{M}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right) y_{kj}^* - \sum_{j=1}^{N}\sum_{k=1}^{M}\left(u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk}\right) x_{kj}^* \right\}. \tag{A.5}$$

Let us denote the last term in (A.5) as $W = \sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} y_{kj}^* - \sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} x_{kj}^*$, where $\beta_{kj} = (u_{kj} - \sum_{i=1}^{M}\lambda_i p_{ijk})$. We define the following sets $H_1 = (X^* \cup Y^*) - Y^*$, $H_2 = (X^* \cup Y^*) - X^*$, and $H_3 = (X^* \cap Y^*)$.

For the elements of H_3, it is clear that W is equal to zero. For the elements of H_1, $\sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} y_{kj}^* = 0$ and $\sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} x_{kj}^* \ge 0$, hence $W \le 0$. As for the elements of H_2, $\sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} y_{kj}^* \le 0$ (since $\beta_{kj} \le 0$) and $\sum_{j=1}^{N}\sum_{k=1}^{M}\beta_{kj} x_{kj}^* = 0$, thus, again $W \le 0$. Therefore, in all cases, we have $W \le 0$, which in conjunction with (A.5) meaning that

$$\sum_{j=1}^{N}\sum_{k=1}^{M} u_{kj}(y_{kj}^* - x_{kj}^*) \le \sum_{k=1}^{M} \lambda_i (P_{Ri} - T_i^*) = \sum_{k=1}^{M} \lambda_i \left(P_{Ri} - \sum_{j=1}^{N}\sum_{k=1}^{n_j} x_{kj}^* p_{ijk}\right), \tag{A.6}$$

which completes the proof. □

ACKNOWLEDGMENTS

This paper was partly presented at the IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication, April 2004, and the IEEE Ninth International Symposium on Computers and Communications (ISCC '04), June 2004. This work was supported in part by Bell University Laboratory, University of Toronto.

REFERENCES

[1] M. L. Honig and J. B. Kim, "Resource allocation for packet data transmission in DS-CDMA," in *Proceedings of the 33rd Annual Allerton Conference on Communication, Control, and Computing*, pp. 925–934, Monticello, Ill, USA, October 1995.

[2] M. L. Honig and J. B. Kim, "Allocation of DS-CDMA parameters to achieve multiple rates and qualities of service," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '96)*, vol. 3, pp. 1974–1978, London, UK, November 1996.

[3] J. B. Kim and M. L. Honig, "Resource allocation for multiple classes of DS-CDMA traffic," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 506–519, 2000.

[4] J. B. Kim, M. L. Honig, and S. Jordan, "Dynamic resource allocation for integrated voice and data traffic in DS-CDMA," in *Proceedings of the 54th IEEE Vehicular Technology Conference (VTC '01)*, vol. 1, pp. 42–46, Atlantic City, NJ, USA, October 2001.

[5] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in energy constrained wireless networks," in *Proceedings of IEEE International Symposium on Information Theory (ISIT '04)*, p. 439, Chicago, Ill, USA, June-July 2004.

[6] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, New York, NY, USA, 2000.

[7] K. Navaie and H. Yanikomeroglu, "Optimal downlink resource allocation for non-realtime traffic in cellular CDMA/TDMA networks," *IEEE Communications Letters*, vol. 10, no. 4, pp. 278–280, 2006.

[8] K. Navaie, D. Y. Montuno, H. Yanikomeroglu, and Y. Q. Zhao, "Optimal downlink resource allocation for cellular CDMA networks," in *Adaptation Techniques in Wireless Multimedia Networks*, Y. Xiao and W. Li, Eds., Nova Science, New York, NY, USA, 2006.

[9] R. D. Yates and C.-Y. Huang, "Integrated power control and base station assignment," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 638–644, 1995.

[10] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1332–1340, 1995.

[11] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

[12] S.-J. Oh and K. M. Wasserman, "Optimality of greedy power control and variable spreading gain in multi-class CDMA mobile networks," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom '99)*, pp. 102–112, Seattle, Wash, USA, August 1999.

[13] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1883–1892, 2001.

[14] M. Xiao, N. B. Shroff, and E. K. P. Chong, "Utility-based power control in cellular wireless systems," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 1, pp. 412–421, Anchorage, Alaska, USA, April 2001.

[15] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 1, pp. 1–14, 2006.

[16] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, 2001.

[17] M. Shabany, K. Navaie, and E. S. Sousa, "Downlink resource allocation for data traffic in heterogenous cellular CDMA networks," in *Proceedings of the 9th International Symposium on Computers and Communications (ISCC '04)*, vol. 1, pp. 436–441, Alexandria, Egypt, June-July 2004.

[18] M. Shabany and K. Navaie, "Joint pilot power adjustment and base station assignment for data traffic in cellular CDMA networks," in *Proceedings of IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, pp. 179–183, Princeton, NJ, USA, April 2004.

[19] R. Vannithamby and E. S. Sousa, "An optimum rate/power allocation scheme for downlink in hybrid CDMA/TDMA cellular system," in *Proceedings of the 52th IEEE Vehicular Technology Conference (VTC '00)*, vol. 4, pp. 1734–1738, Boston, Mass, USA, September 2000.

[20] T.-H. Lee, J.-C. Lin, and T. S. Yu, "Downlink power control algorithms for cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 89–94, 1995.

[21] S. A. Grandhi and J. Zander, "Constrained power control in cellular radio systems," in *Proceedings of the 44th IEEE Vehicular Technology Conference (VTC '94)*, vol. 2, pp. 824–828, Stockholm, Sweden, June 1994.

[22] W. Shih, "A branch and bound method for the multiconstraint zero-one knapsack problem," *The Journal of the Operational Research Society*, vol. 30, no. 4, pp. 369–378, 1979.

[23] M. Akbar, E. G. Manning, G. C. Shoja, and S. Khan, "Heuristic solutions for the multiple-choice multi-dimension knapsack problem," in *Proceedings of the International Conference on Computational Science (ICCS '01)*, pp. 659–668, San Francisco, Calif, USA, May 2001.

[24] M. Moser, D. P. Jokanovic, and N. Shiratori, "An algorithm for the multidimensional multiple-choice knapsack problem," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 80, no. 3, pp. 582–589, 1997.

[25] H. Everett III, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.

[26] H. Harada and R. Prasad, *Simulation and Software Radio for Mobile Communications*, Artech House, Norwood, Mass, USA, 2002.