Wireless Netw
DOI 10.1007/s11276-007-0015-z

Dynamic utility and price based radio resource management for rate adaptive traffic

Leonardo Badia · Michele Zorzi

Abstract This paper explores analytical Radio Resource Management models where the relationship between users and services is mapped through utility functions. Compared to other applications of these models to networking, we focus in particular on specific aspects of multimedia systems with adaptive traffic, and propose a novel framework for describing and investigating dynamic allocation of resources in wireless networks. In doing so, we also consider economic aspects, such as the financial needs of the provider and the users' reaction to prices. As an example of how our analytical tool can be used, in this paper we compare different classes of RRM strategies, e.g., Best Effort vs. Guaranteed Performance, for which we explore the relationships between Radio Resource Allocation, pricing, provider's revenue, network capacity and users' satisfaction. Finally, we present a discussion about Economic Admission Control, which can be applied in Best Effort scenarios to further improve the performance.

Keywords Rate allocation · Utility functions · Service guarantees · Micro-economics · Radio Resource Management · Pricing

Part of this work has been presented at the conference ACM/IEEE MSWiM 2004, Venice (Italy).

L. Badia (✉)
IMT Lucca Institute for Advanced Studies, piazza S. Ponziano 6, 55100 Lucca, Italy
e-mail: lbadia@ing.unife.it

M. Zorzi
Department of Information Engineering, Università degli Studi di Padova, Via Gradenigo 6/B, 35131 Padova, Italy
e-mail: zorzi@ing.unife.it

1 Introduction

The evolution of wireless communication systems is very rapid, and more and more services are being 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 (e.g., 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 rate adaptive traffic [1, 14] 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 these facts into account [6, 10, 17, 19]. In these approaches, utility functions are used to describe the relationship between the users' service appreciation and the allocated resource, which is an effective means 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., their sum).

The introduction of utility functions offers an analytical tool to represent the relationship between users and services. However, it is arguable whether such a connection can be realistic if the pricing issue is neglected. Since the services do not come for free, users would 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 [2], even though the conclusions are more general, i.e., other models can be used as well. The only required aspect 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). The important aspect of the MEDUSA model we will use here is that it introduces a satisfaction probability for each user, and allows the evaluation of every performance metric for the satisfied users only.

In this paper, we expand the previously defined model to gain theoretical insights 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 the effect of service degradation. As is well known, service degradation for ongoing connections might be very annoying, thus penalizing 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 also to refusing many connection) and a best effort strategy where the allocation is adapted dynamically according to the incoming requests from the users. We show how this latter policy depends on the degradation of the utilities when the allocated resource is changed. Thus, our proposed utility model is able to account for the degradation of the utilities due to decreased assignment. The performance of the resulting allocation schemes is discussed and compared by means of an extensive simulation campaign in a wireless cellular scenario.

This comparison is further analyzed under the point of view of Admission Control techniques using economic criteria [8, 11, 17], In particular, we show that such an Economic Admission Control (EAC) can be applied to the Best Effort scenario, being able to significantly improve the performance. At the same time this approach obtains a more efficient resource management with respect to an allocation with fixed guarantees on the QoS.

The application of the model depends on the network capacity, which in turn depends on how multiple users access the channel. Throughout the paper, we will use, where possible, very general capacity models for wireless networks. When specific assumptions are required, e.g., in the numerical evaluation, we will refer to a Code-Division Multiple Access (CDMA) case, since it is often considered in previous work on this subject [5, 14]. However, our approach can be applied without significant changes to any centralized interference-limited system (e.g., to Wireless LANs in [3]). Moreover, due to its generality, it can be applied to other systems with centralized management as well, with slight modifications.

Note that for such systems it is also very important to determine the optimal allocation in the sense of many metrics (throughput, fairness. . .) which is challenging due to the so-called *soft capacity* property [16]. For this reason, contributions employing utility functions often apply optimization techniques and game-theoretical approaches [11] to seek the maximal performance of the system. Since the purpose of this paper is to introduce backwards utilities and to discuss their role in driving the RRM and economic behavior of the system, we will adopt a simpler approach, leaving more refined optimizations (which are of course of interest in this context) as a topic for future research.

The paper is organized as follows: in Section 2 we introduce the model to represent the allocation of the radio resource and the users' appreciation of their assignment. In Section 3 we extend this static model in order to take into account also rate degradations. This results in the original proposal of *backward utilities*, i.e., a memory effect to account for degradations in the assignment. Section 4 theoretically discusses and compares possible strategies to allocate resources, and investigates the concept of Economic Admission Control, whereas Section 5 applies these issues to a given simulation scenario. Finally, in Section 6 some 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. 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 and is a mathematical evaluation of the satisfaction for the ith user.

Similar concepts can be found also in [6, 14]. However, differently from these contributions, our use of A_i aims at the same time at the analytical evaluation of QoS as well as economic quantities. Moreover, we do not investigate optimality conditions, but rather we explore the *dynamic* allocation of radio resource. In fact, we adopt the following statistical interpretation: every user has a probability of reaching satisfaction equal to A_i. Moreover, 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, such as throughput and revenue earned by the provider, which must be evaluated only on users accepting the service. In other words, we implicitly assume that dissatisfied users do not pay for the service. This is true if we focus on the long term objective of the provider: dissatisfied users are lost customers and represent an economic loss.

Several expressions are possible to define a sensible acceptance function. In [2] the following one 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 determine 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 $u_i = u_i(r_i)$ and $p_i = p_i(r_i)$.

The MEDUSA model can be exploited to study different aspects of the Radio Resource Management. In particular, in this work we focus on rate assignment in centralized (e.g., cellular) wireless networks. This choice will allow us to investigate different general policies for service provisioning, as well as Admission Control. Also, we focus on systems where the central provider has the possibility of renegotiating the allocation. In wireless networks carrying multimedia traffic, although rate allocation is performed at the beginning of a connection, reallocation may occur during the connection lifetime as well. In this sense, this framework is more general than more traditional schemes in which calls can only be blocked or dropped.

For the sake of simplicity, we identify r_i with the transmission rate of terminal i considered on average, i.e., without tracking the instantaneous channel variations due to fading. Several statistical metrics coming from each user allocated with rate r_i, with corresponding utility u_i and price p_i, can be evaluated directly. For example, the average revenue R earned by the provider can be computed as

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

and similarly the average number S of admitted users is

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

In these evaluations, A_is are employed with a statistical meaning. Likewise, it would be possible to evaluate other metrics, such as the average amount of allocated resource T or the average total utility or network welfare U, which are obtained by replacing p_i in (2) with r_i and u_i, respectively. The entire analysis which we will develop in the following can be readily adapted to any of these metrics by following the same rationale.

This framework can also be used to compare different objectives of the RRM, e.g., as in [19], where it is shown that the maximization of the total revenue leads to different conclusions than the maximization of the network welfare, or to derive multiple-objective optimizations, since the short-term goal can be different from that in the long-term. However, note also that all these metrics are inherently correlated, since in general the higher the users' satisfaction, the more the allocated resource, and hence the higher the potential revenue. For these reasons, many of the results shown for one metric can be inferred also for the other ones in a qualitatively similar fashion.

Consider the goal of maximizing the revenue,[1] defined in (2). 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 suggests that there exists an optimal pricing policy, i.e., an expression for $p_i(\cdot)$ which maximizes the revenue. 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 $p_i(\cdot)$ should be the same function 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 $r_{\max}$, which is assumed to be the same for all terminals for the sake of simplicity, and which depends on the considered technology. We will also consider additional assumptions for the utility and pricing functions. 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 [17]. The following expression will be employed to represent them:

$$u(r) \triangleq \psi \frac{(r/x_s)^{\zeta}}{1 + (r/x_s)^{\zeta}}. \tag{4}$$

The parameters $\zeta \geq 2$ and $x_s > 0$ tune the utilities, so that they might be different for different users. Note that the value x_s is such that $u(x_s) = \psi/2$. In our simulations, x_s and ζ are random variables for each user. With this definition, utilities are in the range $[0, \psi]$. This aspect will be extended in the next section to allow a broader meaning of utility.

For what concerns the pricing schemes, several contributions [12, 13] have shown that the tariff setup has the double role of achieving revenue and coordinating users. In this

[1] As discussed, similar conclusions can be derived for any other metric.

work we want to discuss a pricing function defined *a priori*, which is applied in the same manner to every user. In general, a realistic pricing function is also often required to be simple, as users usually do not like to deal with complicated tariff plans. For the sake of simplicity, we consider in our practical evaluations a usage-based linear pricing scheme, i.e., we assume $p(r) = \xi r$, with the same constant unit price ξ for all users. However, other more realistic tariffs can be used [5, 7] with the same approach. The reason of our choice is merely its conceptual simplicity.

We assume that users arrive at separate instants. We do not investigate in detail the impact of the different traffic patterns: for our purposes, it is sufficient to think of a list of users to be allocated sequentially in the network. This roughly models the queue of users entering the system, requesting a connection and hence either accepting or refusing the rate proposed by the provider. We also assume that the users accepting the assignment remain in the system for an indefinitely long time, in order to investigate the system performance at saturated capacity. We assume that the resource manager tries to allocate any new user, say user i, by giving an assignment r_i. Since the allocator 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, since $u'(r)$ describes the subjective perception of changes in the rate assignment. We assume that the provider determines a single threshold value $\vartheta > 0$. The initial rate assignment for user i, called r_{i0}, is:

$$r_{i0} \triangleq \max(\{0\} \cup \{r \in]0, r_{\max}] \, : \, u_i'(r) \geq \vartheta\}) \,. \tag{5}$$

Equation (5) means that we choose the highest rate which gives a marginal utility larger than ϑ. If there exists no rate with such marginal utility, we give a rate assignment equal to 0. 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 ϑ imply the allocation of lower rates. Note that this strategy is meant to be simple, and intentionally not related to any optimization of the allocation policy, which is a well known topic in micro-economics, but is beyond the scope of the present paper, which simply discusses a framework to evaluate users' reaction to *any* allocation policy. However, further investigation about how to efficiently select ϑ may be an interesting subject for future work.

The allocation rule operates jointly with the feasibility constraints given by system capacity. Since we mostly refer to interference-limited systems, which are characterized by the so called *soft capacity* property, we take a general approach, in order to avoid further relationships due to power values and interference conditions, which would complicate the description. Thus, we use the well-known Shannon's capacity formula to translate the assigned rate r_i to user i into a signal-to-interference ratio (SIR) value:

$$\gamma_{t,i} = 2^{r_i/B} - 1 \,, \tag{6}$$

where B is the system bandwidth. In this way, we obtain the target SIR $\gamma_{t,i}$ for user i. It is possible to check if the initial value r_{i0} determined for the ith user by Eq. (5) is feasible if considered together with the values assigned to the previously allocated $i-1$ users. To do this, we simply consider the feasibility of solving the Power Control problem [18] when the vector of SIRs for all users is determined through (6) and all propagation parameters (which will be given later in Section 5) are known. For the rest of this paper, we will refer to the uplink of a cellular system, and therefore we will check if the powers found through the Power Control solution are between 0 and a value $P_{\max}$ which is assumed to be the maximum power that a terminal can use. Note that the downlink allows a very similar formulation where this feasibility check is replaced by one on the sum of all powers and the available power at any base station.

If the set of the target SIRs for all users is feasible, this rate assignment is kept. Else, a decision has to be made, according to the different RRM policies as will be shown in the next section. Note that in every case we will come up with an assignment $\mathbf{r} = \{r_1, r_2, \ldots\}$ which is iteratively updated (e.g., by decreasing some of the rates assigned). For generality, we do not introduce any lower bound for the SIR, which can be indefinitely decreased by repeated admissions. However, there is actually an inherent limitation due to the rate degradation, which occurs together with the SIR descent and at a certain point will no longer be accepted by the users. This is evaluated through the MEDUSA framework to determine whether users are satisfied with their rate assignment or refuse the service and leave the system. Since the allocation can be dynamically changed, also this decision can change during time, and this will be addressed in the next section.

3 Dynamic rate allocation and backward utilities

In the previous evaluations, the focus is on user i, which is the one being considered for 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 [16]. A way to see the degradation is to consider that the interference increase might lead to the infeasibility of the SIR requirements for some of the existing users, due to the newly admitted one. We can assume a dynamic soft tuning of the initially offered QoS, but in this case, we also need to see how users react when their assignment (or quality) is reduced. Hence, as a result of a new admission, previously admitted users might even decide to leave the system, if they refuse to accept the degradation in their perceived QoS.

The model previously shown needs to be extended to incorporate also the reaction to a dynamically variable assignment. In particular, this is necessary to capture situations where new admissions cause infeasibility of the assignments for previously connected users. This might be due, in interference-based scenarios, to an SIR below the target given by (6). However, as already said, our model does not explicitly prevent these situations from arising, but rather captures this as a dynamic degradation which is statistically evaluated. In fact, depending on users' reactions, a small degradation might be acceptable, even though in general this is an annoying effect.

To apply rate variations policies which avoid unnecessary degradation, we consider different options. Assume that user i is currently requesting service, and this allocation would cause some of the already existing connections to become infeasible. A first possibility to deal with infeasibility would be to lower the rate assignment for user i. Decreasing r_i would allow to reduce the interference that the new admission causes to the already connected users. In this case there is a slight QoS guarantee for ongoing connections, whereas the rate of new connections can be decreased and become unsatisfactory. This point is eventually reached when a new user can not be admitted at all, or equivalently its rate must be reduced to 0.

From a greedy point of view, the provider might however be interested in exploiting rate adaptation not only to deal with infeasibility of interference constraints, but also to squeeze as many users as possible into the system, to ultimately increase the revenue. This adaptation is possible only if the assignment is not fixed for the whole connection, but might be changed, which again requires to take into account the reaction of the users to assignment variations.

As a first step, consider an extension of the Acceptance probability according to the definition of conditional probability [15]. 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:

$$A^{(1|0)} = \begin{cases} A^{(1)}/A^{(0)} & A^{(1)} \leq A^{(0)} \\ 1 & A^{(1)} > A^{(0)} \end{cases} \tag{7}$$

This equation only exploits the concept of conditional probability with the implicit assumption that users will never refuse a quality improvement captured by higher $A(u, p)$, hence it still works for static assignments.

However, (7) is not completely appropriate to describe *dynamic* variations of the QoS. If the service is improved, the conditional probability of accepting the variation is correctly 1, but if the value of r is decreased during service, this will make the service even less valuable. Whereas (7) states that the probability of accepting a degradation is proportional to the extent of this degradation, we claim that in this case the utility perceived by a decreased service and the probability of accepting the decrease must be even lower. Thus, (7) should consider a lower value of $A^{(1|0)}$ when degradations occur dynamically, i.e., when the service has already started. A more 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.[2] For the sake of simplicity, we will neglect these differences in this discussion.

According to this rationale, we explicitly model the fact that the utility of an assignment is different according to whether it is the result of a degradation or an initial assignment. 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 simplicity, in the following we model the added annoyance of the QoS degradation during a connection, with a term included in the utility and depending on a positive *loss parameter* called L. Its value can be seen as, e.g., the relative weight of the two different annoying events of being served at first with low quality or of experiencing a degradation to that same quality during an ongoing connection. This is a generalization of the well-known trade-off between blocking and dropping probability in Admission Control [4]. If $u(r)$ is the forward utility function, we can define the backward utility as a modified version of it, called $\upsilon(r, u, r^{(0)})$, as follows:

$$\upsilon(r, u, r^{(0)}) = \begin{cases} u(r) - Lu^{(0)}(u^{(0)} - u(r)) & r \leq r^{(0)} \\ u(r) & r > r^{(0)}, \end{cases} \tag{8}$$

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

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

[2] 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.

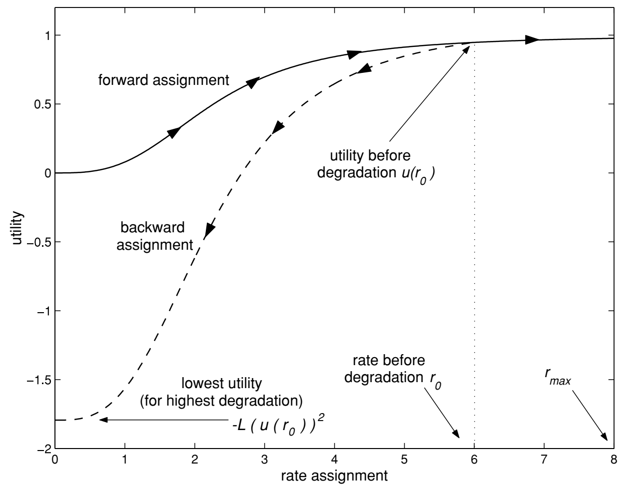

Fig. 1 Forward and backward utility for $L = 2$

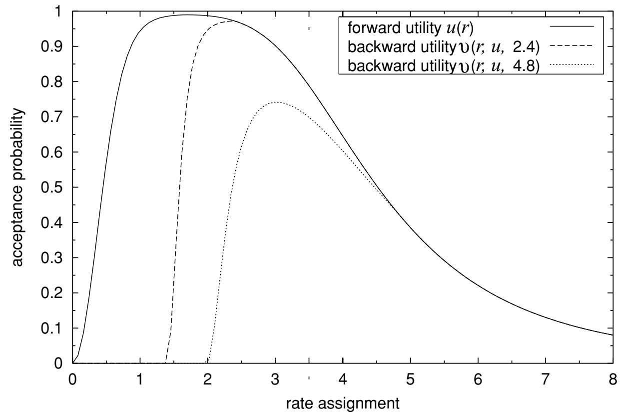

Fig. 2 Forward and backward acceptance prob. for $L = 2$

evaluation. If $L > 0$ we have a *fragile* QoS, i.e., the utility is lower if the current assignment results from a degradation. For an infinitely fragile QoS ($L = \infty$), the utility decreases suddenly to 0 if any degradation occurs, no matter how small. Other values of L obtain intermediate behaviors. At this point, we also need to slightly modify the definition of $A(u, p)$ given by (1), by considering the case in which $u(\cdot)$ is replaced by $\upsilon(\cdot)$, which can assume negative values. Thus, $A(\upsilon, p) = 0$ if $\upsilon < 0$, whereas we use the same definition (and henceforth we still call the function A) when $\upsilon \geq 0$.

Figure 1 reports the backward utility for $\psi = 1$ and $L = 2$ when the rate r is allocated between 0 and $r_{\max} = 8$ units, and the utility is sigmoid-shaped. Figure 2 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., $A(\upsilon(r, u, r^{(0)}), p(r)) \leq A(u(r), p(r))$ 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, remember that when r decreases the price is in general lower. For example, if the assigned rate goes from $r^{(0)} = 4.8$ to $r^{(1)} = 3.2$, Fig. 2 shows that $A(\upsilon(r^{(1)}, u, r^{(0)}), p(r^{(1)}))$ is still considered an improvement with respect to $A(u^{(0)}, p(r^{(0)}))$. However, this happens only since the assignment $r^{(0)} = 4.8$ 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 $L = 2$). Having modeled the (amplified) users' reaction to dynamic assignment in addition to QoS and pricing *per se*, it is possible to compare different policies of dynamic resource management, as we do in the next section.

4 Dynamic RRM strategies and economic admission control

Consider a network where users are characterized by backward utilities when rate degradation occurs. In this scenario, we identify two main classes [1] 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 [9]. 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 [10, 13, 14].

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 it is not possible to re-negotiate the assignment even when this would be beneficial, e.g., when users with low utility or low contribution to the revenue have already been allocated. The BE management instead provides only partial QoS guarantees. On average, we expect this latter policy to allow admission of a larger number of users, though their quality is possibly subject to degradation.

A more precise comparison between these two strategies can be obtained by considering the performance metrics previously defined, e.g., as per (2). The Best Effort RRM is supposed to try to allocate as many users as possible, by exploiting the system soft capacity, which indeed allows to admit new users even when the capacity limit of the system seems to be reached. From a *naïve* point of view, one might erroneously think that the Best Effort RRM would have the advantage of always allowing improvement of the metrics, at least theoretically. This should happen since R, and any similar metric, seem to be non decreasing if N is increased (i.e., more users are considered). Indeed, we must consider that, due to the quality decrease of already admitted users in case i is admitted, there is an increase in these metrics due to user i's term but also a possible decrease for some other users' terms. This depends on how the rate of a generic user j already in the system is decreased as a result of user i's admission 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. Moreover, before the admission of the ith user, we can evaluate, e.g., $S^{(i,0)}$ by simply counting the users allocated before i which have accepted the service conditions. On average, we expect to have $S^{(i,0)} = \sum_{j=1}^{i-1} A(u_j(r_j^{(0)}), p(r_j^{(0)}))$. At this point, it is possible to write:

$$S^{(i,1)} = \sum_{j=1}^{i-1} A(\upsilon_j(r_j^{(1)}, u_j, r_j^{(0)}), p(r_j^{(1)})) + A(u_i(r_i^{(1)}), p(r_j^{(1)})) , \tag{9}$$

which is the total satisfaction (or also the average number of admitted users) 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 $r_j^{(0)} = r_j^{(1)} = 0$.

Similar equations can be written for the metric R. The evaluation after the admission is made 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 for the already admitted users. Equation (9) also indicates that the performance of the BE RRM depends on how the rates are translated from $\mathbf{r}^{(0)}$ to $\mathbf{r}^{(1)}$. When deciding about the actual strategy in which this degradation is performed, as we aim at possibly keeping all the users in the system even after the admission of the new call, we should try not to penalize some users more than others. Moreover, we should avoid to decrease some allocations more than what is necessary, which can be done, e.g., by adopting an iterative approach which decreases the allocation in small steps until a feasible allocation is found. 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 average number is $S^{(i,0)}$, in steps of $1\,\mathrm{dB}/S^{(i,0)}$, until the vector of the powers is feasible. This means that we choose to apply the rationale above to the allocated powers, so as to better account for Power Control issues [18], and the decrease amount is chosen to obtain a good trade-off between a reasonably fast convergence of the degradation procedure and its accuracy. Note however that results produced with other policies still respecting the aforementioned criteria, i.e., fair and small degradations, are still in very good qualitative agreement with the ones shown in the following.

One might argue that the variation in the acceptance probability implies that users may completely refuse the service, considered unsatisfactory due to degradation. Hence, to have a full evaluation of the performance of $\mathbf{r}^{(1)}$, the resource left by the first user who quits the service should 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 user is forced to leave the system without any iteration. This is a conservative approach as it considers the worst case in which all dissatisfied users simultaneously leave the system after the degradation. Note that this only means that no negotiation is considered in order to alleviate the dissatisfaction of the users. However, the resource that dissatisfied users leave is made available to the network manager and can of course be allocated to new incoming users.

Finally, this procedure can be iteratively repeated for every new 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 $\upsilon_j(r, u_j, r_j^{(1)})$ but rather by $\upsilon_j(r, \upsilon_j^{(0)}, r_j^{(1)})$, where $\upsilon_j^{(0)} = \upsilon_j^{(0)}(r, u_j, r_j^{(0)})$.

The BE approach can be further improved by considering a strategy which blocks all users whose admission is estimated to have negative impact on the system. As we are considering users arriving one at a time, it is possible to define a general Admission Control as follows. Assume that we are looking at the admission for the ith user. This user is admitted if:

$$R^{(i,1)} > R^{(i,0)} \tag{10}$$

where $R^{(i,0)}$ and $R^{(i,1)}$ are the values of the revenue before and after the admission of the ith user, respectively. The choice of R among the possible performance metrics is made without loss of generality, as any other metric could be considered instead. Note that $R^{(i,0)}$ can be known exactly, as it depends on the current user behavior, whereas in some cases $R^{(i,1)}$ needs to be estimated as the future behavior of the users following a potential admission is known only statistically.

The online evaluation of (10) can be regarded, by abstracting from the choice of R, as a framework to perform Admission Control, that we will call in the following *Economic*

Admission Control. In light of the above discussion, we are now able to describe the loss due to the quality decrease of already admitted users in case user i is admitted. This value depends on how the rate of a generic user j already in the system must be decreased in order to admit user i into the system. With the same notation used before, a translation of the allocation vector from $\mathbf{r}^{(0)}$ before the admission of user i to $\mathbf{r}^{(1)}$ implies that the new value of the revenue is estimated as:

$$R^{(i,1)} = \sum_{j=1}^{i-1} p\big(r_j^{(1)}\big) A\big(\upsilon_j(r_j^{(1)}, u_j, r_j^{(0)}), p\big(r_j^{(1)}\big)\big) \\ + p\big(r_i^{(1)}\big) A\big(u_i\big(r_i^{(1)}\big), p\big(r_j^{(1)}\big)\big) \; . \qquad (11)$$

The above equation applies to the revenue the same rationale as (9), so that the revenue after the admission is evaluated by considering the backward utilities υ_j instead of u_j for those users that have a dynamically decreased allocation.

If we use the already mentioned mechanism of decreasing the SIR of the $S^{(i,0)}$ connected users in steps of 1 dB$/S^{(i,0)}$ when the admission of the ith user would make the SIR target vector no longer feasible, the Economic Admission Control framework is completely specified by Eqs. (10) and (11). Thus, we can compare it with the two basic GP and BE techniques.

5 Results

We now present comparative results for GP and BE RRM, which show the performance achievable by such strategies. We perform simulations in a cellular wireless system, with N users placed with uniform spatial distribution over an area subdivided into hexagonal cells, 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. In this scenario, we evaluate the RRM performance over a large number of instances, repeating the allocation in the same network instance for every RRM strategy.

Table 1 List of parameters of the simulation scenario

Parameter (symbol)	Value
Number of cells	19
Bandwidth (B)	20 rate units
Max assignable rate ($r_{\max}$)	8 rate units
Max terminal power ($P_{\max}$)	2 W
Cell radius (d)	500 m
Gain at 1 m	−28 dB
Path loss exponent (α)	3.5
Shadowing parameter (σ)	8 dB
Mean SNR at cell border	20 dB

Table 2 List of micro-economic parameters

Parameter	Value
Number of users (N)	160
Utility parameter ζ (curvature)	uniform in [2, 20]
Utility parameter x_s (middlepoint)	uniform in [0.1, 0.9]
Normalization constant k	$-\log 0.9$
Normalization utility ψ	1.0
Normalization price ϕ	1.0
Utility sensitivity μ	2.0
Price sensitivity ϵ	4.0

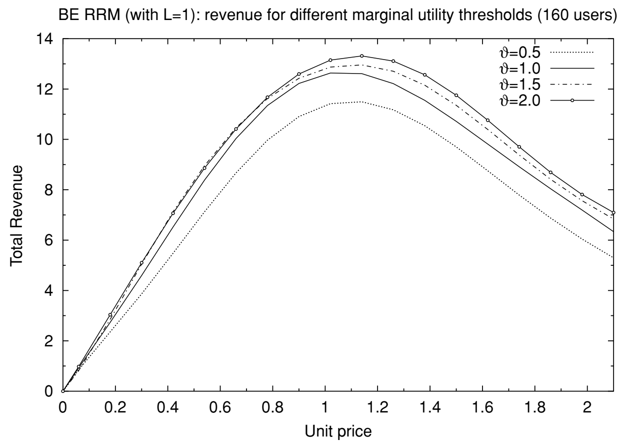

Fig. 3 BE management: revenue R for linear price $p(r) = \xi r$ as a function of ξ

First of all, the revenue (metric R) is considered in Figs. 3–5. Figure 3 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 1. The same analysis is carried out in Fig. 4, 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, for this particular case the larger the allocation threshold, the higher the revenue. However, this conclusion depends on the system parameters, and can not be generalized. In general the best threshold setup depends on many factors, e.g., the number of potential users in the network and the shape of their utilities. A detailed study of these relationships, though a worthwhile effort, is out of the scope of this paper and is left for future research. Figure 5 compares the performance of GP and BE RRM (the latter by considering different values of L, whereas the former does not depend on L). In particular, as will be done in the following for the other metrics, we compare the two approaches for the case $\vartheta = 1.0$ (different values of ϑ present entirely similar results). 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

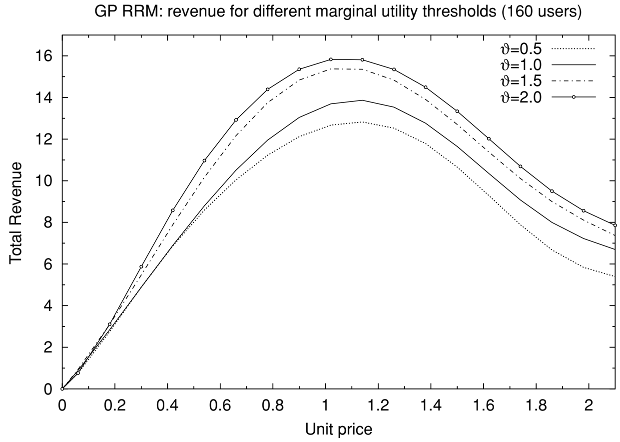

Fig. 4 GP management: revenue R for linear price $p(r) = \xi r$ as a function of ξ

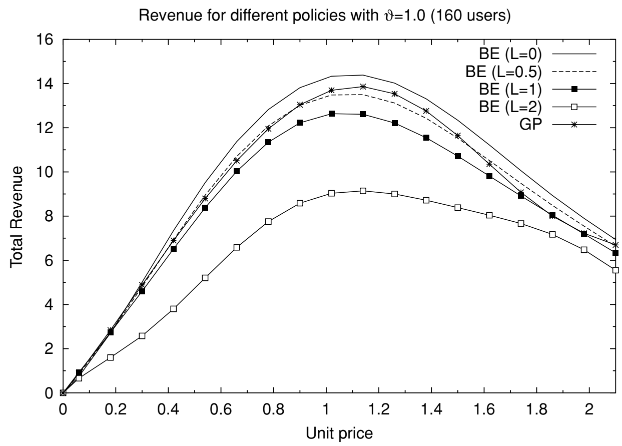

Fig. 5 Revenue R for linear price $p(r) = \xi r$: comparison of different RRM approaches

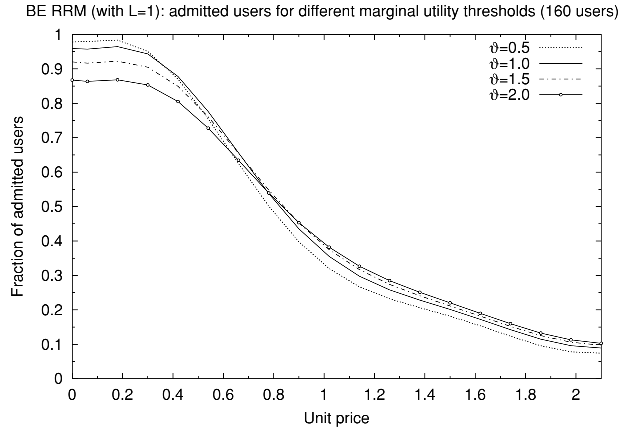

Fig. 6 BE management: admission rate S/N for linear price $p(r) = \xi r$ as a function of ξ

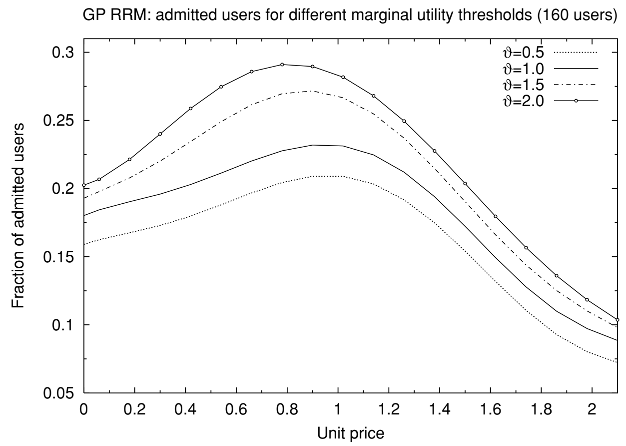

Fig. 7 GP management: admission rate S/N for linear price $p(r) = \xi r$ as a function of ξ

increases. So, the BE RRM performs better only when users are not overly sensitive to dynamic quality degradations.

Secondly, we investigate in Figs. 6 and 7 the average 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 with guaranteed QoS. It can also be said that the two allocation strategies lead to quantitatively *comparable* revenue, as shown in Fig. 5, and in particular the exact value for BE very much depends on L, but the revenue is originated in totally different ways. As shown in Figs. 6 and 7, in the GP case few users are admitted, which therefore pay a considerable amount, in return for performance guarantees. Instead, BE builds the revenue by admitting more users who pay a lower price. These conclusions are summarized in Fig. 8, 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.

To evaluate the performance of Economic Admission Control applied to BE techniques, we show again results by varying ϑ, and also comparisons with the other managements (GP and plain BE). All the simulations have been performed in the same multi-cell scenario by always considering $L = 1$. Figures 9–10 show the evaluation metrics for a BE system with EAC algorithm based on the revenue, as in (10). In more detail, Fig. 9 presents the revenue earned by the BE management with EAC. It is interesting to observe that the behavior is different from Figs. 3 and 4, where, in the range of interest, the larger ϑ the better. Here, variation of ϑ changes the curve but the peak performance is similar, so that

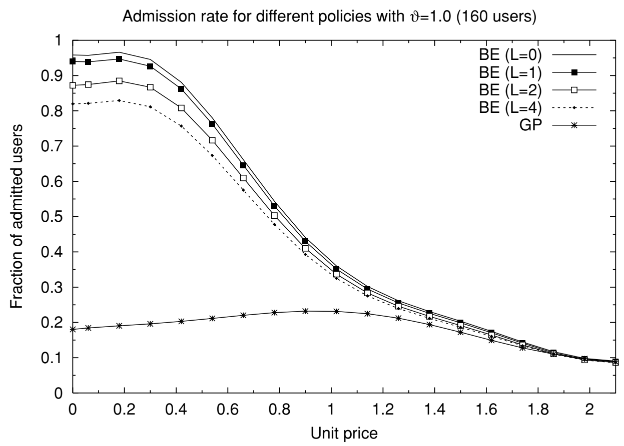

Fig. 8 Admission rate S/N for linear price $p(r) = \xi r$: comparison of different RRM approaches

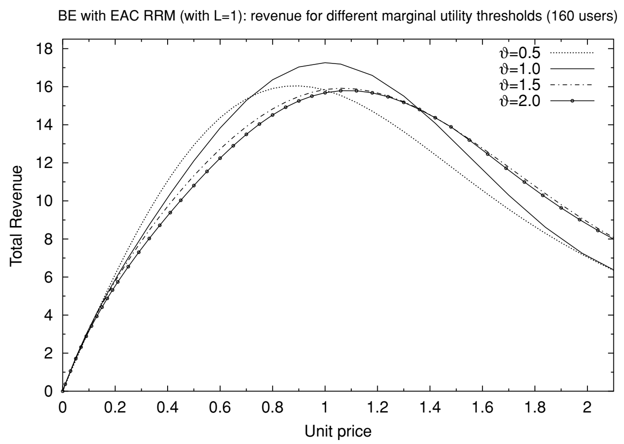

Fig. 9 BE management with EAC: revenue for linear price $p(r) = \xi r$ as a function of ξ

the best value (reached for $\vartheta = 1.0$) is approached anyway. This is reasonable, since the goal of EAC is to increase cooperation among the users. For the previous cases of GP and BE without EAC, the only way to force cooperation is to assign less resource to the users (by choosing a higher ϑ, which is shown to be preferable in Figs. 3 and 4). However, this is limiting for the users, which are bound to experience lower quality, and also it introduces some sensitivity to the parameter values. The EAC approach instead makes the system more robust, and also the maximum achievable revenue less sensitive to the management setting (however, there is still a degree of freedom in the choice of ϑ for what concerns the optimal price variations). Figure 10 presents an oscillatory behavior when the price is changed. The reason for this is in the superposition of two different admission control mechanisms, i.e., EAC and implicit admission control given by the price setup [2, 8]. Increasing the price decreases the number of admitted users at first, but, as long as the price does not increase too much, this is compensated by EAC which tries to admit more users by performing resource sharing in an efficient manner, i.e., by always improving the revenue.

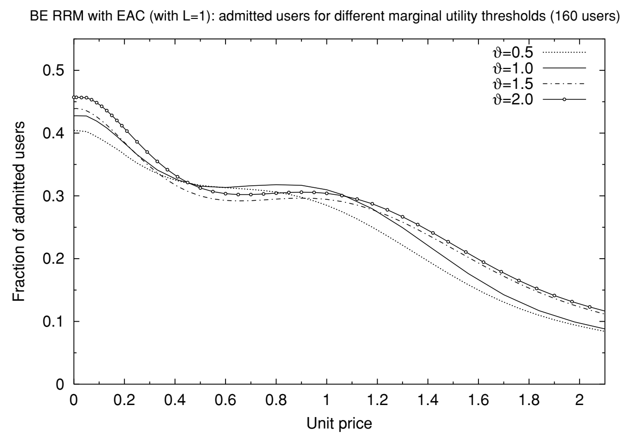

Fig. 10 BE management with EAC: admission rate for linear price $p(r) = \xi r$ as a function of ξ

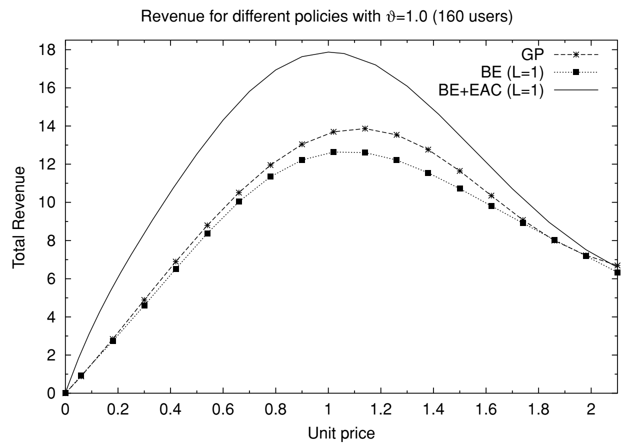

Fig. 11 Comparison of the total revenue obtained with $\vartheta = 1$

This compensation results in the almost constant number of admitted users as the price increases. When the price is very high, the EAC capabilities are overtaken by the mechanism of dissatisfied users leaving the system, but in this case a strong decrease of the number of admitted users can also be observed for the plain BE case, see Fig. 6, which clearly gives an upper bound for the number of admitted users with respect to the case with EAC.

Finally, Figs. 11–12 compare the three approaches for the case $\vartheta = 1.0$ (different values of ϑ present entirely similar results). EAC is shown to be able to significantly improve the total revenue (analogous results also hold for other metrics). The admission rate is lower than in the simple BE management, but this trivially follows from the goal of BE to obtain an admission rate as high as possible. BE and BE + EAC tend also to coincide when the price increases. In any case, the number of admitted users for BE + EAC is significantly higher than for the GP policy. For what concerns the comparison with BE instead, also note that large admission rates are inefficient, since they decrease the total revenue, which is caused by the admission of users with poor QoS that do not pay high tariffs and cause more

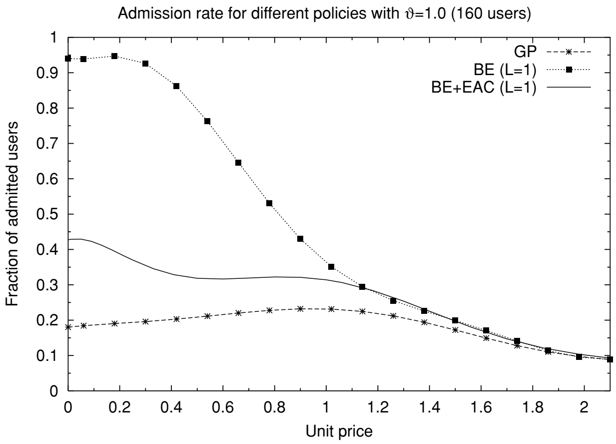

Fig. 12 Comparison of the admission rates obtained with $\vartheta = 1$

interference to other users. For this reason, Economic Admission Control offers a very interesting possibility of improving network management. Importantly, the results shown here represent only one among the several possibilities to perform Admission Control based on micro-economic criteria. Besides this simple revenue-based approach, other possibilities are still to be studied and can be investigated by future research.

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 strategies. 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 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.

A possible strategy can be to conservatively admit users with a smarter criterion, such as the Economic Admission Control, which tries to predict the impact of the admission on the system. In this way the system is still Best Effort, since there is no guarantee on the achievable performance. However, our results prove that this methodology allows to combine the advantages of the Best Effort and of the Guaranteed Performance techniques. In other words, it allows to improve the performance of the BE management by increasing reliability, while still being able to satisfy a larger number of users. As a consequence, the revenue is significantly increased with respect to both GP and BE approaches.

References

1. E. Altman, A. Orda, and N. Shimkin, Bandwidth allocation for guaranteed versus best effort service categories, in: *Proceedings INFOCOM*, (1998) vol. 2 pp. 617–624.
2. L. Badia, M. Lindström, J. Zander, and M. Zorzi, Demand and pricing effects on the radio resource allocation of multimedia communication systems, in: *Proceedings IEEE Globecom*, (2003) vol. 7 pp. 4116–4121.
3. L. Badia and M. Zorzi, Radio resource management with utility and pricing for wireless LAN hot-spots, Kluwer Wireless Personal Communication Journal 34(1–2) (July 2005) 127–142.
4. L. Badia, M. Zorzi, and A. Gazzini, A model for threshold comparison call admission control in third generation cellular systems, in: *Proceedings ICC*, (2003) vol. 3 pp. 1664–1668.
5. 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.
6. D. Famolari, N. Mandayam, D. Goodman, and V. Shah, *A New Framework for Power Control in Wireless Data Networks: Games, Utility and Pricing*, (Kluwer Academic Publishers, 1999) chapter 1 pp. 289–310.
7. R.J. Gibbens and F.P. Kelly, Resource pricing and the evolution of congestion control, Automatica 35 (1999) 1969–1985.
8. J. Hou, J. Yang, and S. Papavassiliou, Integration of pricing with call admission control to meet QoS requirements in cellular networks, IEEE Transactions on Parallel and Distributed Systems 13(9) (Sept 2002) 898–910.
9. 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.
10. F. Kelly, A. Maulloo, and D. Tan, Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 49 (1998) 237–252.
11. H. Lin, M. Chatterjee, S.K. Das, and K. Basu, ARC: An integrated admission and rate control framework for competitive wireless CDMA data networks using non-cooperative games, IEEE Transactions on Mobile Computing 4(3) (2005) 243–258.
12. C. Saraydar, N. Mandayam, and D.J. Goodman, Efficient power control via pricing in wireless data networks, IEEE Transactions on Communications 50(2) (2002) 291–303.
13. S. Shenker, D.D. Clark, D. Estrin, and S. Herzog, Pricing in computer networks: Reshaping the research agenda, ACM Computer Communication Review 26 (1996) 19–43.
14. V. Siris, Resource control for elastic traffic in CDMA networks, in: *Proceedings ACM MobiCom* (2002) pp. 193–204.

15. R.M. Spiegel, *Theory and Problems of Statistics* (McGraw-Hill, New York, 1996).
16. A.M. Viterbi and A.J. Viterbi, Erlang capacity of a power controlled CDMA system, IEEE Journal on Selected Areas in Communications 11(6) (1993) 892–900.
17. 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.
18. J. Zander and S.-L. Kim, *Radio Resource Management for Wireless Networks* (Artech House, Norwood MA, 2001).
19. 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 of the International Conference on Computer Communications and Networks* (2002) pp. 485–488.

Leonardo Badia received a Laurea degree (with honors) in electrical engineering and a Ph.D. in information engineering from the University of Ferrara, Italy, in 2000 and 2004, respectively. He was a Research Fellow at the University of Ferrara from 2001 to 2006. During these years, he also had collaborations with the University of Padova, Italy, and Wireless@KTH, Royal Institute of Technology, Stockholm, Sweden. In 2006, he joined the "Institutions Markets Technologies" (IMT) Institute for Advanced Studies, Lucca, Italy, where he is currently a Research Fellow. His research interests include wireless ad hoc and mesh networks, analysis of transmission protocols, optimization tools and economic models applied to radio resource management.

Michele Zorzi received a Laurea degree and a Ph.D. in electrical engineering from the University of Padova in 1990 and 1994, respectively. During academic year 1992–1993, he was on leave at UCSD, attending graduate courses and doing research on multiple access in mobile radio networks. In 1993 he joined the faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. After spending three years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, where he became a professor in 2000. Since November 2003 he has been on the faculty at the Information Engineering Department of the University of Padova. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, energy constrained communications protocols, and broadband wireless access. He was Editor-In-Chief of IEEE Wireless Communications, 2003–2005, and currently serves on the Editorial Boards of IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, Wiley's Journal of Wireless Communications and Mobile Computing, and ACM/URSI/Kluwer Journal of Wireless Networks, and on the Steering Committee of the IEEE Transactions on Mobile Computing. He has also been a Guest Editor of special issues in IEEE Personal Communications (Energy Management in Personal Communications Systems) and IEEE Journal on Selected Areas in Communications (Multimedia Network Radios).