QoS Provisioning and Radio Resource Allocation in OFDMA based WiMAX Systems

Raj Iyengar, Koushik Kar, Biplab Sikdar, Xiang Luo
ECSE Department
Rensselaer Polytechnic Institute, Troy, NY 12180
Email: iyengr@rpi.edu

***Abstract*— Broadband Wireless Access based on the IEEE 802.16 suite of standards is quickly gaining importance as the basis for 4G technology, with the capability of providing high data rate QoS provisioned services for static as well as highly mobile users. Amongst the different PHY layers in the specification, Orthogonal Frequency Division based Multiple Access (OFDMA) is likely to emerge as the preferred PHY, due to the flexibility it offers to the system designer. The standard [6] does not specify the resource allocation technique to be used in the system but leaves this open to allow vendors to differentiate their products. Since scheduling over OFDMA throws open a number of interesting and novel problems, we discuss different aspects of this issue in this article. We also provide computationally efficient approaches for solving some key channel and power assignment questions in an OFDMA based single IEEE 802.11 cell, with the goal of maximizing system throughput.**

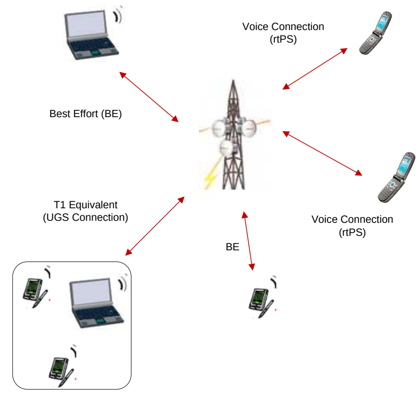

Fig. 1. An 802.16 Base Station serving users at different QoS Levels.

INTRODUCTION

Broadband wireless networks that can provide high speed wireless links to users are increasingly gathering interest in the academic community as well as the industry. An important example are networks based in the IEEE 802.16 suite of standards for point to multi-point broadband wireless access. Initial deployments are likely to be based on fixed point users or at best users with limited mobility or nomadicity, but the 802.16e amendment allows support for highly mobile users.

The standard [6] specifies three different PHYs: Single Carrier (SC), Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Frequency Division based Multiple Access (OFDMA). Amongst these, OFDMA is of most interest in research as well as practice, due to the inherent flexibility it offers to the system designer. This relates directly to the resource allocation (scheduling) question which is of fundamental importance in these systems, since the resource manager controls the efficiency of the system.

Consequently, in this article we focus on resource (OFDM subcarriers in frequency and time slots) allocation problems over an OFDMA PHY. Another major motivating factor behind our study is that the 802.16 draft standard [6] does not specify the exact scheduling technique to be used in the system. While it does specify mechanisms for ways the resource manager can gather information and means to effect the resource allocation decisions made, the actual algorithm is not specified, and allows vendors to differentiate their products. In this article, we focus on a single 802.16 based cell as a first step, although in practice, it is likely that multiple cells will co-exist resulting in more complicated and interesting challenges, resulting in a multi-channel multi-user interference constrained problem for the radio resource manager. The objective of this article is to provide initial solution directions for the scheduling problems likely to be encountered in next generation wireless systems. A typical 802.16 cell operating with users at different levels of QoS as specified in the 802.16 draft standard is shown in Figure 1. A good introduction to the QoS classes in IEEE 802.16 can be found in [5].

The initial portion of an 802.16 frame (the control information) consists of the Downlink Map (DL-Map) and the Uplink Map (UL-Map). These specify information about the allocations made for each client on Uplink/Downlink. Broadly speaking, these maps contain information about which subcarriers and which time slots are allocated to a given user, in a given frame. The Downlink portion of the frame is followed by the Uplink portion. The horizontal axis denotes time and the vertical denotes subcarriers used in OFDMA (hence this axis denotes frequency). The frame structure typically used in IEEE 802.16 based wireless systems is shown in Figure 2, assuming a system with 4 subcarriers for ease of exposition, though realistic systems are expected to have upto 2048 subcarriers.

Figure 2 shows a Time Division Duplex (TDD) frame, for an OFDMA PHY, with allocations made for 3 users on the Uplink Subframe. At time T1, user 1 (U1) is allocated all 4 subcarriers, whereas at time T2, U3 is allocated 3 subcarriers with the remaining subcarrier allocated to U2. In the remainder of the article, we focus on the uplink sucarrier/time allocation problem, although the techniques developed are applicable for downlink

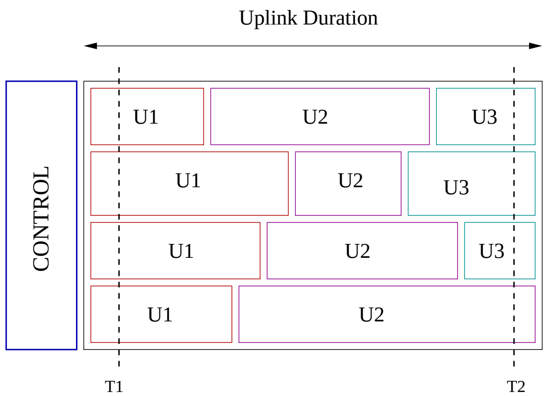

Fig. 2. Typical Frame Structure in 802.16 Scenarios. (The vertical axis represents the different subcarriers used in the system.)

scheduling as well.

We note that the IEEE 802.16 standard specifies a number of techniques for users to request bandwidth resources on the uplink, and a suite of QoS options. We focus on the specific QoS requirement that the user requires an allocation which allows the user to transfer a certain amount of data, over a certain time interval. This is applicable to the UGS (Unsolicited Grant Service) and rtPS (Real Time Polling Service) QoS classes.

The resource allocation problem is to allocate time slots on a subset of the subcarriers available (frequency resource) to meet client demands and maximize system throughput. The time interval T over which these demands must be satisfied can be equal to the frame duration T_F or some other value, and can be interpreted as a time horizon over which the QoS requirements must be met. For example, if the channel conditions are quickly changing, then T can be assumed to be a small value, of the order of a few frame times. On the other hand, if the channel conditions are slowly varying, T can be a larger value.

There is very little literature on resource allocation for wireless networks built around the IEEE 802.16 [6] standard. The draft [6] specifies a number of hooks and features that can be used but does not specify the exact scheduling algorithm, to allow vendors to differentiate their products. Note that in a centralized wireless network, the scheduler is one of the most important components of the system. To the best of the author's knowledge, no published work using the techniques described in this paper exist. While a number of papers on the problem of *bit loading* for OFDMA systems exists, these prior work do not explore in detail the related resource allocation problem.

The Model of an IEEE 802.16 Cell

The time and frequency resources that must be shared between clients are represented in Figure 2. This serves as an abstraction of the OFDMA PHY used. Unlike an OFDM system, the OFDMA system provides the added flexibility of allocating a subset of available subcarriers to a user for some time duration. A Physical Slot (PS) is the basic unit of time at the PHY layer. In the case of OFDMA, this corresponds to a symbol time. A number of PSes together constitute a *slot*. Since resource allocation is typically a MAC layer problem, we focus on the *slot and subcarrier allocation* problem. We note that the techniques developed in this article assume that channel state measurements are made at the individual subcarrier level.

The channel conditions perceived by each station are captured in a channel conditions matrix of dimension $n \times m$ (m subcarriers and n users). The entries in this matrix are a measure of the rate achievable by user i on subcarrier j. For example, these entries may have units of bits/sec which is intuitively useful since the allocations for each user are time durations on subcarriers.

QoS Classes in 802.16

The algorithms presented in this article are applicable to the *Unsolicited Grant Service* (UGS) and *Real Time Polling Service* (rtPS) QoS Classes defined in the 802.16 draft standard [6]. We consider that each user in the UGS class requires the capability to transfer a fixed minimum amount of data over a given time interval T. We note that there is an inherent assumption that it is sufficient to find or be given as input a value T over which the possible satisfaction of demands is acceptable to the applications being used at the user, or as defined in a form of a *Service Level Agreement* (SLA) with the user. We use an identical notion of the rtPS service, with implicit assumptions on the periodicity of the polling intervals for the minimum rate required by an rtPS flow. The Best Effort (BE) and related nrtPS (non-real-time Polling Service) QoS classes are not considered here. The analysis of 802.16 based BE type services are the topic of another article altogether.

Subchannelization Techniques in 802.16

IEEE 802.16 allows for different mappings of subcarriers, which are the elemental frequency level components of an OFDM/OFDMA system, to subchannels, which are logical groupings of subcarriers. An example of a frequency selective mapping is Partially Utilized Subcarrier (PUSC), where the subcarriers that constitute a subchannel are randomly selected from all over the available bandwidth. Owing to the inherent diversity of the constituents of a subchannel, the conditions seen by any user on any given subchannel are roughly the same. Consequently, subcarrier and slot allocation in such scenarios is a much easier problem in this scenario than scenarios with widely varying channel conditions across users. In the case of *Band AMC*, the subcarriers that constitute a subchannel are adjacent to each other, and the channel conditions seen by a user vary across subchannels and further, these vectors vary across users as well.

We note that the techniques developed in this article are applicable to systems that allocate resources at a subchannel level as well. An introduction to subchannelization techniques can be found in [1].

Subcarrier Allocation with QoS Constraints

In this section we present a formulation for the subcarrier allocation problem, where the goal is to maximize the sum throughput across all users so that the demand for each user is satisfied. The motivation for this formulation is that it captures the conflicting objectives of the network (maximizing throughput over the entire network) and the users (satisfying demands of the individual users).

Formulation

The discrete version of the problem (where the time axis is divided into a number of discrete time slots) is in general NP-Hard [3], which motivates the study of the continuous-time relaxation. In the discrete case, the subcarrier allocation problem reduces to the allocation of the subcarriers in each of the slots to individual stations to achieve certain objectives. In the continuous relaxation, the subcarrier allocation problem is to allocate continuous-time chunks to users across the available subcarriers to satisfy demand and maximize throughput.

Let α_{ij} represent the rate achievable by user i on subcarrier j in bits/sec (for m subcarriers and n users). In the current problem formulation, we assume that the achievable rates α_{ij} remain the same irrespective of how many subcarriers are allocated to each user. The goal of the subcarrier allocation problem is to assign subcarriers to users so as to maximize the overall amount of data bits transmitted. This can be posed as

$$\max \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_{ij} x_{ij}, \quad (1)$$

where the variables x_{ij} represent the time duration allotted to station i on subcarrier j to transmit data (for example, refer Figure 2). Additionally, we have the obvious constraint that the total time allocated across all stations on any subcarrier cannot exceed the duration T. This can be written as: $\sum_{i=1}^{n} x_{ij} \leq T \quad \forall\, j = 1 \cdots m$. Furthermore, we have the QoS constraint, that specifies that the total data transmitted by a station i in time T must at least equal the demand d_i in bits, and can be formulated as: $\sum_{j=1}^{m} \alpha_{ij} x_{ij} \geq d_i \quad \forall\, i = 1 \cdots n$. Note that the time duration T represents the time horizon over which QoS guarantees must be provided. Also note that this formulation makes an inherent assumption that channel conditions do not vary significantly over an update interval. In the case of uplink traffic, uplink channel conditions on the different subcarriers can be measured (or estimated based on previous measurements) at the base station every T seconds. For downlink traffic, such channel condition measurements/estimates are obtained at the user, and communicated to the base station. Once a solution to the subcarrier allocation problem posed above is obtained, the exact position of this time chunk allocated to a user is communicated to the user by the base station using control messages that are broadcast to all users, at the start of each frame (these are referred to as Uplink Map and Downlink Map in the standard [6]).

CFS: A Solution Approach for the Subcarrier Allocation Problem

We now develop a heuristic approach that solves the subcarrier allocation problem posed above quite effectively, and in a computationally efficient manner. Note that the subcarrier allocation problem is a linear program (LP), but solving it using generalized LP solving techniques may be computationally too expensive to be of practical interest. We therefore propose a solution approach based on generalized concurrent flow. Using the concurrent flow approach, the time required to satisfy all user demands can be minimized. The rest of the time in the time horizon T can be used to assign subcarriers to users so as to maximize throughput (without any consideration of the user demands), i.e., a subcarrier can be greedily allocated to a user that is currently experiencing the best channel quality for that subcarrier. If the user demands are satisfiable, then our algorithm is able to provide a solution where all user demands are satisfied, although it may not necessarily maximize system throughput (experimental results, some of which are shown below, however show that the attained throughput is very close to the optimum). In case the problem is infeasible, then although all user demands cannot be satisfied, our approach provides a solution which is weighted max-min fair (with the weights being proportional to user demands) across all users. Our solution approach, which we refer to as *Concurrent Flow based Solution (CFS)*, is explained intuitively in Figure 3. More details of this solution technique can be found in [3].

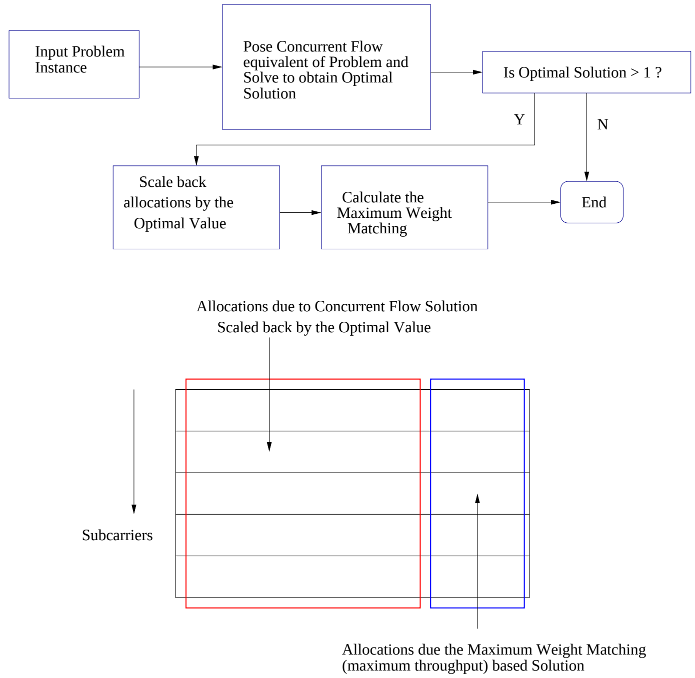

Fig. 3. CFS: Solution Approach Outline and Nature of Solution.

Numerical Results: In this section we present simulation results for the CFS heuristic outlined earlier. We assume an 802.16 system with frame time 5msec. The system operates over 5 subcarriers, and all users have the same demand in these experiments. The achievable rate on each subcarrier for each client is chosen randomly between 1 and 10Mbps. The optimal solution is computed by solving the linear program using CPLEX. From Table I, we note that CFS performs very well for the problem instances considered, and closely approximates the optimal solution.

Joint Subcarrier Allocation and Power Assignment

In general, the achievable rate of a user on a subcarrier is a function of the SINR of that user on that subcarrier, thereby depending on the power allocated on that subcarrier. In uplink transmission, the total power of a user must be split across

Sample	Total bits (CFS)	Total bits (Optimal)	% Difference from Optimal
1	209481	213397	1.8
2	202752	205125	1.1
3	234676	236142	0.6
4	188224	190933	1.4
5	211184	213285	1.0
6	205837	208170	1.1
7	200496	203696	1.5
8	216975	221786	2.1
9	221638	223643	0.9
10	203580	206879	1.6

TABLE I

PERFORMANCE OF CFS; 20 NODES, DEMAND = 9000 BITS/USER.

the different subcarriers used by it, and therefore, the effective achievable rate of a user on a subcarrier typically diminishes as more subcarriers are assigned to it. We now consider this effect in studying the uplink scheduling problem, by posing it as a joint subcarrier allocation and power assignment problem with the goal of maximizing aggregate user throughput.

Formulation

We consider a simple version of the joint subcarrier and power allocation problem in this article, where there are no minimum throughput requirements (demands) for individual users. Even without per-user QoS constraints, the problem considered is difficult to solve optimally, although some interesting solution techniques can be provided in extremal SINR regimes (the SINR is either high or low). The problem considered is as follows: given a measure of the noise power n_{ij} perceived by user i on subcarrier j, across all i, j, we need to find an allocation of subcarriers to users, and the corresponding power assignments on the allocated subcarriers, sum throughput over all users is maximized. Note that while multiple subcarriers can be allocated to a single user, multiple users cannot share a single subcarrier. It is easy to argue that the solution in this case can be represented as a *poly-matching*, an extension of a standard matching, where the users can be matched to multiple subcarriers, but not vice versa. This is represented diagrammatically in Figure 4.

As in our previous discussion, the objective is to maximize sum throughput across all users as stated below:

$$\max \sum_{i=1}^{L} \sum_{j \in \phi_i} \log(1 + \frac{p_{ij}}{n_{ij}}). \quad (2)$$

Here ϕ_i represents the set of subcarriers allocated to user i, where the allocations across all users result in a valid poly-matching. Also, p_{ij} denotes the power allocated by user i on subcarrier j. The constraints to this problem are that user i has a fixed amount of power, P_i, to divide between allocated subcarriers: $\sum_{j \in \phi_i} p_{ij} \leq P_i$.

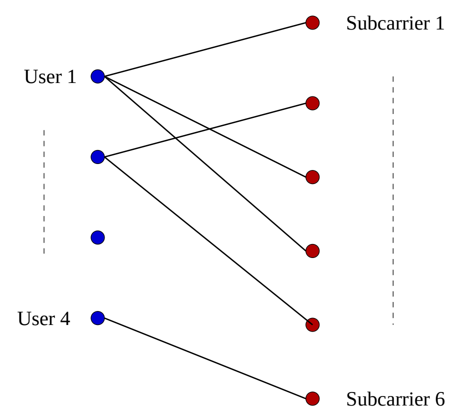

Fig. 4. A *poly-matching:* The figure shows one valid poly-matching for 4 users and 6 subcarriers. (Note that the poly-matching is represented by the edges.)

HSO: Efficient Solution Technique in High SINR Regime

The number of polymatchings is in general exponential in the size of the user-subcarrier bipartite graph, and naive approaches are typically not likely to result in computationally efficient approaches of finding the optimal allocation. However, considering the fact that the SINR values seen by the users are typically very high, it is possible to construct a bipartite graph of users and subcarriers with appropriate edge weights so that the maximum throughput subcarrier allocation is obtained by solving a simpler maximum weighted matching problem, which is well known to be solvable in polynomial time. Experiments in [4] show that this technique closely tracks the optimal solution computed by a brute force search of poly-matchings. The algorithm developed in the High SINR regime is referred to as the *High SINR Optimal (HSO)* algorithm. We note that under HSO, once the subcarrier allocations have been decided, power is divided by a user amongst its subcarriers using standard waterfilling techniques. An example solution (optimal) of the joint subcarrier allocation and power assignment problem is shown in Figure 5.

Numerical Results: Figure 6 shows the simulation results for a system consisting of 5 users and 10 subcarriers. In these simulations, we choose $\sqrt{n_{ij}}$ from Gaussian distribution $N(0, \sigma^2)$; thus we have $E(n_{ij}) = \sigma^2$. For each user i, the maximum

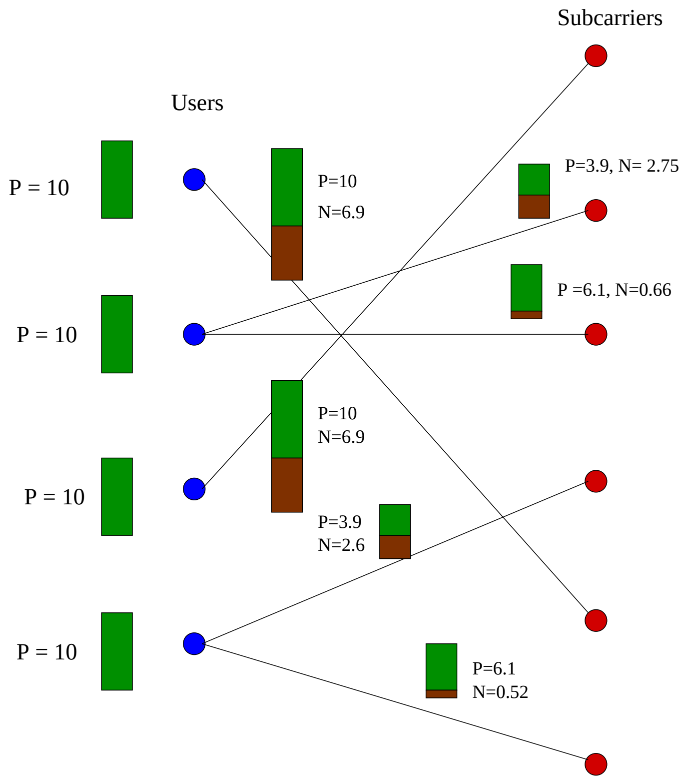

Fig. 5. Example Solution of the Joint Subcarrier Allocation and Power Assignment Problem.

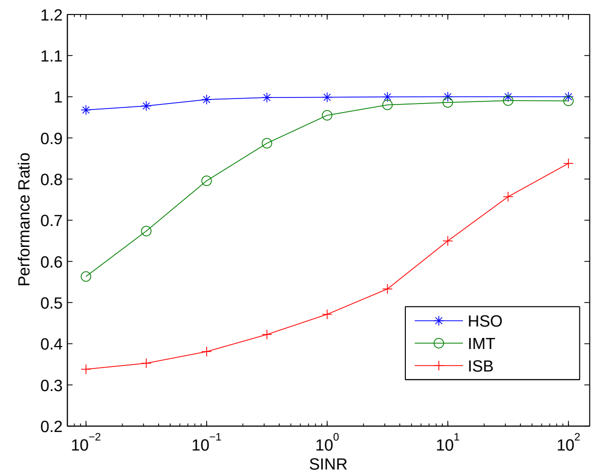

Fig. 6. Simulation Results for 5 Users and 10 Subcarriers.

power P_i is chosen from the uniform distribution $U(0.5, 1.5)$. In the simulations, σ^2 is changed (keeping P_i fixed) to generate a wide range of SINR environments.

The performance numbers shown in the figure across different SINR values correspond to the average performance for that SINR. In the figure, the x-axis corresponds to the SINR, plotted in a semi-log scale. The y-axis corresponds to the ratio of the average throughput attained by an algorithm and the maximum throughput attainable (obtained by complete enumeration over all possible poly-matchings).

From the figure, we see that the HSO algorithm achieves the optimal solution (performance ratio is 1) under high SINR. In fact, the performance ratio of HSO it almost optimal when SINR is close to unity or higher.

This performance is also considerably better than that of two simple greedy subcarrier allocation heuristics (referred to as IMT and ISB in Figure 6). In the *Incremental Max-Thruput (IMT)* heuristic, subcarriers are assigned to users one by one, in any order, with the user chosen such that the assignment yields the maximum additional throughput across all users. In the *Incremental SINR Balancing (ISB)* heuristic, subcarriers are again assigned to users one by one, with the user chosen such that the ratio of the total power and the total noise is balanced across all users, as much as possible. In both cases, user power is divided between the subcarriers using waterfilling.

Summary and Future Work

In this article we discuss some key resource allocation and power control questions in IEEE 802.16 based wireless networks. More specifically, we consider the QoS-constrained scheduling problem and the joint scheduling and power allocation problem for an 802.16 cell, and present low-complexity algorithms that attain the best achievable data rate in the system. The area of frame based, multi-tone, QoS constrained scheduling and resource allocation is relatively new and has direct applications in next generation wireless networks.

An interesting area of future work is to consider per-user demand constraints in the joint subcarrier and power allocation problem. Compared with the simpler version of this problem discussed in this article, the above problem is much more difficult to solve in general, even under the assumption that the SINR values seen by all users are very high. Computationally efficient approaches for solving this QoS-constrained joint subcarrier and power assignment question are currently being investigated. Scheduling and power allocation problems in multi-cell scenarios also remain open for future investigation.

References

[1] H. Yaghoobi, *Scalable OFDMA Physical Layer in IEEE 802.16 WirelessMAN*, Intel Technology Journal, Vol. 8, Issue 3, 2004.
[2] T. H. Cormen, C. E. Leiserson and R. L. Rivest, Introduction to Algorithms, McGraw-Hill, 1990.
[3] R. Iyengar, K. Kar, B. Sikdar, *Scheduling Algorithms for PMP Operation in IEEE 802.16 Networks*, RAWNET 2006 workshop , in Conjunction with WiOPT '06, Boston, Ma.
[4] X. Luo, R. Iyengar, K. Kar, *Throughput Optimal Channel Allocation in Multichannel Wireless Networks*, accepted for publication, WCNC 2007.
[5] C. Cicconetti et al., *Qality of Service Support in IEEE 802.16 Networks*, IEEE Network, Vol. 20, Issue 2, pp. 50-55.
[6] Draft IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems
[7] L. Fleischer, K. Wayne *Fast and Simple Approximation Schemes for Generalized Flow*, in Math. Programming 91 (2002), no. 2, pp. 215-238.
[8] G. Kulkarni, S. Adlakha, M. Srivastava *Subcarrier Allocation and Bit Loading Algorithms for OFDMA based Wireless Networks*, IEEE Transactions on Mobile Computing, Vol. 04, no. 6, pp. 652-662, Nov. 2005.
[9] Christos Papdimitriou, Kenneth Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, 1982.
[10] N. Young *Sequential and Parallel Algorithms for Mixed Covering and Packing*, Proc. FOCS 2001, pp: 538.
[11] M. Garey, D. Johnson *Computers and Intractability* W.H. Freeman, 1979.
[12] H. W. Kuhn, The Hungarian Method for the assignment problem, Naval Research Logistic Quarterly, 2:83-97, 1955.