An Optimization Framework for Radio Resource Management Based on Utility vs. Price Tradeoff in WCDMA Systems

Leonardo Badia*, Cristiano Saturni†, Lorenzo Brunetta†, Michele Zorzi*†
{lbadia,zorzi}@ing.unife.it , {saturni,brunetta}@dei.unipd.it
* Department of Engineering, University of Ferrara, via Saragat 1, 44100 Ferrara, Italy
† Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy

***Abstract*— In this paper we investigate Radio Resource Management strategies for multimedia networks driven by economic aspects such as users' utility and service pricing. To this end, we discuss a framework in which the impact of both QoS and pricing is accounted for in the users' acceptance rate of the service. The model is general enough to be adapted to different situations and optimization goals. Thus, we discuss possible objectives for the network management under the constraints of meeting Quality of Service requirements, by satisfying at the same time technical conditions of feasibility. We employ utility functions, in order to account for the additional characteristic represented by the traffic elasticity, and consider the effect of price to have a realistic characterization of economic quantities. The resulting optimization problem is then discussed and analyzed, to derive general insight and identify possibilities for enhancement.**

***Index Terms*— Optimization, Radio Resource Management, Pricing, Utility Functions, Mathematical programming, Economics.**

I. Introduction

The diffusion of Wireless Communication Systems is nowadays widespread, especially Wideband Code Division Multiple Access (WCDMA) systems, which are able to provide users with different services and data rates. At the same time, this means huge investments by the network operators and the need for an efficient management of the radio spectrum, which is a scarce and expensive resource. There are mainly two points-of-view from which the efficiency can be regarded: users' satisfaction and overall goodness of the network management. The first point is made challenging by the fact that users of Next Generation cellular systems are expected to be highly differentiated in terms of access technology and service requirements. Utility functions can be used in this kind of problem as general means to describe Quality of Service (QoS) [1–3].

However, a pure description of the problem in terms of user-centric utilities might be inefficient, since approaches based on game-theory [4] demonstrate the need for a centralized supervision by the network operator. This explains why it is necessary to introduce coordination among the users, and a possibility to do so is by maximizing a given objective. For example, the global management can be aimed at maximizing the provider revenue, in order to have a sustainable business model, or the number of admitted users, or the total utility coming from the assignment. These objectives have been proven to be contrasting [2], thus the design choice in this sense is very important.

To have a realistic description of the above metrics, also the price paid for the service must be taken into account. The relationship between economics and RRM is twofold, since the pricing and allocation strategies of the provider determine the behavior of the users. Users without adequate QoS are likely dissatisfied; however, the users' feelings also depend on the price paid for the service. This means that users offered with exaggerated QoS at a very high price may want to refuse the service as well. Thus, both single users' and provider's satisfaction levels depend on both allocation and pricing policies.

For this reason, in the present paper we adopt a micro-economic model for the Radio Resource Management (RRM) which explicitly takes into account the trade-off between utility and price, called the MEDUSA model [5]. Its key aspect is that it allows the evaluation of several metrics, which can be included in the management objective. Within this framework, we incorporate economic considerations, by integrating pricing and utility functions. Note also that some recent contributions [6, 7] deal with pricing, since it can be proven in many different ways that charging users for network usage can improve the efficiency of network management. The contribution we bring here is different, since we are interested in economic quantities as possible goals of the optimization, not only as heuristic ways to improve the RRM.

With the presented model it is possible to analyze from a theoretical point-of-view the RRM issue seen as an optimization problem. Our contributions in the present paper are the following: first of all, we propose an optimization framework which is able to account for different goals of the RRM. Secondly, we discuss accurate linearizations of the problem and investigate solutions based on both exact and heuristic algorithms, with considerations about optimality and complexity. Finally, we argue possible extensions of the work, which include possibilities of improving the management at a more global level with the introduction of further RRM schemes (e.g., utility-aware Admission Control) where the optimization problem is considered in different ways.

The rest of the paper is organized as follows: in Section II we discuss related models of utility-based RRM and optimization framework. Section III presents the micro-economic model for integrating utility and pricing in the users' behavior and discusses the network constraints under which the allocation is performed. Section IV synthesizes these elements into the formulation of an optimization problem, which is studied with several techniques. In particular, we present two methods, based on Inter Linear Programming (ILP) and Constrain Programming (CP), which lead to exact solutions. Due to the intrinsic high complexity of exact approaches, a heuristic method based on genetic algorithm is also introduced. Section V evaluates all the methods in a given scenario, by also bringing a discussion about comparative results. Section VI concludes the paper.

II. Related Works

Many papers, appeared in the literature so far, have dealt with the problem of exploiting economic theoretical instruments, like utility functions and virtual pricing, to model RRM and derive analytical studies. In [1] a utility-based framework is introduced to study the Power Control problem, seen as the maximization of a network metric. However, utilities are used in a different way, since they do not model the offered QoS but take into account soft capacity constraints, which we address instead explicitly. Also a cost function is considered, related to the power consumption and hence not generating revenue.

In [8] a utility-based RRM is also considered, where the utilities are instead related to power consumption. In this paper the pricing issue is also considered, but again differently from here, since it is a technical instrument introduced at a virtual level to improve cooperation of the users. This virtual pricing, used mainly as a way to prevent congestion from arising in the network, is also considered in [6,7,9]. Rather, what we call *pricing* here is related to the money exchange between the users and the provider, which implies for example the need for a definition made a priori, since the tariff must be known in advance by the users. Also, this meaning of pricing implies revenue generation as an outcome for the provider.

Other contributions where the RRM framework is extended beyond the power control are [3] and [10], where the structure of the cellular networks is exploited, in the former to separate optimality conditions of Power and Rate control, and in the latter to derive a hierarchical Power/Rate control. In the present paper, instead, a simplified Power Control is assumed and we focus mainly on Rate Control, which is the subject also of [2, 11].

For what concerns the optimization framework applied to the WCDMA capacity constraint, some relationship with the present work are present in [12], even though the scope is different, since here we are more interested in applying directly optimization techniques. Instead, [4] and [13] apply optimization techniques to find theoretical conditions by means of dual problems, but our approach is more practical since our investigation is centered on the WCDMA capacity constraint.

III. Radio Resource Management with Utility and Price

The utility-based approach to RRM assumes for each user the availability of a non-decreasing, quasi-concave function to describe the QoS coming from the assignment of different amounts of resource. This mathematical formulation allows a number of analytical results to be derived. In this paper we are interested in considering an extended version of this framework, in which the competition among users for the resource is made more interesting since resource does not come for free. Thus, Subsection III-A will present the framework used to represent users' behavior. Additionally, we discuss how to model the integration of the microeconomic goals with a technical constraint given by the network capacity. In this paper we focus on a single-cell WCDMA network, whose analysis is given in Subsection III-B.

A. The MEDUSA Model

We refer to the model proposed in [5], which accounts jointly for two contrasting aspects: utility functions to represent the QoS perceived by the users, and a *pricing function* which determines the price paid by the users. It is assumed that both of them are non-negative non-decreasing functions of the assigned resource, and they also concur together in determining users' choices. To identify and cut this trade-off, an *Acceptance value* is defined, which depends on the utility u_i and on the paid price p_i assigned to each user i, thus we will call this function $A_i(u_i, p_i)$. We assume that both u_i and p_i are functions of the allocated resource r_i; thus, A_i depends also on r_i. However, looking at A_i as $A_i(u_i, p_i)$ allows to better emphasize the trade-off, since it is easier to represent the properties of dependence on utility and price that a suitable acceptance function must have. In fact, it is intuitive that A_i must be an increasing function of the utility and a decreasing function of the price.

This model can be applied to the RRM for networks where the terminal have multi-rate capabilities by identifying r_i with the transmission rate assigned to terminal i. The assumption of traffic elasticity [3] corresponds to considering r_i as continuously tunable.

To model the RRM, we assume that all users adopt similar criteria to evaluate their service appreciation, hence we use the same function $A_i(\cdot) = A(\cdot)$ for every user, without the index i. For fairness reasons, it is also sensible that the users know the tariff plan a priori, that is, also the pricing $p_i(\cdot) = p(\cdot)$ is the same for all users. Hence, the subscript i will be omitted for the functions $A(\cdot)$ and $p(\cdot)$, while it will be kept when speaking of the actual values A_i (or p_i) achieved (or paid) by user i, which may be different for different users.

The utility function is instead assumed to be different for every user to account for the variability of services and terminals. Being a subjective factor, the users' evaluation of the service can not be controlled by the resource manager. Instead, it is mainly impacted by the kind of service enjoyed and the terminal used. Hence, we assume a different $u_i(\cdot)$ for every user. However, the utilities are not completely arbitrary, first of all because they are usually assumed to behave according to certain economic properties, like the law of diminishing marginal utilities, which states that the first derivative of the utility, which is non-negative, tends to zero for large r. From the perspective of a RRM utility, this might mean, e.g., that the quality experienced from radio services can not be indefinitely high, i.e., it saturates, at least when a maximum amount of achievable resource $\bar{R}$ is reached.

Moreover, in the following we will model the utilities as parametric functions, so that we are able to tune the utilities by simply changing their internal parameters. This can also be regulated so that there is less or more variability of subjective parameters among the users.

m_i	resulting Economic Metric	symbol
p_i	total revenue	R
1	# of admitted users	S
r_i	total assigned rate	T
u_i	total utility	U

TABLE I

ECONOMIC METRICS OF INTEREST FOR THE RRM

Importantly, note that these specifications about utilities and pricing do not affect the model that will be developed in the following, since they simply impact on the goal function of the optimization problem. Thus, different models for the utilities or the pricing function, or even for the satisfaction $A(\cdot)$ can be adopted without affecting the procedure.

To quantify the network performance, we focus on particular values, called in the sequel *Economic Metrics*, which are meaningful only if evaluated on the users accepting the service conditions. An immediate example is the number of satisfied users itself. These metrics can be evaluated by giving a statistical meaning to A_i. The expected value of an Economic Metric M is in fact:

$$M = \sum_{i=0}^{N-1} m_i A(u_i, p_i)\,, \tag{1}$$

where m_i is the contribution of user i to the Economic Metric and N is the number of users.

Table I reports some examples of Economic Metrics. The total utility is sometimes referred to as *network welfare*, whereas the total assigned rate can be considered as a measure of the resulting throughput. These quantities can be used as goal functions for an optimization problem. In this way it is possible to adopt a wide approach, which considers different Economic Metrics, or even a combination of them. This means that in general it is possible to support different reasonable choices of M.

Note that this discussion specifies the first part of the optimization problem, i.e., the goal function. What is still open is the constraint set, which is the part most specific to the telecommunication networks properties. The constraints we have to consider are related to the network capacity. In other words, if the goal function is, for example, the revenue R, the optimization is achieved by obtaining the highest possible revenue provided that the vector of the allocated r_i's is *feasible* with the capacity of the network, which can be different according to the kind of system and also to the direction of the link (uplink/downlink).

In the next subsection, we will focus on the downlink of a single-cell WCDMA system (note, however, that the same rationale also applies to the uplink with few changes). In fact, it is common to assume the forward link as bottleneck for the connection in multimedia networks, hence we will assume that the capacity is limited in this way.

B. WCDMA Capacity Constraint

We focus on the *soft-capacity* constraint, which assumes the system to be interference limited, as in WCDMA networks. For the sake of simplicity, we will refer to the single-cell case only, even though the reasoning developed in the following can be adapted easily to a multi-cell case with simple modifications. In particular, a very common way of approximating the multi-cell case with a single-cell approach is to consider the inter-cell interference as fixed, so that it is merged into the thermal noise[1].

Modelling the WCDMA soft capacity requires to introduce auxiliary variables representing the allocated power (let us call w_i the power allocated to the ith user). It is well known [14] that in Code Division systems the constraints based on the probability of finding the resource busy are very loose. On the other hand, users can not be admitted when the already connected calls would be subject to an excessive degradation in case of a new admission.

A possible model of this constraint, which follows an approach commonly used in the literature [12], considers the Signal-to-Interference Ratio (SIR). For the downlink of a WCDMA system, the SIR is defined as:

$$SIR_i = \frac{g_i w_i}{\sum_{j=0, j\neq i}^{N-1} \xi_{ij} g_i w_j + \eta_i} \tag{2}$$

where g_k is the power link gain for the kth user, $0 \leq \xi_{ij} \leq 1$ is the normalized cross-correlation between i and j at receiver i and the term η_i is due to background noise. In the following, we will assume this term as equal to a constant η for all users.

The bit-energy-to-interference ratio is equal to

$$\left(\frac{E_b}{I_0}\right)_i = \frac{\mathcal{B}}{r_i} SIR_i \tag{3}$$

where $\mathcal{B}$ is the spreading bandwidth. These relationships offer a matching between data rates and SIR. In fact, if a constant BER requirement is fixed for all mobiles, i.e., $(\frac{E_b}{I_0})$ is equal to a constant $\mathcal{Z}$ for all users and rates, a target for the SIR, called γ, can be defined for every value of rate r, i.e.:

$$\gamma = r\mathcal{Z}/\mathcal{B}.$$

The above equations can be used to map a rate assignment equal to r_i to a SIR-target assignment equal to γ_i via a linear relationship.

With the above notation, it is possible to use the limitation on the total power, which is bound to be below a given value $\bar{W}$, as the feasibility constraint. This will be used in the next section, which is devoted to the formulation of the optimization problem.

IV. THE OPTIMIZATION FRAMEWORK

The formulation of the *WCDMA optimization problem* follows these steps. Let us focus on the revenue maximization problem (for other Economic Metrics, the procedure can be repeated with minor changes).

[1] We will implicitly adopt this approach when, in the scenario, we will adopt a large thermal noise, which can be seen as including also interference coming from surrounding cells.

At first, we have the following nonlinear optimization problem with continuous variables:

$$
\begin{aligned}
\max \quad & \sum_{i=0}^{N-1} p(r_i)A(u_i(r_i), p(r_i)) \\
\text{s.t.} \quad & \\
& r_i = \frac{\frac{\mathcal{B}g_i}{\gamma_i} w_i}{g_i \sum_{j=0\dots N-1, j\neq i} w_j + \eta} \qquad \forall i = 0\dots N-1 \\
& \sum_{i=0}^{N-1} w_i \leq \bar{W} \\
& 0 \leq r_i \leq \bar{R} \qquad \forall i = 0\dots N-1 \\
& w_i \geq 0 \qquad \forall i = 0\dots N-1
\end{aligned} \tag{4}
$$

where $u_i(r_i)$'s are quasi-concave non-decreasing non-negative functions, $p(r_i)$ is a non-decreasing non-negative function, $A_i(u_i, p_i)$'s are compound functions with values in $[0, 1]$ which are decreasing in p_i and increasing in u_i. Moreover, the values of $\gamma_i, g_i, \mathcal{B}, \eta, \bar{R}, \bar{W}$ are all positive constants.

In the following we present two exact methods to solve this problem. In particular, in Subsection IV-A we will present an approach based on Integer Linear Programming (ILP), whereas in Subsection IV-B we solve the problem by means of Constraint Programming (CP). To deal with the high complexity of these approaches, in Subsection IV-C we outline a procedure to derive also a Genetic Algorithm (GA).

A. The ILP approach

The variables w_i are allowed to take value only on a finite subset of $[0, \bar{W}]$. To deal with the difficulties that the functions

$$\theta_i(r) = p(r)A(u_i(r), p(r)) \qquad \forall i = 0, \dots, N-1$$

brings, we can think of approximating them with piece-wise constant functions Θ_i. We can think of Θ_i's as a "sample & hold" version of the θ_i.

Thus, if Θ_i is defined on a finite set of pairwise disjoint closed intervals of $[0, \bar{R}]$, and is constant in every interval, we can approximate this problem by a boolean integer program with linear objective and linear and quadratic constraints, which in turn can be transformed into an equivalent 0-1 Integer Linear Program (ILP).

This ILP can then be solved by a standard Branch&Cut algorithm [15]. Branch&Cut implements Branch&Bound by using linear programming to derive valid bounds during construction of the search tree. It is employed for solving mixed integer linear programs. The relaxation of the integrality constraints (LP-relaxation), which is usually done in the Branch&Bound algorithms implemented by commercial ILP solvers, often provides lower bounds that are very poor.

In Branch&Cut the LP-relaxation is instead successively tightened by the addition to the model of "valid" inequalities, where a valid inequality is in general a plane that cuts off a portion of the region of the feasible solutions, so it is also referred to as cutting plane.

Inequalities that can be generated are in general exponentially many, but usually only a subset of them is sufficient to significantly strengthen the lower bound. These subset of constraints are generated by a so-called cutting plane procedure. This makes Branch&Cut a much more powerful method than Branch&Bound alone. In fact, since cuts are generated dynamically throughout the search tree, the solution can be found for large-scale instances. This can be done by using a generic ILP solver, such as CPLEX [16], which has been employed to test the accuracy of the solution.

In our heuristic approach, by improving the approximation of the θ_i and by increasing the number of w_i samples, we can improve the approximation of the original nonlinear problem as desired. In the following we show how the approximating integer linear model is obtained.

If we require the variables w_i to take value in $\{W_0, .., W_{m-1}\}^N$, being $0 \leq W_0 \leq W_1 < \cdots < W_{m-1} \leq \bar{W}$, and the variables r_i to take value only in $[R_0^-, R_0^+] \cup \cdots \cup [R_{k-1}^-, R_{k-1}^+]$, with $0 \leq R_0^- < R_0^+ \leq \cdots \leq R_{k-1}^- < R_{k-1}^+ \leq \bar{R}$, and $\forall i = 0, \dots, N-1$ and $\forall j = 0, \dots, k-1$ we approximate θ_i in $[R_j^-, R_j^+]$ with the constant $\Theta_{ij} = \theta_i(R_j^+)$, then we can build the following boolean quadratic program:

$$
\begin{aligned}
\max \quad & \sum_{i=0}^{N-1} \sum_{j=0}^{k-1} y_{ij}\Theta_{ij} \\
\text{s.t.:} \quad & \\
& \sum_{j=0}^{k-1} y_{ij} = 1, \quad \sum_{h=0}^{m-1} x_{ih} = 1 \qquad \forall i = 0\dots N-1 \\
& \sum_{i=0}^{N-1} \sum_{h=0}^{m-1} x_{ih} W_h \leq \bar{W} \\
& y_{ij} \in \{0, 1\} \qquad \begin{cases} \forall i = 0\dots N-1, \\ \forall j = 0\dots k-1 \end{cases} \\
& x_{ih} \in \{0, 1\} \qquad \begin{cases} \forall i = 0\dots N-1, \\ \forall h = 0\dots m-1 \end{cases} \\
& \frac{\frac{\mathcal{B}g_i}{\gamma_i} \sum_{h=0}^{m-1} W_h x_{ih}}{g_i \sum_{l=0\dots N-1, l\neq i} \sum_{j=0}^{m-1} W_j x_{lj} + \eta} \geq \sum_{h=0}^{k-1} R_h^- y_{ih} \\
& \qquad \forall i = 0\dots N-1 \\
& \frac{\frac{\mathcal{B}g_i}{\gamma_i} \sum_{h=0}^{m-1} W_h x_{ih}}{g_i \sum_{l=0\dots N-1, l\neq i} \sum_{j=0}^{m-1} W_j x_{lj} + \eta} \leq \sum_{h=0}^{k-1} R_h^+ y_{ih} \\
& \qquad \forall i = 0\dots N-1
\end{aligned} \tag{5}
$$

where the boolean variables have the following meaning:

$$x_{ij} = 1 \Leftrightarrow w_i = W_j$$

$$y_{ih} = 1 \Leftrightarrow r_i \in [R_h^-, R_h^+].$$

Note that, if (x, y) is a solution of problem (5), then $w_i = \sum_{h=0\dots m-1} W_h x_{ih}$ for $i = 0\dots N-1$, the SIR for each user can be computed from Equation (2) and by using it in Equation (3) the r_i's can be derived, being certain that $r_i \in$

$[\sum_{h=0}^{k-1} R_h^- y_{ih}, \sum_{h=0}^{k-1} R_h^+ y_{ih}]$, for $i = 0 \dots N-1$. Therefore, (5) approximates the original problem (4).

By substituting the product $x_{\ell j} y_{ih}$ with the variables $z_{\ell jih}$, $\forall \ell, j, i, h$, and adding the corresponding constraints:

$$2z_{\ell jih} \leq x_{\ell j} + y_{ih}$$

$$z_{\ell jih} \geq x_{\ell j} + y_{ih} - 1$$

that force $z_{\ell jih}$ to be equal to $x_{\ell j} y_{ih}$, we obtain the following boolean ILP:

$$
\begin{aligned}
&\max \quad \sum_{i=0}^{N-1} \sum_{j=0}^{k-1} y_{ij} \Theta_{ij} \\
&\text{s.t.:} \\
&\sum_{j=0}^{k-1} y_{ij} = 1, \quad \sum_{h=0}^{m-1} x_{ih} = 1 \qquad \forall i = 0 \dots N-1 \qquad (6) \\
&\sum_{i=0}^{N-1} \sum_{h=0}^{m-1} x_{ih} W_h \leq \bar{W} \\
&y_{ij} \in \{0,1\} \qquad \begin{cases} \forall i = 0 \dots N-1, \\ \forall j = 0 \dots k-1 \end{cases} \\
&x_{ih} \in \{0,1\} \qquad \begin{cases} \forall i = 0 \dots N-1, \\ \forall h = 0 \dots m-1 \end{cases} \\
&\left.\begin{array}{l} z_{ljih} \in \{0,1\} \\ 2z_{ljih} \leq x_{lj} + y_{ih} \\ z_{ljih} \geq x_{lj} + y_{ih} - 1 \end{array}\right\} \qquad \begin{cases} \forall i = 0 \dots N-1, \\ \forall h = 0 \dots k-1, \\ \forall l = 0 \dots N-1, \\ \forall j = 0 \dots m-1 \end{cases} \\
&\sum_{h=0}^{m-1} \frac{\mathcal{B} g_i}{\gamma_i} W_h x_{ih} \leq \sum_{h=0}^{k-1} \eta R_h^+ y_{ih} + \\
&+ \sum_{h=0}^{k-1} \sum_{l=0}^{N-1} \sum_{\substack{j=0 \\ l \neq i}}^{m-1} g_i W_j R_h^+ z_{ljih} \qquad \forall i = 0 \dots N-1 \\
&\sum_{h=0}^{k-1} \sum_{l=0}^{N-1} \sum_{\substack{j=0 \\ l \neq i}}^{m-1} g_i W_j R_h^- z_{ljih} + \\
&+ \sum_{h=0}^{k-1} R_h^- \eta y_{ih} \leq \sum_{h=0}^{m-1} \frac{\mathcal{B} g_i}{\gamma_i} W_h x_{ih} \qquad \forall i = 0 \dots N-1
\end{aligned}
$$

which is equivalent to problem (5), in the sense that the set of the (x, y) components of solutions (x, y, z) of (IV-A) are the solutions of (5).

By increasing m and k, we can improve the capability of the linear discrete problem (IV-A) to approximate the original continuous nonlinear problem (4). In our experiments we have chosen equidistant W_i's and $R_i^+ = \frac{\bar{R}}{k}(i+1)$ and $R_i^- = \frac{\bar{R}}{k} i + \delta$ for $i = 0 \dots k-1$, with δ being a small positive number $< \frac{\bar{R}}{k}$. Note that the set of intervals allowed for variable r_i need not be the same for every i, as we did, and we can choose them according to the function θ_i.

B. The CP approach

In this section we will present an exact algorithm to solve the original non linear problem (4), assuming that the vector $\mathbf{w} = (w_0, \dots, w_{N-1}) \in \{W_0 \dots W_{m-1}\}^N$ and the functions θ_i are approximated by the piece-wise constant functions Θ_i, as before. Hence we can compare the solution given by the approach described in this section with that given by the ILP model.

With this assumption, problem (4) is:

$$
\begin{aligned}
&\max f(\mathbf{w}) \\
&\textstyle\sum_{i=0}^{N-1} w_i \leq \bar{W} \\
&r_i(\mathbf{w}) \leq \bar{R} \qquad \forall i = 0 \dots N-1 \\
&\mathbf{w} \in \{W_0 \dots W_{m-1}\}^N
\end{aligned}
$$

where:

$$r_i(\mathbf{w}) = \frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\sum_{j=0 \dots N-1, j \neq i} w_j)}$$

and

$$f(\mathbf{w}) = \sum_{i=0}^{N-1} p(r_i(\mathbf{w})) A(u_i(r_i(\mathbf{w})), p(r_i(\mathbf{w})))$$

Our algorithm is based on a standard backtracking technique (akin to those used in Constraint Programming) to generate all the solutions and to find one with the highest value. To limit the number of solutions actually evaluated, we use two pruning mechanisms based on feasibility check.

One mechanism detects infeasibility of partial variable assignments. First it exploits the fact that each variable has some values that cannot be used. In fact, if $\mathbf{w}$ is a feasible point then:

$$r_i(\mathbf{w}) = \frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\sum_{j=0 \dots N-1, j \neq i} w_j)} =$$

$$= \frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\sum_{j=0}^{N-1} w_j - w_i)} \geq \frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\bar{W} - w_i)}$$

So, for the feasibility of w, it will never happen that

$$\frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\bar{W} - w_i)} > \bar{R},$$

which implies that:

$$\frac{\frac{\mathcal{B} g_i}{\gamma_i} w_i}{\eta + g_i(\bar{W} - w_i)} \leq \bar{R},$$

or equivalently:

$$w_i \leq \frac{\bar{R}\eta + \bar{R} g_i \bar{W}}{g_i(\frac{T}{\gamma_i} + \bar{R})} = \bar{W}_i.$$

Hence, variable w_i must be assigned to values $W_h \leq \bar{W}_i$ From our computational experience, this mechanism reduces by ten times the size of the solution space to be generated.

A second procedure works as follows: if we have assigned variables $w_0 \dots w_k$ and:

$$\tau(w_0 \dots w_k) := \sum_{i=0}^{k} w_i + (N-k-1) L_W > \bar{W}$$

```
EXPLORE(k)
{
    if (k ≥ 1 and k ≤ N − 1)
        if (τ(w_0 . . . w_k) > W̄)
            return;
    if (k ≥ N)
    {
        if(w = (w_0 . . . w_{N−1}) is feasible)
        {
            calculate f(w);
            update the best w*;
        }
    }
    else
    {
        for(h = 0 . . . m − 1)
            if(W_h ≤ W̄_k)
            {
                w_k := W_h;
                EXPLORE(k + 1);
            }
    }
}
```

TABLE II

THE EXPLORE PROCEDURE FOR THE CP ALGORITHM

where:

$$L_W = \min\{W_h | h = 0 \ldots m-1\}$$

there are no values left for the remaining $w_{k+1} \ldots w_{N-1}$ so as to obtain a feasible assignment of all the variables. Hence, a feasible assignment of the variables $w_0 \ldots w_k$ must satisfy the condition:

$$\tau(w_0 \ldots w_k) \leq \bar{W}$$

Our algorithm assigns values to variables, starting from the first variable to the last; as soon as it finds out that the current partial assignment of the variables cannot be extended to the others in a feasible way, it stops searching for values of the remaining variables. Then it goes backtracking and tries a different partial assignment for the same subset of variables. More precisely the algorithm consists in the recursive procedure outlined in Table II.

C. A Genetic Algorithm

The methods used in the previous section provide an exact solution in finite time. Unfortunately, this time is often too long for the approach to be usable in practice.

To test the algorithms under this aspect, we solve several problem instances with N users, $6 \leq N \leq 16$, $m = 20$ and $k = 20$ on an average notebook computer. Within 10000 seconds the CP implementation always returned the optimal solution, while the solver CPLEX often could not solve the correspondent ILP in time. When both approaches returned a solution, the solutions were identical.

In this interval of N, the CP approach turned out to be faster. When $N < 6$ the advantage in using CP for what concerns the computational time is of several order of magnitude; in fact, on our test computer the problem is solved in less than one second. However, the computational time of the CP algorithm increases rapidly with N, until it fails to supply a solution in an acceptable amount of time with $N > 16$.

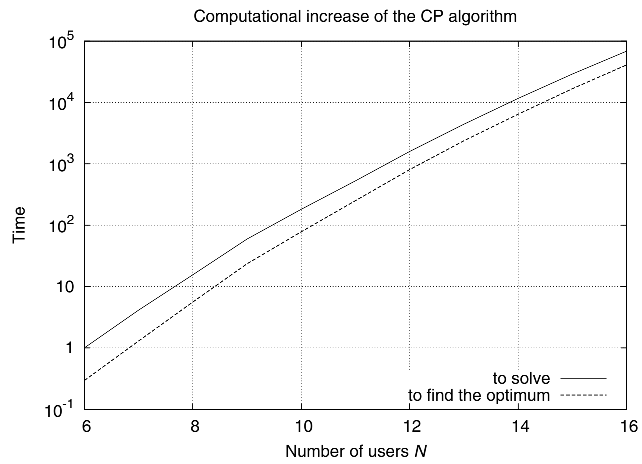

Fig. 1. Normalized computational time of the CP approach as a function of N

To show the trend of this computational complexity increase, consider Fig.1. Here, two curves are plotted: the time required by the CP algorithm to explore the whole solution space and the time to find the optimal solution; the values are normalized to the time to explore the solution space when $N = 6$. Both values increase exponentially with the number of users, which makes the CP approach very fast for small N but inapplicable to large networks.

Thus, especially if one wants to implement an optimization technique directly at the Access Point of the cellular system, more efficient techniques from the computational complexity perspective are needed. To this end, we outline a Genetic Algorithm (GA), which give results very close to the optimal solution but with lower computational complexity.

A GA for an optimization problem finds a feasible point with good value by imitating the *Natural Selection*, the process of adaptation to the environment of living beings. We have developed a standard Genetic Algorithm (GA) for the original problem. In our GA an *individual* is a vector of feasible powers. The *value* of an individual w is the value of the objective function on w. The *gene* of an individual is a component of the vector. An initial *population*, i.e., a set of individuals, is obtained by generating random vectors in $[0, \bar{W}]$ and *repairing* them if they are not feasible. This means that they are modified as little as possible, so as to make them feasible.

The population is iteratively modified by applying the *mating*, the *mutation* and the *selection* operations in this order. This is done for a fixed number of iterations. The mating operation consists of choosing some of the individuals and *mating* them, that is, substituting them with their *children*. The children of two individuals are obtained by swapping the second half components and repairing the two possibly infeasible resulting vectors. The mutation operation consists in mutating all the individuals. A mutated individual is obtained by altering some of its components with a random value $\delta \in [-\Delta, +\Delta]$, where $\Delta < \frac{\bar{W}}{2}$. The selection operation produces the population of the next iteration. This new population has the same number of individuals of the current one. Its individuals are chosen from the current population, possibly more than once. The probabil-

Parameter (symbol)	value
spreading bandwidth ($\mathcal{B}$)	2.5 rate units
max rate per terminal ($\bar{R}$)	1 rate unit
cell min radius (d_0)	5 m
cell max radius (d_1)	100 m
gain at 1 m (A)	-28dB
Hata path loss exponent (β)	3
Shadowing parameter (σ)	6dB
max transmission power ($\bar{W}$)	20dBm
Thermal noise floor (η)	-38dBm

TABLE III
PARAMETERS OF THE PHYSICAL SCENARIO

Parameter	value
k	$-\log 0.9$
ψ	1.0
ϕ	1.0
μ	2.0
ϵ	4.0

TABLE IV
PARAMETERS OF THE ECONOMIC SCENARIO

ity of an individual being chosen for the next *generation* (the population of the next iteration) depends on the *rank* of the individual within the current population, the higher the rank the higher the probability. The rank of an individual within a population is the position of the individual in the population ordered by increasing value. The returned individual is the best individual found in all the generated populations.

V. RESULTS

First of all, let us briefly outline the working assumptions used for deriving results. For analytical convenience, we will consider a simple usage-based linear pricing scheme, i.e., we assume $p(r) = \alpha r$ with constant α. This scheme is sensible, as it is simple and hence it might be appreciated by the users. Also, it is likely that data-transfer services tariff users according to the amount of downloaded data [7]. On the other hand, the framework can be applied to other kinds of pricing as well. The graphs presented in the following have been plotted by considering different α's, hence emphasizing the dependence of the Economic Metrics on the pricing setup. Utilities are assumed to be sigmoid-shaped functions [1], which span from 0 to 1. The mathematical expression is

$$u(r) = (r/r_s)^\zeta/(1 + (r/r_s)^\zeta),$$

where the concavity is regulated by an exponent ζ uniformly distributed in the range $2 \div 4$ and the flex point r_s is taken as a random value uniformly distributed in the range $0.1 \div 0.4$.

The acceptance-probability functions are chosen as follows:

$$A(u_i, p_i) \triangleq 1 - e^{-k\cdot(u_i/\psi)^\mu\cdot(p_i/\phi)^{-\epsilon}}, \qquad (7)$$

where k, μ, ϵ, ψ, ϕ are positive constants. This is justified by the ideas presented in [5]. In particular, the parameters k, ψ and ϕ are normalization constants, whereas the role of ϵ and μ is to make the acceptance probability similar to the Cobb-Douglas curves [17], widely used in economics to represent in general a relationship of economic nature, where an exponent, called "elasticity", is appropriately chosen in order to have a good match with experimental measurements curve. Then, in our case, the parameters ϵ and μ represent the sensitivities of the users to the price and the utility. These data describing the technical parameters of the scenario can be found in Table III.

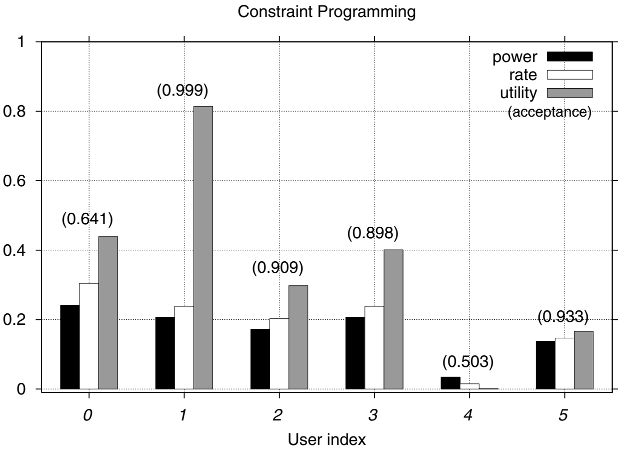

Fig. 2. Values allocated by CP for an instance with $\alpha = 1.0$.

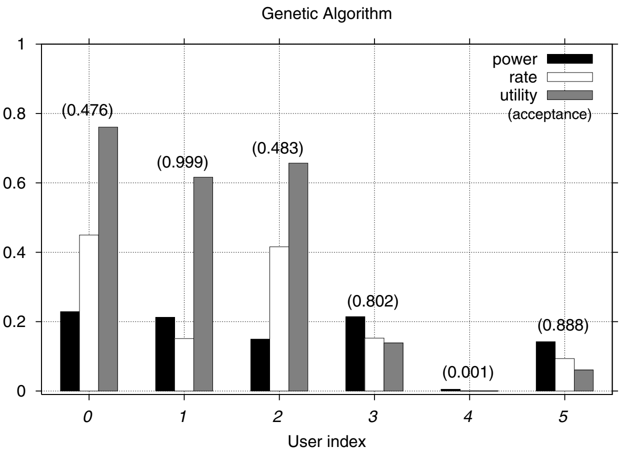

Fig. 3. Values allocated by GA for an instance with $\alpha = 1.0$.

In the following we will always consider the goal function of the optimization to be the Economic Metric R, i.e., the revenue. However, this choice, like all the purely technical choices made about pricing and utility functions, can be replaced without affecting the general optimization framework.

Under this framework, the proposed approaches are able to find the optimal solution in terms of rate and power allocation, which determine the economic metrics. In Figs. 2 and 3 we report the solutions found by the exact approach (in particular, the CP algorithm) and the Genetic Algorithm in terms of allocated values. Being the price equal to 1, the allocation vector of the rates weighed by the acceptance probabilities can be also seen as revenue generation due to a single user.

The optimal solution found by the CP approach and the GA are quite similar, at least for what concerns the allocated power, which is the main allocated quantity. However, a small variability of the allocated power reflects in higher variations of the achieved rate and even more of the perceived utility. This

N	A	B	C	D	E	F
6	1.00	0.29	522.01	43.12	28.94	146.60
7	4.12	1.29	2075.8	983.1	66.52	759.68
8	15.53	5.59	2672.5	11.76	6.294	2.1001
9	59.41	23.53	2742.5	11.17	4.117	0.4749
10	181.42	78.00	2440.0	6.882	1.118	0.0881
11	527.61	250.82	681.94	1.235	1.235	4.92e-3
12	1592.2	805.88	1954.4	1.328	1.328	1.67e-3
13	4427.3	2371.9	889.71	1.464	1.464	6.20e-4
14	11664	6434.7	4160.6	1.518	1.518	9.17e-5
15	29099	16788	16.597	1.653	1.653	9.81e-5
16	68611	41005	745.83	1.764	1.764	4.29e-5

TABLE V

COMPUTATIONAL TIMES: A = CP TIME TO SOLVE. B = CP TIME TO FIND THE OPTIMUM. C = GA TIME TO SOLVE. D = GA TIME TO REACH 99% OF CP VALUE. E = GA TIME TO REACH 90% OF CP VALUE. F = TRADEOFF CP VS. GA (RATIO BETWEEN D AND B).

means that a power allocation which is optimal for one Economic Metric does not guarantee itself efficiency also for the other Economic Metrics. It is also emphasized that the optimal solution is obtained with extreme unfairness in the revenue-optimal solution, since some users (in particular, user 4 in the figures) achieve low data rates.

However, apart from the exact determination of the optimal solution to the problem, our investigation has also the practical goal of determining efficient solutions (not necessarily the optimal ones) in a reasonable time. In other words, in realistic situations the resource manager needs to find a good feasible solution for the original problem. With this purpose in mind, we have seen that, among the strategies compared, the GA is the best choice, because it finds a good feasible solution of the original problem in a very small amount of time.

Table V compares the GA with the CP algorithm in terms of computational time, by reporting several time metrics. The values for the ILP problem are not reported since CPLEX was able to provide solutions only for instances with low N. The last column plots also the tradeoff between the CP and the GA, seen as the ratio between the time required by the CP algorithm to find the optimum and the time required by GA to find the 99% of this value. This value represents the capability of the GA vs. the CP algorithm of finding a good solution: the lower the value, the better the relative performance of the GA. We can see that, when $N \geq 9$, the GA finds a feasible solution whose value is $\geq 99\%$ of the value of the best solution given by the CP while taking less time. When $N > 12$, the GA gives a solution which is at least 99% as good as the CP one, in less than $\frac{1}{1000}$ of the time required by the CP to give that solution.

However, an important remark needs to be kept in mind: the GA and the CP do not face exactly the same problem. The CP chooses the powers only on a finite set of values, while the GA chooses them on a continuous interval. Because the objective function is the same and the CP's possible powers are feasible for the GA, it can happen that the GA finds a feasible solution with a higher value than the optimal solution found by the CP. If the number of iterations is sufficiently large, the GA always finds a feasible solution with higher value.

With this in mind, we can now compare the average performance obtained by the CP approach and the Genetic Algorithm. Figs. 4–7 show the evaluation of the Economic Metrics for the optimal solutions found by the CP and the GA. In these figures, only networks with $N = 6$ have been considered. Note that the ILP procedure always obtains the same results of the CP. In these results, the GA has been run for 300 iterations only, in order to have quick evaluation time, by considering the best individual found up to the last iteration.

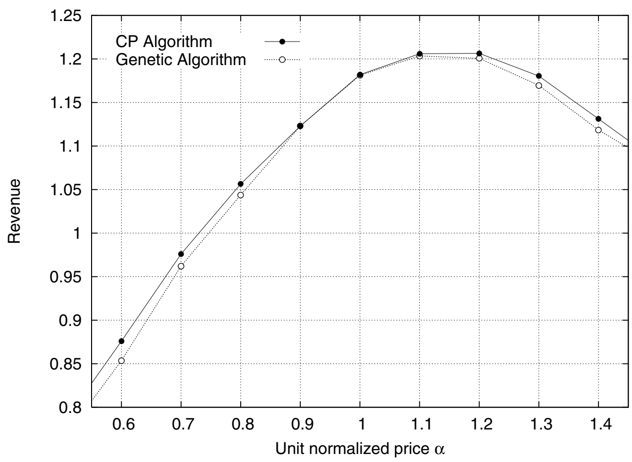

Fig. 4. Revenue R as a function of the unit price α.

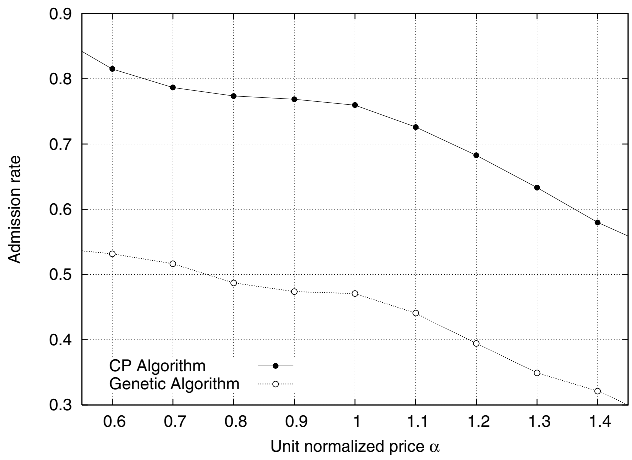

Fig. 5. Admission rate S/N as a function of the unit price α.

Note that all these evaluations are affected by strumental error which arises from the approximation in the discretization of rates and powers. Due to this error, for example, it might happen than the solution found by the heuristic Genetic Algorithm is even better than the exact optimal one, since the models are different. Indeed, this can be eliminated almost completely through an appropriate increase of k and m. However, this of course increases the time required for the solution.

Let us discuss the results in more detail: Fig. 4 shows that the Genetic Algorithm exhibits good match with the optimal solutions found in the exact manner by the CP approach. Note that in all our simulations we run the Genetic Algorithm always converges to the global maximum, even though this conclusion might not be true in general.

Also note that the optimal solutions for the revenue-maximization problem presents a very smooth behavior which highlight the presence of an optimal price setting for the provider, in order to maximize the revenue. Also, the confi-

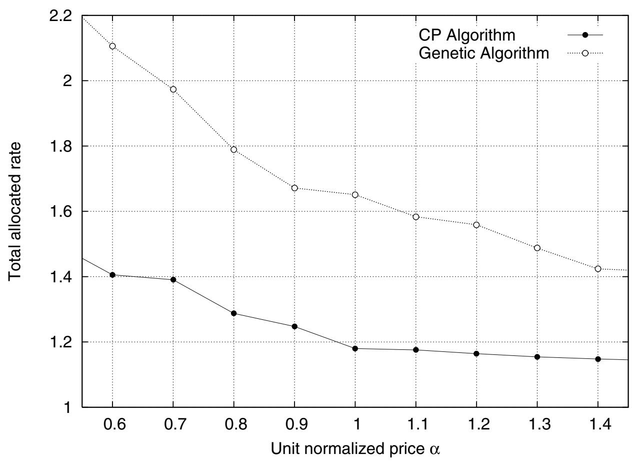

Fig. 6. Total allocated rate T as a function of the unit price α.

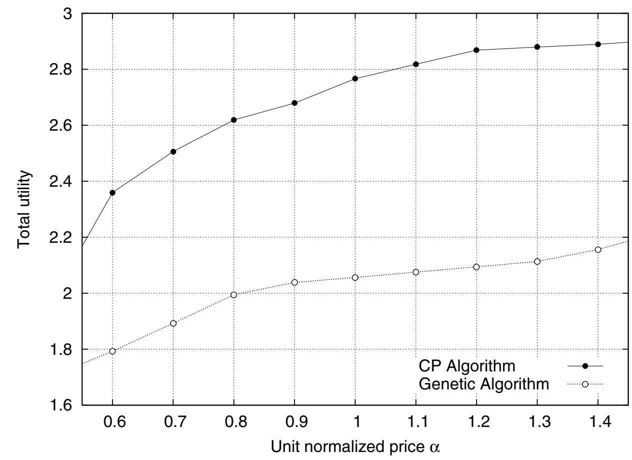

Fig. 7. Total utility U as a function of the unit price α.

dence of the results is very good, with slightly larger oscillations in the Genetic Algorithm due to its heuristic nature and also its usage of randomly generated points. On the other hand, the curves presented in Figs. 5–7 are more subject to dispersion, since the plotted metrics are not directly involved in the optimization procedure.

In particular, as is clear from Fig. 6 the optimum found by the CP approach allocated less resource than the Genetic Algorithm, where conversely Figs. 5 and 7 show that the number of admitted users and the total utility of the users are higher in the solution found with the CP approach. This means that the CP approach still succeeds in better using the resources. However, the results of the Genetic Algorithm for what concerns the metric under optimization are very close to the optimum. Thus, for an effective allocation with applicability to real cases, it is worth considering also to include several Economic Metrics in the goal function, in order to have an efficient RRM under many points of view.

VI. Conclusions and Future Work

The contribution of this paper is a unified framework to study the optimization of RRM when the trade-off between utility and price is taken into account. We formalize the optimization problem for the rate allocation which aims at maximizing a given Economic Metric and also propose linearized versions, which are solved with several approaches.

In particular, we used both exact and heuristic methods. Due to the intrinsic complexity of the exact procedures, the heuristic approach seems to be a more practical solution to solve such problems. In particular, an interesting possibility is offered by strategies based on genetic algorithms, which exhibit a good trade-off between accuracy and computational complexity.

As a future research topic, it might be interesting to investigate how these results can be useful in terms of gaining understanding about the properties of the Radio Resource Management, in particular in which way they can be applied to the design of realistic and effective strategies.

To this end, we think that a possible alternative, not directly explored here but left for further research, involves Admission Control strategies to decrease the problem complexity. In fact, the main issue to address when evaluating optimization is that the complexity of the problem strongly increases with the number of users. Thus a strategy, even heuristic, to decide first of all if for some of the users it is not even worth trying the resource allocation, can be identified, and this seems to be a valid alternative to the problem of the computational capacity which arises in such large networks.

References

[1] M. Xiao, N. B. Shroff, and E. K.-P. Chong, "Utility-based power control in cellular wireless systems," in *Proceedings INFOCOM 2001*, 2001, pp. 412–421.

[2] C. Zhou, M. L. Honig, S. Jordan, and R. Berry, "Utility-based resource allocation for wireless networks with mixed voice and data services," in *Proceedings 11th International Conference on Computer Communications and Networks*, 2002.

[3] V. Siris, "Resource control for elastic traffic in CDMA networks," in *Proceedings ACM MobiCom 2002*, 2002, pp. 193–204.

[4] H. Yaïche, R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.

[5] L. Badia, M. Lindström, J. Zander, and M. Zorzi, "Demand and pricing effects on the radio resource allocation of multimedia communication systems," in *Proceedings Globecom 2003*, 2003, vol. 7, pp. 4116–4121.

[6] T. Heikkinen, "On congestion pricing in a wireless network," *Wireless Networks*, vol. 8, no. 4, pp. 347–354, 2002.

[7] R. J. Gibbens and F. P. Kelly, "Resource pricing and the evolution of congestion control," *Automatica*, vol. 35, pp. 1969–1985, 1999.

[8] D. Famolari, N. Mandayam, D. Goodman, and V. Shah, *Wireless Multimedia Network Technologies*, chapter 1: A New Framework for Power Control in Wireless Data Networks: Games, Utility and Pricing, pp. 289–310, Kluwer Academic Publishers, 1999.

[9] C. Saraydar, N. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 291–303, 2002.

[10] L. Song and N. Mandayam, "Hierarchical SIR and rate control on the forward link for CDMA data users under delay and error constraints," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1871–1882, 2001.

[11] K. Kar, S. Sarkar, and L. Tassiulas, "A simple rate control algorithm for maximizing total user utility," in *Proceedings INFOCOM 2001*, 2001, vol. 1, pp. 133–141.

[12] F. Berggren, S.-L. Kim, R. Jäntti, and J. Zander, "Joint power control and intracell scheduling of DS-CDMA nonreal time data," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1860–1870, 2001.

[13] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[14] A. M. Viterbi and A. J. Viterbi, "Erlang capacity of a power controlled CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 892–900, 1993.

[15] L. A. Wolsey, *Integer Programming*, John Wiley and sons, New York, 1998.

[16] "ILOG Cplex 9.0, user's manual and reference manual," ILOG, S.A., `http://www.ilog.com/`, 2004.

[17] H. R. Varian, *Intermediate Microeconomics: A Modern Approach*, Norton, New York, 1999.