On Utility-based Radio Resource Management with and without Service Guarantees

Leonardo Badia*, Michele Zorzi*†
* Dept. of Engineering, University of Ferrara, Via Saragat 1, 44100 Ferrara, Italy
† Dept. of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy
{lbadia,zorzi}@ing.unife.it

ABSTRACT

In this paper we discuss utility functions models to study Radio Resource Management. Our goal is to identify the characteristics of the wireless systems which make such theoretical models, though challenging, very useful, as they allow to quantify the Quality of Service and to analytically investigate the users' satisfaction. Moreover, we show how, within a utility-based framework, it is possible to also study economic issues, besides more conventional technical aspects such as throughput or system capacity. Thus, when economics are taken into account by considering the financial needs of the provider and the users' reaction to prices, we are able to study wireless systems in a more realistic and appropriate way. Another key contribution of this paper is a discussion on how utility functions should be applied to the particular case of the radio resource. To this end, we extend classic economic concepts with an original proposal, better able to model the nature of the wireless services. Finally, by giving both analytical insight and numerical results, we compare different classes of RRM strategies and explore the relationships between Radio Resource Allocation, pricing, provider's revenue, network capacity and users' satisfaction.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: network communications, wireless communication; I.6.5 [**Model Development**]: modeling methodologies

General Terms

Management, Performance, Economics

Keywords

Rate allocation, Utility functions, Service Guarantees

1. INTRODUCTION

The evolution of wireless communication systems is very rapid and more and more services are offered through the wireless medium. This implies an increasing importance of the Radio Resource Management (RRM), which must consider not only technical efficiency but also whether different constraints given by users' preferences are met. This and other economic aspects (like the provider's income, which makes the business model sustainable) can have a heavy impact on the entire system. In particular, this is a strong point when elastic traffic [1,2] is considered, i.e., the system allows a tunable allocation, with possibly different degrees of service.

In the recent literature, several researchers have proposed to introduce utility-based RRM to take into account these facts [3–6]. The approach consists in defining utility functions to represent the service appreciation of the users, depending on the amount of allocated resource, which is an effective way to mathematically represent the Quality of Service (QoS) provided. In this way, in fact, one can both control the QoS requirements of every user and also evaluate the overall network welfare, defined as an aggregate of the utilities (e.g., the sum, if they are considered to be additive).

The introduction of utility functions offers an analytical tool to represent the relationship between users and services. However, it is arguable that such a connection can be realistic if the pricing issues is neglected. Since the services do not come for free, users would also likely prefer to have a cheap service. Hence, we need to include into the model the trade-off between offered QoS (seen through utilities) and price paid. In the present paper we will refer to the MEDUSA (Micro-economic Elastic Decentralized Users' Service Acceptance) model proposed in [7], even though the conclusions are quite general, i.e., other models can be used as well. The only thing which is requested by the study presented here is to represent in some way the choices of each user according to both utility and price in a reasonable manner (i.e., users always prefer higher utility and lower price). In particular, the MEDUSA model introduces a satisfaction probability for each user, and allows the evaluation of every performance metric for the satisfied users only.

In this paper we present the following novel contributions: we expand the previously defined model to gain theoretical insight about the RRM in the case of dynamic resource assignment. We propose in particular to introduce what we call *backward utility functions*, which effectively represent service degradation. As is well known, service degradation for ongoing connections might be very annoying, thus it may penalize the perceived QoS. Under this framework, we analyze the goals of achieving both satisfactory revenue for the provider and welfare for the users, implying that a trade-off has to be cut between offered QoS and pricing.

Moreover, we consider two possible approaches to utility-based RRM, which are a reservation scheme in which the initial allocation is kept constant (which leads to no degradation, but potentially refuses many connection) and a best effort strategy where the allocation is adapted dynamically according to incoming requests of the users. We show how this latter policy depends on the degradation of the utilities when the allocated resource is changed. Then, an an-

MSWiM'04, October 4-6, 2004, Venezia, Italy.

alytical model is proposed to take into account the degradation of the utilities due to decreased assignment. Finally, the performance of the resulting allocation schemes is discussed and compared by means of an extensive simulation campaign.

The paper is organized as follows: in Section 2 we introduce the model used to represent the allocation of the radio resource and the users' appreciation of their assignment. In Section 3 we discuss how to extend this model, defined in a static manner, to take into account also rate degradations. This will result in the original proposal of *backward utilities*, i.e., a hysteresis effect to account for degradations in the assignment. Then, Section 4 theoretically discusses and compares the possible strategies to allocate resources, whereas Section 5 presents the results applied to a given simulation scenario. Finally, in Section 6 the conclusions are drawn.

2. THE MEDUSA MODEL

We give here a short summary of the MEDUSA model, whose basic idea is to quantify the level of satisfaction for each of the N users, which depends on both the perceived QoS and the price paid. Thus, for each user i, an *Acceptance value* $A_i \in [0,1]$ is defined, that depends on the utility u_i and on the paid price p_i. This is a mathematical evaluation of the degree of satisfaction for the ith user, which can be used in an analytical investigation, by assuming the following statistical meaning: every user has a probability of reaching satisfaction equal to A_i. Only the resource allocated to satisfied users is efficiently used. In the following, we will assume that allocation to unsatisfied users is wasted, since, for example, they leave the service. This impacts on every metric related to resource allocation, like throughput and revenue earned by the provider, which must be evaluated by considering only users accepting the service.

Several expressions are possible to define a sensible acceptance function. In [7] the following was proposed:

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

where k, μ, ϵ, ψ, ϕ are appropriate positive constants. The exponents μ and ϵ regulate the sensitivity to utility and price, respectively, whereas ψ, ϕ and k are merely normalization constants (a reference utility, a reference price, and the opposite of the logarithm of a reference value for A, respectively). Note that both u_i and p_i depend on the allocated resource r_i. Thus, the shape of the acceptance probability as a function of r_i depends on the functions $u_i = u(r_i)$ and $p_i = p(r_i)$. The MEDUSA model can be exploited to study different aspects of the Radio Resource Management. In particular we focus on rate assignment for CDMA–like networks. For the sake of simplicity, let us identify r_i with the transmission rate of terminal i considered on average, i.e., channel variations due to fading are neglected. Several statistical average metrics coming from each user allocated with rate r_i, which implies utility u_i and price p_i, can be evaluated directly and simply. For example, the revenue R earned by the provider, the number S of admitted users, the amount T of allocated resource and the total utility U for admitted users can be computed as:

$$R = \sum_{i=1}^{N} p_i A(u_i, p_i)\,, \tag{2}$$

$$S = \sum_{i=1}^{N} A(u_i, p_i)\,, \tag{3}$$

$$T = \sum_{i=1}^{N} r_i A(u_i, p_i)\,, \tag{4}$$

$$U = \sum_{i=1}^{N} u_i A(u_i, p_i)\,, \tag{5}$$

respectively. In this sense, we can choose different alternative goals for the RRM, like the maximization of the total revenue which leads to different conclusions than the welfare maximization, as shown in [6]. Within this framework, it is also possible to jointly evaluate several metrics of interest: for example, it can be assumed that the short-term goal of the provider is to improve the revenue earned from the RRM, whereas in the long run it is desirable to improve the social welfare, i.e., the total utility coming from the assignment. This also represents a possibility of supplying a better service, being the same resource better utilized. From a general point of view, all these metrics concur to the objective of an efficient RRM, seen from either the customer's or the provider's perspective (or both). In fact, these two goals of satisfying the provider and the customers are not independent: the higher the users' satisfaction, the more the allocated resource, and hence the higher the potential revenue. On the other hand, pricing and allocation strategies of the provider determine the behavior of the users. We might exploit the simplicity of a direct evaluation of the four previously defined metrics, which allows to test the performance of RRM by means of simulation.

Consider the goal of maximizing the revenue by referring to Equation (2) [1]. The following intuitive property is implicitly represented: too high prices drive customers away (A_i decreases) and yield very little revenue, whereas too low prices can easily be afforded by all users, but also with low revenue as a result. This can be formalized by stating the existence of an optimal pricing policy, i.e., an expression for $p_i(\cdot)$, which is the one that corresponds to the maximum revenue. Note however that, when the resource to allocate is scarce, as is usually assumed, this optimal pricing is also achieved when the capacity is fully utilized.

In general a purely analytical investigation of the problem is hard. Moreover, if done under the assumption of having p_i's as general as possible it will result in an unrealistic model, since $p_i(\cdot)$ is reasonably the same at least within the same service class, and in general similar for all users. Note that technological constraints impose that r_i is between 0 and a maximum value r_{max}, assumed for the sake of simplicity equal for all terminals, which depends on the considered technology. We will also consider additional assumptions for the utility and pricing functions. Note that these choices are only for the sake of analytical convenience, but other assumptions can be used as well (provided that the basic properties previously discussed are satisfied).

To model the utilities, we employ sigmoid curves, which are well-known functions often used to describe QoS perception [5]. The following expression will be employed to represent these curves:

$$u(r) \triangleq \psi \frac{(r/x_s)^{\zeta}}{1+(r/x_s)^{\zeta}}. \tag{6}$$

The parameters $\zeta \geq 2$ and $x_s > 0$ tune the utilities, so that they might be different for every user. Note that the value x_s is such that $u(x_s) = \psi/2$. Within simulations, x_s and ζ are random variables for each user. This definition results in having utilities which are bound in the range $[0, \psi]$. This aspect will be extended in the next Section to allow a broader definition of the concept of utility.

For what concerns the pricing schemes, several contributions [2, 8, 9] have shown that the tariff setup has the double role of achieving revenue and coordinating users. In this work we want to discuss a pricing function defined a priori, which is applied in the same manner for every user. We consider a simple usage-based linear pricing scheme, i.e., we assume $p(r) = \xi r$ with constant ξ. In fact, a realistic pricing function is also often required to be

[1] Similar conclusions, not discussed here for lack of space, can be derived from Equations (3)–(5) for the other metrics.

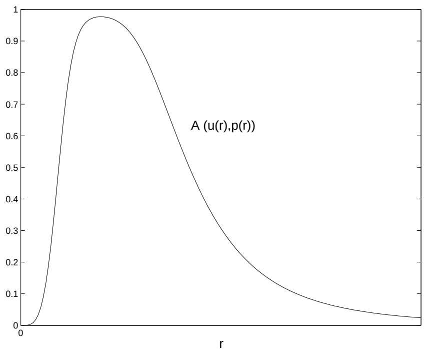

Figure 1: Behavior of the acceptance probability $A(u_i, p_i)$ as a function of r for price $p(r_i) = \xi r_i$ and sigmoid utility

simple, as users usually do not like to deal with complicated tariff plans. The linear pricing policy, i.e., each user pays proportionally to the allocated resource, is easily understandable and hence suitable for our purposes. However, other data-transfer services tariff considering the amount of downloaded data can be used [10, 11].

Figure 1 presents an example of a possible resulting acceptance probability curve, by integrating Equations (1), (6) and the linear pricing, which presents many similarities with the utilities involving battery life and perceived error probability presented in [4], for which several theoretical results have been derived. These results are not discussed here in detail; however, they only rely on simple regularity properties for $u_i(\cdot)$ and $p_i(\cdot)$, e.g., being quasi-concave non decreasing functions [2].

Prior to starting this analysis, we also need a general framework for the RRM scheme which determines the resource allocation in a tunable way. We assume that users arrive at separate instants. After each arrival, the resource manager tries to allocate the new user, say user i, by giving an assignment r_i. Since the resource manager can not exactly predict the requests of the upcoming users in the queue, we might think that it adopts a greedy strategy fixed a priori, for example the one described as follows (this reasoning is applicable however to any other conceptually equivalent strategy). We consider a rate allocation strategy based on the derivative of the utility In fact, $u'(r)$ describes the subjective perception of changes in the rate assignment. We assume that the provider performs a greedy-algorithm allocation by determining a priori a threshold value $\vartheta > 0$, which numerically summarizes the general bandwidth management strategy into a single parameter. The initial rate assignment for user i, called r_{i0}, is:

$$r_{i0} \triangleq \max(\{0\} \cup \{r \in]0, r_{max}] \, : \, u_i'(r) \geq \vartheta\}) \,. \tag{7}$$

Due to the saturation of the utility functions, the greater the value of ϑ, the lower the initial rate r_{i0} proposed to user i. This implies that $\vartheta \to 0$ means, roughly speaking, very high utility supply, whereas higher values of threshold imply the allocation of lower rates.

This allocation rule is however not independent of feasibility constraints. We must in fact take into account the soft capacity of a CDMA system [12], which means to consider the feasibility of the rate assignment in an interference-limited system. This can be done in many ways. In the present paper we simply translate the rate to signal-to-interference ratio (SIR) by means of the well-known Shannon's capacity formula:

$$\gamma_{t,i} = 2^{r_i/W} - 1 \,, \tag{8}$$

[2]Usually, technological constraints also impose additional properties, like upper-limits for the utilities, which simplify the problem even further.

where $\gamma_{t,i}$ is the target SIR for user i. In this way it is possible to check if the initial value r_{i0} determined for the ith user by Equation (7) is feasible if considered together with the values assigned to the previously allocated $i - 1$ users. If the set of the target SIRs for all users is feasible, this rate assignment is kept. Else, a decision has to be taken, according to the different RRM policies as will be shown in next Section. Note that in every case we will come up with an assignment $\mathbf{r}$ which is iteratively updated. At each step however, we must use the MEDUSA framework to determine if users are satisfied with their assignment or leave the service. Since the allocation can be dynamically changed, also this decision can change during time, and this will be addressed in the next Section.

3. ELASTIC TRAFFIC AND BACKWARD UTILITIES

In the previous evaluations, the focus is on user i, which is the one currently under admission. To have a realistic analysis however, we should consider that in interference-limited systems every new admission may decrease the quality of already connected users [12]. Hence a new admission can be damaging, if previously admitted users have decreased resource assignments or more likely if they refuse the degradation in the QoS due to the new admission.

In more detail, different approaches are possible to deal with traffic elasticity. For example, we could exploit the possibility of tuning the assigned rate only once, i.e., when the user is allocated for the first time. In this way, the allocation is tunable but static. Else, it is possible to consider a dynamic soft tuning of the offered QoS, which can be exploited to improve the efficiency of the assignment, even during connection. In this case, we need also to see how users react to variations in the QoS, which can occur when their assignment is reduced. This is not the only possibility of experiencing a degradation: if a new user is admitted, the QoS might be decreased even for the users whose rates are not explicitly reduced. In this case, the degradation is rather due to the interference increase, which might cause the infeasibility of the SIR requirements.

We want therefore to extend the MEDUSA model to describe the reaction to a dynamic assignment, i.e., to consider variations from the initial assignment. The framework outlined previously is fully satisfactory in the case of a conservative approach which does not perturb the allocation of already admitted users. Hence, already admitted users are untouched until the end of their transmission and there is no allocation dynamics to account for. However, since resource is scarce, after a certain number of admissions a new one will be infeasible. This impossibility of admitting a new user i can be seen as related to the infeasibility of the vector of the transmitted powers. Hence, if the traffic were inelastic we would have to block every new user. Since we are instead considering elastic traffic, it is possible to adjust the transmission rates, and we can exploit this property to try to increase the number of admitted users. By considering the framework of Section 2, we have two basic choices: the first one is to try an assignment lower than r_i only for user i. In this way, the already established connections enjoy the same quality. Hence, in this case there is a slight QoS guarantee, of course subject to the condition of finding available resource. From a greedy point-of-view, the provider might however be interested in trying to "squeeze" as many users as possible into the systems, to ultimately increase the revenue. Such an analysis might be seen also from the perspective of Congestion Control, that means adaptive RRM under Rate Control to improve the allocation and allow correct network operability, as discussed in Chapter 9 of [13]. This adaptation is possible only if the assignment is not fixed for the whole connection, but might be changed, which requires to take into account the users' reaction to assignment variations.

As a first step, consider an extension of the Acceptance probability according to the definition of conditional probability [14]. Assume what follows: if two assignments $r^{(0)}$ and $r^{(1)}$ are characterized by a value of Acceptance probability equal to $A^{(0)}$ and $A^{(1)}$, respectively, we define a *conditional Acceptance probability* of accepting $r^{(1)}$ given that $r^{(0)}$ was acceptable, called $A^{(1|0)}$ and equal to:

$$A^{(1|0)} = \begin{cases} A^{(1)}/A^{(0)} & A^{(1)} \le A^{(0)} \\ 1 & A^{(1)} > A^{(0)} \end{cases} \tag{9}$$

This Equation only exploits the concept of conditional probability users with the implicit assumption that users will never refuse quality improvement captured by higher $A(u,p)$, hence it offers a good model to analyze assignment variations made a priori. However, for the purpose of considering *dynamic* variations of the QoS, Equation (9) is inappropriate. That is, it is unable to describe rate control performed during the connection, where the impact of rate modification can be heavier. On the one hand, where the QoS is increased, it is obvious that none of the users already in the system will be disappointed. Thus, the conditional probability of accepting the variation is always 1, so we keep this part of Equation (9) unchanged. On the other hand, Equation (9) says that the probability of leaving the service because of a service degradation is proportional to the degradation amount. However, real services are heavily affected by QoS decreases during an ongoing connection and it is likely that the correct relationship is more than proportional to the degradation amount. In other words, if the value of r is decreased during service supply, this will make the service even less valuable, so that the utility and the probability of accepting the decrease must be even lower. Thus, Equation (9) should consider a lower value of $A^{(1|0)}$ when degradations occur dynamically, i.e., when the service is already started. Of course a detailed scheme would consider the duration of the interval in which the service evaluation has been equal to $A^{(0)}$, and take into account that different services behave differently in this respect[3]. For the sake of simplicity, we will neglect these differences.

What we will consider is that the utility of an assignment is different if it results from a degradation or it is the same assignment done at the beginning. In the following we will speak of *backward utility*, i.e., we have different utility curves, for increasing or decreasing quality. The initial assignment always increases the resource, hence $u(r)$ is the forward utility, which is an increasing function such that $u(0)=0$, as discussed in Section 2. On the other hand, the backward utility can even go below 0. If this happens, the acceptance probability will be surely 0.

For the sake of simplicity, in the following we model the added annoyance of the QoS degradation, when it occurs during connection, with a term included in the utility and depending on a positive *loss parameter* called L. Its value can for example be seen as the relative weight of the two different annoying events of being served at first with low quality or experiencing degradation to low quality during an ongoing connection. This is a generalization of the well-known trade-off between blocking and dropping probability in Admission Control [15]. If $u(r)$ is the forward utility function we can define the backward utility as a modified version of the utility, called $\upsilon(r,u,r^{(0)})$, as follows:

$$\upsilon(r,u,r^{(0)}) = \begin{cases} u(r) - Lu^{(0)}(u^{(0)} - u(r)) & r \le r^{(0)} \\ u(r) & r > r^{(0)}, \end{cases} \tag{10}$$

where $u^{(0)} = u(r^{(0)})$.

[3]For example, data transfer sessions are probably kept alive if the degradation occurs almost at the end. However, for real time data, like a sport match, such a quality decrease might be very annoying.

When $L = 0$, the reaction of the users is always the same, regardless of the time in which the service is re-evaluated. Hence, the a priori evaluation is the same as the real-time evaluation. If $L > 0$ we have a *fragile* QoS, i.e., the utility is lower if the current assignment results from a degradation. An infinitely fragile QoS ($L = \infty$) decreases suddenly to 0 if any degradation occurs, no matter how small. In this sense, a reasonable range of values for L is around the maximum utility ψ (which will be confirmed by simulations in Section 5). Approximately, values of L are meaningful if between 0 and 2ψ. At this point we also need to slightly modify the definition of $A(u,p)$ given by Equation (1) by considering the case in which $u(\cdot)$ is replaced by $\upsilon(\cdot)$, which can assume negative values. Thus,

$$A(\upsilon,p) = 0 \quad \text{if } \upsilon < 0 \;, \tag{11}$$

whereas we use the same definition (and henceforth we still call the function A) when $\upsilon \ge 0$.

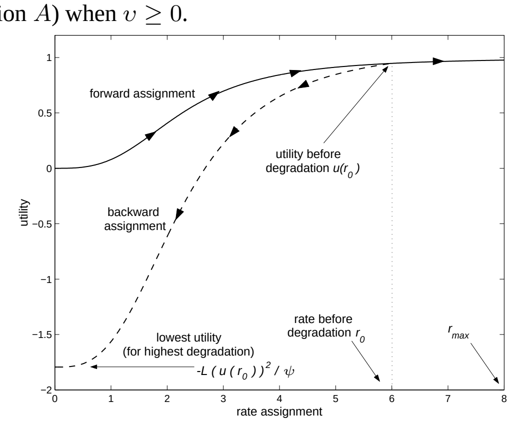

Figure 2: Forward and backward utility for $L = 2\psi$.

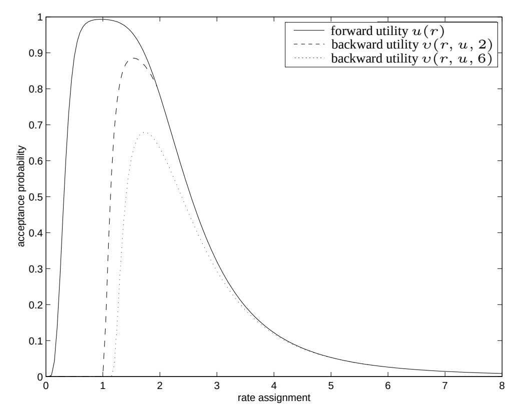

Figure 3: Forward and backward acceptance prob. for $L = 2\psi$.

Figure 2 reports the backward utility for $\psi = 1$ and $L = 2\psi = 2$ when the rate r is allocated between 0 and $r_{max} = 8$ and the utility is sigmoid-shaped. The behaviour of such a curve which vaguely reminds, to some extent, the hysteresis of magnetic materials. Figure 3 shows instead the behavior of the acceptance value $A(\upsilon,p)$ with backward assignments from different $r^{(0)}$'s. Note that the backward acceptances are always below those in the forward case, i.e., $A(\upsilon(r,u,r^{(0)}),p) \le A(u(r),p)$ for every $r^{(0)}$. It is also true that a decrease of the rate does not always imply a lower value for $A(u,p)$. In fact, one should remember that when r decreases the price is in general lower. For example, if the assigned rate goes from $r^{(0)} = 6$ to $r^{(1)} = 3$, Figure 3 shows that $A(\upsilon(r^{(1)},u,r^{(0)}),p)$ is still considered an improvement with respect to $A(u^{(0)},p)$. However, this happens only since the assignment $r^{(0)} = 6$ is not the most preferred by the user. This effect is mainly due to the fact that the price is decreased, even though the quality is slightly decreased too (and this latter effect is emphasized by having $L = 2\psi$). Now we have modeled the users' reaction to dynamic assignment in addition to QoS and pricing *per se*, which implies an amplification of QoS degradation. This allows us to compare different policies of dynamic resource management, which will be done in the next Section.

4. COMPARISON BETWEEN DYNAMIC RRM STRATEGIES

Consider a network where users are characterized by backward utilities when rate degradation occurs. In this scenario, we identify two main classes [16, 17] of management, characterized by a different behavior with respect to congestion control and elasticity. The first one is called *Guaranteed Performance* (GP), since it assures a reservation of a fixed resource for the whole connection. Reservations or priorities are widely used in communication networks [18]. In particular, a GP approach can neglect the backward utilities, since the resource assignment is never decreased. Also, GP management implements only a simplified form of congestion control, since users are conservatively admitted but the assignment is static. The second class is a *Best Effort* (BE) allocation, which provides no guarantee about possible future variations of the QoS after admission [1–3].

It is easy to understand that the GP management has the advantage of keeping a constant grade of service, i.e., the satisfaction level of the connected users is fixed. On the other hand, the main drawback is that the resource might be wasted, by allocating resource to users either with lower utility or lower contribution to the revenue. The BE management instead provides only partial QoS guarantees. On average, we might expect that this latter policy is more effective, since it allows admission to a larger number of users, though their quality is possibly subject to degradation.

In the following we will compare these two strategies both from the theoretical point of view and, after having discussed possible implementations of the Best Effort strategy, by means of simulations using the performance metrics previously defined. From a *naïve* point of view, the Best Effort RRM has the advantage of always allowing improvement of the metrics, at least theoretically. This would happen since R, S, T and U as defined in Equations (2)–(5) seem to be non decreasing if N is increased (i.e., more users are considered). Indeed, we must consider the loss due to the quality decrease of already admitted users in case that i is admitted. This value depends on how the rate of a generic user j already in the system is decreased to admit user i into the system. Assume that the resource allocation vector (considered to have i elements) is $\mathbf{r}^{(0)} = (r_1^{(0)}, r_2^{(0)}, \ldots r_{i-1}^{(0)}, r_i^{(0)} = 0)$ before the admission of user i and $\mathbf{r}^{(1)} = (r_1^{(1)}, r_2^{(1)}, \ldots r_{i-1}^{(1)}, r_i^{(1)})$ afterwards. Hence, it is possible to write:

$$S^{(i,0)} = \sum_{j=1}^{i-1} A(u_j(r_j^{(0)}), p(r_j^{(0)})) \ , \qquad (12)$$

$$S^{(i,1)} = \sum_{j=1}^{i-1} A(\upsilon_j(r_j^{(1)}, u_j, r_j^{(0)}), p(r_j^{(0)})) + +A(u_i(r_i^{(1)}), p(r_j^{(1)})) \ , \qquad (13)$$

which is the total satisfaction (or also the average number of admitted users) before and after the admission of the ith user. The sum can be made over all users which have requested admission before i, since if a user $j < i$ has already terminated its call or has refused the proposed QoS we assume to have $r_j^{(0)} = r_j^{(1)} = 0$.

Parameter (symbol)	value
number of cells	19
bandwidth (W)	20 rate units
max assignable rate (r_{max})	8 rate units
cell radius (d)	500 m
gain at 1 m (A)	-28dB
path loss exponent (α)	3.5
shadowing parameter (σ)	8dB
mean SNR at cell border	20dB

Table 1: List of parameters of the simulation scenario

Parameter	value
N (number of users)	160
ζ	$2 \div 20$
x_s	$0.1 \div 0.9$
k	$-\log 0.9$
ψ	1.0
L	1.0
ϕ	1.0
μ	2.0
ϵ	4.0

Table 2: List of micro-economic parameters

Similar Equations can be written for the metrics R, T, U. The evaluation after the admission is done not only by considering the contribution of the ith user, which is always an improvement, but also a possible degradation, taken into account by considering the backward utilities υ_j instead of u_j. Equations (12)–(13) also indicate that the performance of the BE RRM depends on how the rates are translated from $\mathbf{r}^{(0)}$ to $\mathbf{r}^{(1)}$. However, our investigations on this matter have shown that this choice only marginally affects the performance, provided that the degradation of the rate is fair and relatively small for every user. In the following, we will consider a degradation from $\mathbf{r}^{(0)}$ to $\mathbf{r}^{(1)}$ by decreasing the SIR of connected users with rate larger than 0, whose number is called $\mathcal{N}$, in steps of $1 \text{ dB}/\mathcal{N}$, until the vector of the powers is feasible. Results produced with other policies are still in agreement with the ones shown in the following. One might argue that the variation in the acceptance probability implies that users might completely refuse the service, considered unsatisfactory due to degradation. Hence, to have a full description of the model, the resource left by users who quit from the service could be reassigned to the users who stay in the system. This study however would imply iterative (and possibly long) evaluations of these negotiations, hence it is left for further research. In the sequel, we will always assume that if a user j accepts the assignment $r_j^{(0)}$ but considers the degradation to $r_j^{(1)}$ unacceptable, this resource is left unused by the system, which is a conservative approach.

Finally, note that this procedure can be iteratively repeated (but we avoid to give the whole formula, whose notation would be cumbersome) for every successive user which can not be feasibly allocated. At this point however, we should consider "backward backward utilities". That is, assume that also user $i+1$ can not be admitted into the system and we want to represent the degradation of the service for a user j who has already experienced a degradation from $r_j^{(0)}$ to $r_j^{(1)}$ during admission of user i. In this case, the backward utility to consider for user j is not represented by $\upsilon(r, u, r_j^{(1)})$ but by $\upsilon(r, \upsilon^{(0)}, r_j^{(1)})$, where $\upsilon^{(0)} = \upsilon^{(0)}(r, u, r_j^{(0)})$.

5. RESULTS

In the following we will present comparative results for GP and BE RRM, which show the achievable performance by such strategies. We perform simulations in a CDMA system, with N users placed with uniform spatial distribution over hexagonal cells, which are wrapped around as usually done to avoid border effects. Table 1 shows the propagation parameters of the simulation scenario and Table 2 reports the parameters of the MEDUSA model.

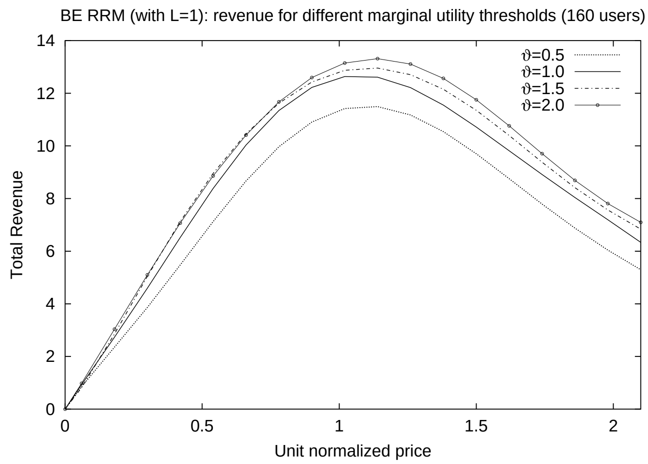

Figure 4: BE management: revenue R for linear price $p(r) = \xi r$ as a function of ξ.

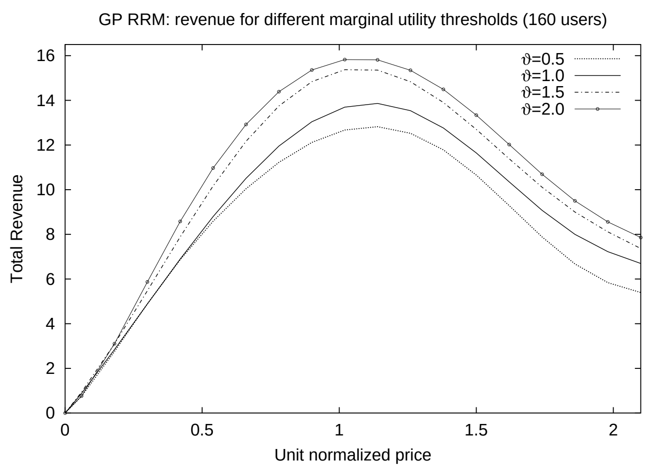

Figure 5: GP management: revenue R for linear price $p(r) = \xi r$ as a function of ξ.

First of all, the revenue (metric R) is considered in Figures 4–6. Figure 4 investigates the BE resource management, with different values of the threshold ϑ. In this case the impact of the users' reaction to dynamic allocation has been taken into account by letting the loss parameter L equal ψ. The same analysis is carried out in Figure 5, but for the GP strategy. Here, the value of L has no impact on the results, being the allocation for every user untouched while it is in the system. As can be seen, both strategies exhibit better performance when $\vartheta = 2.0$ is considered. More in general the larger the allocation threshold, the higher the revenue. This means that in the considered scenario it is better to allocate a small amount of resource per user to a large number of users, rather than the opposite. However, from other considered scenarios (whose results are not shown here since this investigation is outside the scope of the present paper) it emerges that this conclusion can not be generalized. Hence, in general the optimal threshold setup depends on many factors, in particular on the number of users in the network. More interesting is to look at Figure 6, where the performance of GP and BE RRM is compared (the latter by considering different values of L). In particular, as will be done in the following for the other metrics, we evaluate the BE RRM for different values of L and the GP RRM, which does not depend on L, and we compare the two approaches for the case $\vartheta = 1.0$ (different values of ϑ present entirely similar results).

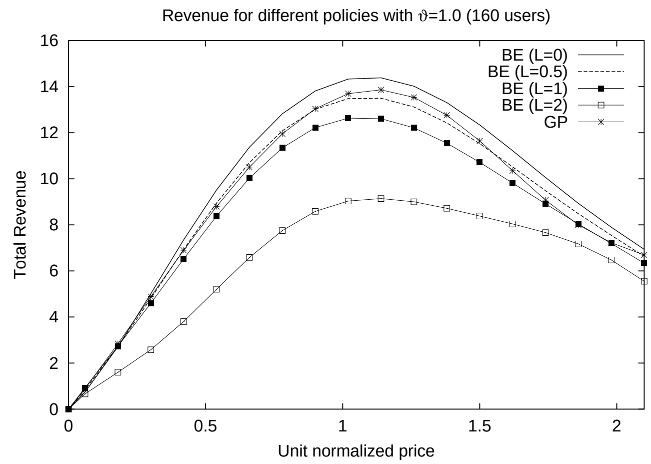

Figure 6: Revenue R for linear price $p(r) = \xi r$: comparison of different RRM approaches.

It is emphasized that when no loss occurs, i.e., $L = 0$, the BE strategy outperforms GP in terms of revenue. However, the revenue decreases as L increases. So, the BE RRM is a good choice only when users are not so sensitive to dynamic quality degradations.

Secondly, we investigate in Figures 7 and 8 the number of admitted users S as a function of the price, for different values of the threshold. The two Figures represent the normalized value of S for the BE and GP strategies, respectively. It is clearly emphasized that the BE approach outperforms GP in this respect. In particular, if the price is relatively low, BE can admit a percentage of users close to 90%, even though the capacity is scarce. For more reasonable price values, the number of admitted users is anyway larger for the BE RRM policies than for GP. Hence, the only way to have an efficient RRM in terms of number of admitted users is to fully exploit the traffic elasticity by applying rate tunability every time it is possible. In other words, if having a large number of customers is included among the provider's goals, it is very difficult to allocate them in a guaranteed way. These conclusions are summarized in Figure 9, where the two RRM policies are compared for the case of $\vartheta = 1.0$. It is shown that this general conclusion is still valid even if L has a higher value, i.e., when users are very sensitive to service degradations.

Figure 10 investigates instead the comparison between BE and GP for the metric T, which represents the amount of allocated resource. In this case the detailed curves for the two policies are not shown, since they are very similar. In fact, what Figure 10 points out is that both strategies are able to allocate almost the same amount of resource, thus indicating that both strategies are efficient in not wasting network capacity. This means that both of them are well designed to allocate as much resource as possible, compatibly with users' satisfaction. In particular, when the price is low the BE approach allocates slightly more resource, and this is reasonable, as the elasticity of the traffic is better exploited. On the other hand, as the price increases, GP becomes more appreciated by the users, which accept to pay high tariffs only if the performance is guaranteed. For higher values of L, the amount of resource allocated by BE RRM is slightly decreased, which means that the performed assignment may become less efficient for high sensitivity of the users to quality degradations.

Finally, the total utility (Figure 11 vs. Figure 12, and comparison in Figure 13) is higher for the GP strategy, and this is again due to the absence of guaranteed performance in the BE approach. Note also that the total utility of the BE strategy increases when L in-

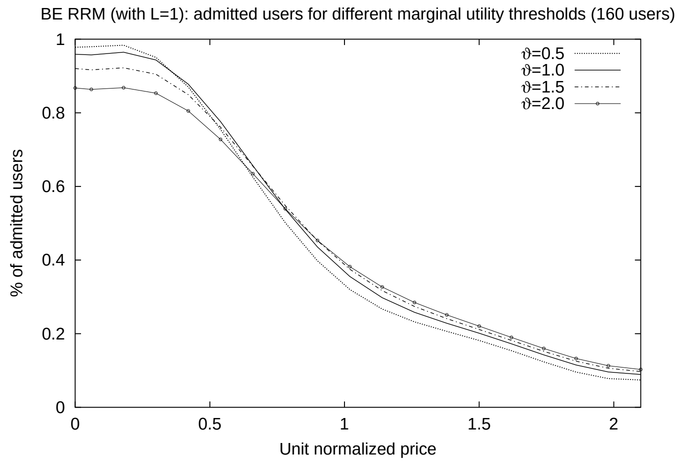

Figure 7: BE management: admission rate S/N for linear price $p(r) = \xi r$ as a function of ξ.

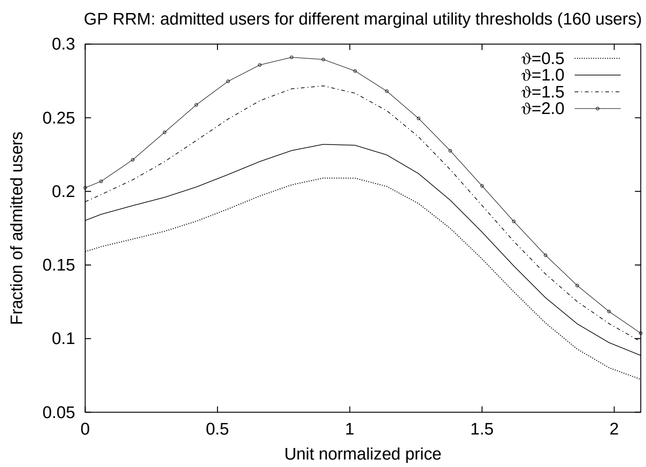

Figure 8: GP management: admission rate S/N for linear price $p(r) = \xi r$ as a function of ξ.

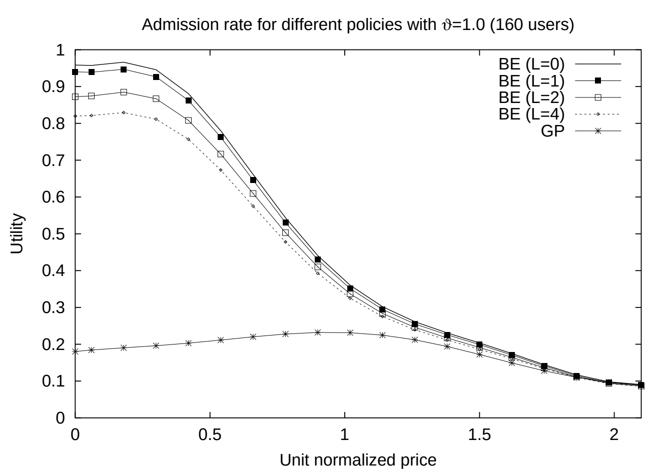

Figure 9: Admission rate S/N for linear price $p(r) = \xi r$: comparison of different RRM approaches.

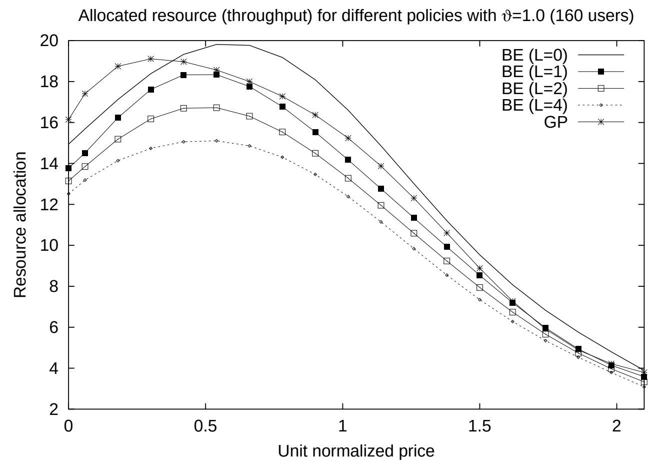

Figure 10: Allocated resource (Throughput T) for linear price $p(r) = \xi r$: comparison of different RRM approaches.

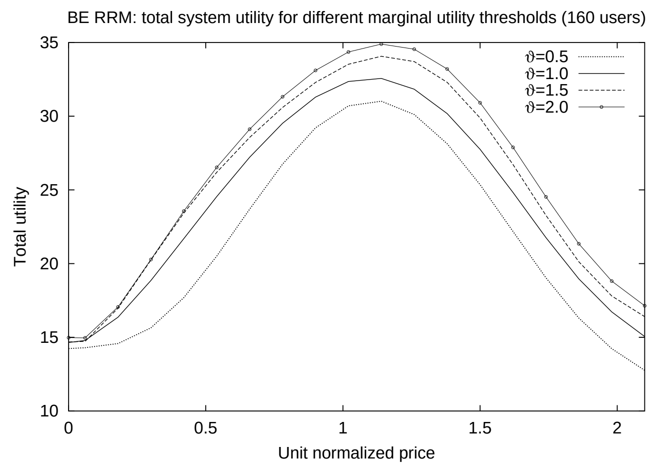

Figure 11: BE management: total utility U for linear price $p(r) = \xi r$ as a function of ξ.

creases, which might appear as counterintuitive. This is due to the fact that when L is higher, there are also more users pulled out of the system. The remaining ones are the users with relatively higher utility. Hence, the higher L, the higher the total utility but also the lower the number of admitted users S. This can be seen as a more general consequence of the fact, shown in [6] for a simpler case, that revenue and total utility in certain cases are contrasting goals.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have explored some capabilities of utility-based RRM and have presented a model to describe a user-centric management of rate allocation and highlight pros and cons of different allocation classes. This theoretical approach is useful from the analytical point of view to design efficient RRM strategies. As a further contribution, our work has strengthened the analysis with practical examples, by investigating in particular the comparison between two possible allocation strategies, the RRM with Guaranteed Performance (GP) and the totally Best Effort (BE) RRM.

In general, the choice between these two approaches might be related to choices made a priori by the provider. It is true that the BE strategy allocates more users, but the quality is poorer and especially the revenue is lower. In the analyzed scenario, the BE is

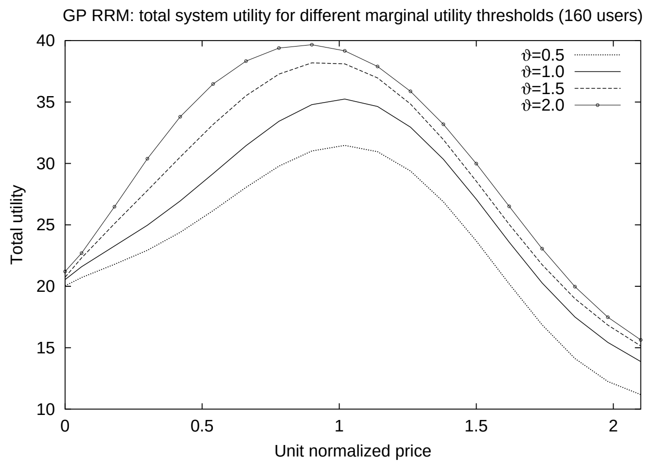

Figure 12: GP management: total utility U for linear price $p(r) = \xi r$ as a function of ξ.

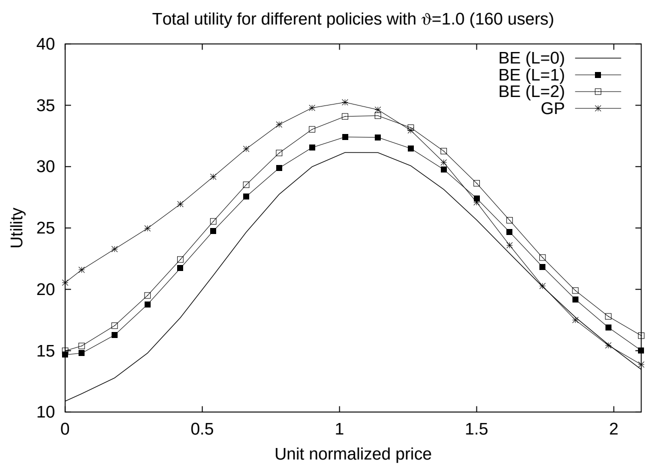

Figure 13: Total utility U for linear price $p(r) = \xi r$: comparison of different RRM approaches.

penalized by the introduction of the backward utilities. In other words, when decreasing the allocated resource has a strong negative effect on users' satisfaction, it seems that a Guaranteed Performance could be the only solution. Otherwise, users may want to leave the system, and the provider's revenue is decreased. On the other hand, if the provider's goal is mainly to keep its own users and to also acquire new customers, the GP RRM is clearly inappropriate. Thus, none of the two approaches can be the ultimate solution, rather a trade-off between them could be more suitable.

To cut the trade-off, a possible strategy can be to preventively admit users with a smarter criterion, which can, with an exact analysis or even simply by means of heuristic rules, try to forecast the impact of the admission on the entire system. In this way the philosophy will still be Best Effort, since there is no guarantee on the achievable QoS, but the performance metrics can be improved, at least on average. This strategy can lead to developing a Microeconomic Admission Control, which can be the goal of future research.

7. REFERENCES

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

[2] S. Shenker, D. D. Clark, D. Estrin, and S. Herzog, "Pricing in computer networks: Reshaping the research agenda," *ACM Computer Communication Review*, vol. 26, pp. 19–43, 1996.

[3] 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.

[4] D. Famolari, N. Mandayam, D. Goodman, and V. Shah, "A New Framework for Power Control in Wireless Data Networks: Games, Utility and Pricing," in: *Wireless Multimedia Network Technologies*, R. Ganesh, K. Pahlavan, and Z. Zvonar, eds., pp. 289–310. Kluwer Academic Publishers, 1999.

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

[6] 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.

[7] 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.

[8] 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.

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

[10] C. Comaniciu, N. Mandayam, D. Famolari, and P. Agrawal, "Admission and flow control for multimedia CDMA," in *Proceedings IEEE Conference on Multimedia and Expo (ICME 2000)*, vol. 3, pp. 1265–1268, 2000.

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

[12] 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.

[13] H. Holma and A. Toskala, *WCDMA for UMTS*. John Wiley and sons, New York, 2000.

[14] R. M. Spiegel, *Theory and Problems of Statistics*. McGraw-Hill, New York, 1996.

[15] L. Badia, M. Zorzi, and A. Gazzini, "A model for threshold comparison call admission control in Third Generation cellular systems," in *Proceedings IEEE ICC'03*, vol. 3, pp. 1664–1668, 2003.

[16] E. Altman, A. Orda, and N. Shimkin, "Bandwidth allocation for guaranteed versus best effort service categories," in *Proceedings INFOCOM '98*, vol. 2, pp. 617–624, 1998.

[17] A. Thavarajah and W. H. Lam, "Comparison study of two channel assignment schemes [cellular mobile systems]," *Electronics Letters*, vol. 36, no. 4, pp. 354–355, 2000.

[18] H. Jiang and S. Jordan, "A pricing model for high speed networks with guaranteed quality of service," in *Proceedings INFOCOM 1996*, vol. 2, pp. 888–895, 1996.