Integer Linear Programming Optimization of Joint RRM Policies for Heterogeneous Wireless Systems

M.C. Lucas-Estañ[a,*], J. Gozalvez[a] and J. Sánchez-Soriano[b]

[a] Ubiquitous Wireless Communications Research (Uwicore) Laboratory
[b] The Operations Research Center
University Miguel Hernandez of Elche, Avda. de la Universidad s/n, 03202, Elche, Spain.

Abstract—**Wireless systems will be characterized by the coexistence of heterogeneous Radio Access Technologies (RATs) with different, but also complementary, performance and technical characteristics. These heterogeneous wireless networks will provide network operators the possibility to efficiently and coordinately use the heterogeneous radio resources, for which novel Joint Radio Resource Management (JRRM) policies need to be designed. In this context, this work proposes and evaluates a JRRM policy that simultaneously determines for each user an adequate combination of RAT and number of radio resources within such RAT to guarantee the user/service QoS requirements, and efficiently distribute the radio resources considering a user fairness approach aimed at maximizing the system capacity. To this aim, the JRRM algorithm, which takes into account the discrete nature of radio resources, is based on integer linear programming optimization mechanisms.**

Index Terms—**Joint Radio Resource Management, heterogeneous wireless systems, integer linear programming, multimedia.**

1. INTRODUCTION

The continuous evolution of wireless and mobile technologies is creating wireless communication ecosystems where several Radio Access Technologies (RATs) physically coexist. In addition, new user applications with diverse and more restrictive Quality of Service (QoS) requirements are emerging, raising the challenge of carrying out a more efficient use of the scarce available radio resources. In this context, there is a wide consensus in the research community that future heterogeneous wireless networks will require the coordinated management of the radio resources from the coexisting RATs. Such management will be conducted through Joint Radio Resource Management (JRRM) techniques, also

* Corresponding author. Tel.:+34 965222031, fax: +34 966658903
E-mail address: m.lucas@umh.es (M. C. Lucas-Estañ)

referred to as Common Radio Resource Management (CRRM) techniques, aimed at efficiently distributing the available heterogeneous radio resources in order to satisfy the user QoS demands, while increasing the operators' system revenue. The JRRM concept was defined by the 3GPP (3rd Generation Partnership Project) in [1] and [2], where also different supporting network architectures were presented to ensure the interoperability between the different RATs.

JRRM policies are responsible for assigning incoming calls the optimum RAT over which to convey them. Several studies can be found in the literature investigating proposals to address the RAT selection dilemma. [3] proposes some initial JRRM RAT selection mechanisms based on pre-established service-to-RAT assignments, and a load balancing criterion (the user is assigned to the lowest loaded RAT). Another interesting initial RAT selection strategy [4] defines a suitability factor for selecting each available RAT based on aspects such as the radio resources availability, the interference level, and the terminal and network capabilities. An innovative contribution is the evaluation of the suitability factor's dependency with varying operators' policies. Several RAT selection principles based on the signal strength (coverage) and instantaneous load are also suggested in [5]. Other initial RAT selection policies base their JRRM decisions on each RAT's load and the multi-technology capability of user terminals. In fact, it is possible to take advantage of the presence of multi-mode terminals by switching their assigned RATs in order to free the capacity required to accept new calls from single-mode terminals. Several strategies to perform this traffic rearrangement in a heterogeneous wireless framework are discussed in [6] and references therein. The load balancing strategies employed in [7] are aimed at achieving a uniform traffic distribution, which is pointed out to be desirable in order to maximize the trunking gain and minimize the probability of making unnecessary vertical handovers of multi-technology terminals. For non real-time services, the load balancing is performed based on the measured buffer delay, while for real-time services a load balancing principle based on load thresholds is proposed. In [8], the authors evaluate a vertical handover strategy to simultaneously decide the target network for several users requesting a handover in the same base station. The vertical handover algorithm will decide the best target network for each user based on its QoS requirements and individual interests, but would also take into account the networks' load resulting from the handover process. Consequently, the

algorithm improves its RAT selection decision when simultaneously receiving a high number of handover petitions. Finally, a centric RAT selection algorithm is proposed in [9]. The proposed approach selects the RAT providing each user the highest utility defined in terms of cost, power consumption, achievable data rate, network load and link quality. [9] also demonstrates the need and benefits of exploiting context-aware preferences in the selection process.

In addition to decide the RAT over which each incoming call will be conveyed, certain JRRM policies have recently proposed to also determine the resources needed by each user to satisfy its QoS demands. An example is the work reported in [10], which proposes a Joint Call Admission Control (JCAC) algorithm based on the arrival rate of each class of calls. This algorithm simultaneously addresses the RAT selection and Call Admission Control (CAC) dilemmas, with the aim of reducing the call blocking and dropping probabilities, and ensuring fairness in the allocation of radio resources. Other interesting JRRM proposals jointly addressing the RAT selection and intra-RAT Radio Resource Management (RRM) dilemmas have been reported in [11] and [12]. These proposals are based on neural networks, and simultaneously determine the most appropriate RAT and bit rate allocation. In [11], the proposed JRRM algorithm is based on neural networks and fuzzy logic, and considers factors such as the signal strength, resource availability and mobile speed. However, this JRRM algorithm does not incorporate mechanisms to take into account the diverse QoS requirements of different service types. The JRRM proposal reported in [12] bases its decision criterion on user QoS demands in terms of required bit rate or maximum delay, and employs Hopfield neural networks to find its JRRM solutions. This algorithm simultaneously decides the optimum RAT for the incoming call, and the necessary radio resources at the assigned RAT. However, the diverse nature and characteristics of radio resources in heterogeneous environments is not actually considered in [12]. Since such diversity can considerably impact the capacity of the selected RAT to satisfy the user QoS demands, and therefore the optimum JRRM solutions to meet such demands, this work extends the current JRRM state of the art by proposing a JRRM algorithm that based on the radio resources' diversity, simultaneously determines for each user an adequate combination of RAT and number of discrete radio resources within such RAT to guarantee its QoS requirements while trying to maximize the system's capacity in terms of users satisfactorily served. In addition, the proposed

JRRM algorithm also tackles the CAC dilemma given that the JRRM solution determines the most suitable RAT with available radio resources to support the new incoming call considering the current system conditions (load and QoS requirements). The proposed JRRM algorithm implements a user fairness policy, and aims to equally satisfy all users in the system. To this aim, the proposal is based on integer linear programming and optimization techniques.

The rest of the paper is organized as follows. Section 2 presents the integer linear programming based JRRM algorithm proposed in this work. Section 3 evaluates its performance, and proposes additional variants to further optimize its operation and the resulting user perceived QoS. The implementation and computational cost of the proposed algorithm is evaluated in Section 4, while Section 5 summarizes the contributions from this work and draws final conclusions.

2. INTEGER LINEAR PROGRAMMING JRRM POLICY

This paper considers a multimedia scenario where email (background), web (interactive), and real-time video (with different mean bit rates) services coexist, and require different number of radio resources to achieve equal QoS levels. In this context, the proposed JRRM policy tries to exploit this QoS/resource flexibility to provide all users present in a multimedia environment with the highest possible homogeneous user satisfaction levels. The study is conducted considering a heterogeneous wireless ecosystem where the GPRS (General Packet Radio Service), EDGE (Enhanced Data rates for GSM Evolution), and HSDPA (High Speed Downlink Packet Access) RATs physically coexist, and provide the same radio coverage from a multi-RAT base station. A radio resource is equivalent to a timeslot for GPRS and EDGE, and a code for HSDPA.

2.1. Traffic Class Utility Functions

This work is based on utility functions that try to characterize the QoS satisfaction level experienced by a user based on the requested traffic service and the radio resources it has been assigned (combination of RAT and number of radio resources assigned within that RAT). This is a challenging task because user satisfaction is a subjective concept that heavily depends on user perceptions. The defined utility functions try to express the perceived user QoS as the link quality, and therefore data rate, varies. To establish the utility functions, the minimum, mean, and

maximum QoS levels demanded by users are first defined per service class as illustrated in Fig. 1. This work considers a multimedia traffic scenario with email (background), web (interactive) and real-time H.263 video (with different mean bit rates) users.

For web and email services, utility values are expressed in terms of the user throughput. The minimum, mean and maximum QoS levels for web users have been defined as the throughput needed to satisfactorily transmit 90%, 95% and 97.5% of web pages in less than 4 seconds as established by the 3GPP TS 22.105 recommendations [13]. These high percentiles have been selected due the high transmission reliability requirements of non-real time data services. Web traffic is here modeled using the work reported in [14], whereas [15] has been used to model email traffic. The email model considers the transmission of emails with and without attachments, which makes it difficult to successfully transmit emails with large attachments within the 4 seconds 3GPP recommendations. Consequently, the email QoS thresholds have been established based on the throughput required to satisfactorily transmit 65%, 75% and 80% of the emails (with or without attachments). Once the QoS satisfaction thresholds have been established for web and email services, the utility functions have been defined so that users perceive a null utility value if their minimum QoS demand is not satisfied. This condition avoids assigning radio resources to users who would experience very poor QoS levels. Web and email user satisfaction linearly grows with the experienced throughput between the minimum and maximum QoS thresholds. Utility values equal to one have been
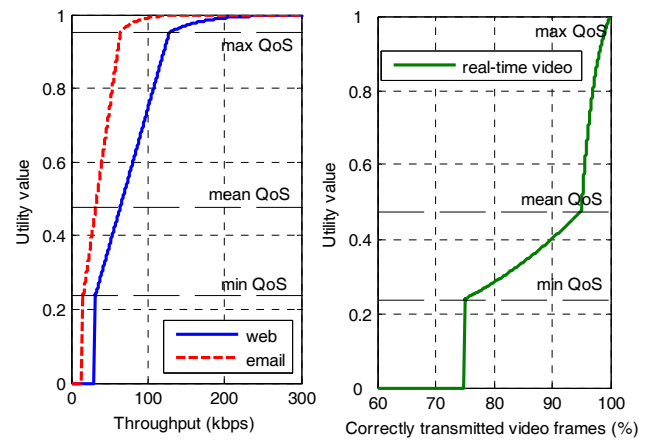

Fig. 1. Utility functions per traffic service.

avoided for web and email transmissions to account for the transmission reliability requirements of these services, and the dependence of the achievable throughput levels on the experienced channel quality conditions.

For real-time video services, video frames are considered to be satisfactorily transmitted if they are transmitted before the next video frame is to be transmitted. Consequently, the utility functions for real-time video services have been defined based on the percentage of correctly transmitted video frames, and the real-time video utility functions are independent of the mean video bit rates. The real-time video QoS satisfaction thresholds have been established considering the H.263 traffic model described in [16] and the indications provided in [17]. The studies reported in [17] show that a 25%, or even higher, dropping rate does not have a catastrophic effect on the QoS perceived by H.263 video users, and that dropping rates as high as 5% can be overcome if appropriate transmission techniques are invoked. Based on these results, the minimum and mean QoS satisfaction levels correspond to guaranteeing that 75% and 95% of video frames are transmitted before the next video frame needs to be transmitted. The maximum utility value for real-time video users has been set equal to one, and is achieved when all video frames are transmitted before the next video frame is to be transmitted. Although the 5% difference between the mean and maximum QoS levels might look negligible, this 5% includes the H.263 I-frames. These frames include information of independently coded images in a video sequence, and are also used to code/decode other images exploiting temporal redundancy. As a result, I-frames have a significant impact on the user perceived QoS level, and require high transmission rates due to their potential large size. Similarly to web and email users, real-time video users also perceive a null utility value below the minimum QoS threshold. Following the indications in [17] that highlight that an acceptable video quality requires a high percentage of correctly received video frames, the video utility increases slowly with the percentage of transmitted frames until the mean QoS level is achieved, and then rapidly until the maximum QoS level.

Once the utility functions are established, it is then necessary to relate the utility values with the different radio resource assignments. To establish this relation, the throughput achieved by each RAT and number of radio resources combination must be considered. However, it is difficult to estimate the throughput that could be achieved with a given number of radio resources given that the simulated radio access technologies implement link adaptation schemes. These schemes dynamically vary the used transmission mode (i.e., modulation and coding scheme) based on the experienced channel quality conditions. To account for these variations, and considering the difficulty to predict the achievable throughput in adaptive radio interfaces, the relation between the utility values and radio resource assignments has been established considering the data rate of the transmission modes providing a balance between high data rates and high error correction capabilities. In this context, average throughput values of 13.4 kbps and 22.4 kbps per timeslot (TS) have been selected for GPRS and EDGE, respectively, corresponding to the data rates of the coding scheme 2 (CS2) in GPRS, and the modulation and coding scheme 5 (MCS5) in EDGE [18]. In HSDPA, a high number of transmission modes are defined depending on the number of assigned codes. This work considers the transmission modes related to the 30 CQI (Channel Quality Indicator) values for User Equipment category 10 [19]. To achieve the sought balance between high data rate and high error correction capabilities, the selected transmission rate per number of assigned HSDPA codes corresponds to that achieved by the 'intermediate' transmission mode out of all possible modes for a given number of codes. The resulting data rates for each possible radio resource assignment are shown in Table I. It is then possible to directly relate for web and email services the utility values that would correspond to each radio resources assignment. For example, if we consider a radio resource assignment corresponding to two EDGE timeslots, this assignment can achieve an average throughput of 44.8 kbps. Considering the utility functions defined in Fig. 1, such throughput would then correspond to a 0.33 utility value for web users, and 0.67 for email users.

For real-time H.263 video services, an additional step is necessary. A cumulative distribution function (CDF) of the throughput needed to transmit each video frame before the next video frame is to be transmitted is derived following the implemented H.263 video model [16]. Through these CDFs, the percentage of video frames reported in Fig. 1 can be related to the corresponding necessary throughputs for the various video bit rates considered in this work. Once the utility values are expressed as a function of the throughput, the utility values can be related to radio resources using the previously discussed relation between throughput and radio resources. Table II shows an

TABLE I
DATA RATE PER RADIO RESOURCE ASSIGNMENT

Resources per RAT	Data rate (kbps)	Modulation
n GPRS TS	$n \times 13.4$	GMSK
n EDGE TS	$n \times 22.4$	8PSK
1 HSDPA codes	116.5	QPSK
2 HSDPA codes	396	QPSK
3 HSDPA codes	741	QPSK
4 HSDPA codes	1139.5	QPSK
5 HSDPA codes	2332	16-QAM
7 HSDPA codes	4859.5	16-QAM
8 HSDPA codes	5709	16-QAM
10 HSDPA codes	7205.5	16-QAM
12 HSDPA codes	8618.5	16-QAM
15 HSDPA codes	11685	16-QAM

TABLE II
UTILITY VALUES FOR 64KBPS VIDEO USERS

Res./ RAT	Throughput (kbps)	Utility value	Res./ RAT	Throughput (kbps)	Utility value	Res./ RAT	Throughput (kbps)	Utility value
1G	13.4	0.00	4E	89.6	0.29	3H	741	1.00
1E	22.4	0.00	7G	93.8	0.31	4H	1139.5	1.00
2G	26.8	0.00	8G	107.2	0.35	5H	2332	1.00
3G	40.2	0.00	5E	112	0.37	7H	4859.5	1.00
2E	44.8	0.00	1H	116.5	0.38	8H	5709	1.00
4G	53.6	0.00	6E	134.4	0.44	10H	7205.5	1.00
5G	67	0.00	7E	156.8	0.93	12H	8618.5	1.00
3E	67.2	0.00	8E	179.2	0.98	15H	11685	1.00
6G	80.4	0.00	2H	396	1.00			

example of the utility values obtained for the real-time 64 kbps H.263 video users with the different radio resource assignments; the utility values are listed according to the throughput provided by the corresponding RAT/radio resources combination. In this table, the assignments (RAT and number of radio resources) are denoted as xY, corresponding to x radio resources (timeslots or codes) from RAT Y (GPRS is represented as G, EDGE as E, and HSDPA as H). It is interesting to note that certain assignments cannot achieve utility values greater than zero.

2.2. JRRM Objective Function

The integer linear programming based JRRM policy proposed in this work has been designed with the objective of achieving an optimum distribution of radio resources in a heterogeneous and multimedia framework following a user fairness approach. In particular, the proposed algorithm seeks to provide similar, and highest possible, utility levels for all service classes, and only when the number of available radio resources is lower than the demand, will the implemented policy prioritize certain traffic classes. In this context, the JRRM proposal seeks to maximize the

lowest utility value assigned to a user in a joint radio resources distribution round. This JRRM algorithm is then referred to as MAXILOU (MAXImize Lowest Utility), and can be defined through the following objective function:

$$\max \quad \min_{j \in \{1,...,N\}} u_j \tag{1}$$

with N representing the number of active users requesting radio resources, and u_j being the utility value obtained by user j in the current radio resource distribution round:

$$u_j = \sum_{r=1}^{3} \sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}, \quad \text{where}$$

$$y_j^{r,s} \in \{0,1\} \quad \forall j \in \{1,...,N\}, \forall r \in \{1,2,3\}, \forall s \in \{1,...,c^r\} \tag{2}$$

In (2), $U_j(r,s)$ represents the utility value obtained by user j when assigned s radio resources (codes or timeslots) of RAT r (r is equal to 1, 2 or 3 for GPRS, EDGE and HSDPA respectively), and $s \in [1,c^r]$ with c^r corresponding to the maximum number of radio resources available at each RAT; the $U_j(r,s)$ values were established in Section 2.1. $y_j^{r,s}$ is a binary variable equal to one if user j is assigned s radio resources of RAT r, and equal to 0 if not.

In order to apply linear programming techniques to solve the established problem, (1) must be expressed as a linear equation. To this aim, a new real variable denoted as z, and equal to the smaller utility value assigned to a user has been defined, which results in the following JRRM objective function subject to the first JRRM problem constraint:

$$\max \quad z \tag{3}$$

$$s.t.:$$

$$z \leq u_j, \quad \forall j \in \{1,...,N\} \tag{4}$$

$$z \in \Re$$

2.3. System and Service Constraints

The aim of this work is to design a JRRM algorithm that provides an optimum radio resource distribution solution based on the current system conditions, users QoS requirements, and the established user fairness policy. In this context, the system and service constraints need to be also included in the problem statement and expressed as linear equations.

This work considers that users can only receive simultaneously resources from one RAT (of course, a user can be assigned different RATs for different sessions or calls). As a result, only one $y_j^{r,s}$ variable

can be equal to one for each user, which results in the following system constraint:

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} y_j^{r,s} \leq 1, \quad \forall j \in \{1,...,N\} \tag{5}$$

The JRRM policy must then decide for each user which $y_j^{r,s}$ variable is equal to one, considering that only $y_j^{r,s}$ variables achieving a utility value greater than zero are allowed, i.e. resources are not assigned to a user if they will not result in achieving at least the user minimum QoS level. This system constraint is expressed in (6), where K represents a positive constant value large enough:

$$y_j^{r,s} \leq K \cdot U_j(r,s),$$
$$\forall j \in \{1,...,N\}, \forall r \in \{1,2,3\}, \forall s \in \{1,...,c^r\} \tag{6}$$

Given the limited number of available radio resources in a system, which is expressed in (7), if a high system load cannot guarantee that all users are granted their minimum QoS requests, the problem would not have a satisfactory solution. When there are not enough resources to satisfy the minimum QoS level to all active users, the JRRM objective function will take the null value. In that case, whatever radio resource distribution with at least one user perceiving the zero utility value is an optimum solution to the integer linear programming problem, as for example, when all users do not receive resources, i.e. $\sum_{r=1}^{3}\sum_{s=1}^{c^r} y_j^{r,s} = 0, \quad \forall j \in \{1,...,N\}$. To avoid this situation, it has been imposed that one $y_j^{r,s}$ variable must be equal to one for each user, and the system constraint defined in (5) is then replaced by (8). When (8) is considered, the problem would be infeasible if there are not enough resources to satisfy the minimum QoS level to all active users in the system. In that case, some users should be eliminated from the radio resource distribution process following the service prioritization constraint that will be later discussed.

$$\sum_{j=1}^{N}\sum_{s=1}^{c^r} s \cdot y_j^{r,s} \leq c^r, \forall r \in \{1,2,3\} \tag{7}$$

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} y_j^{r,s} = 1, \quad \forall j \in \{1,...,N\} \tag{8}$$

The JRRM policy is here applied to distribute all the heterogeneous radio resources whenever a user requests access to the system, or a user ends its transmission and frees its radio resources (a different policy could have been defined without modifying the JRRM proposal). In this case, only real-time video active users who were assigned resources in the previous JRRM distribution round can maintain the minimum number of their assigned resources ($s_{min,j}$ radio resources from RAT $r_{min,j}$) that guarantees their established minimum QoS level. These real-time video users have to compete with the rest of users for other radio resource assignments improving their QoS satisfaction level. This condition can be expressed as follows:

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s} \geq U_j(r_{\min,j}, s_{\min,j})$$
$$\forall j \in \{1,...,N\} \mid t_j = 3\} \tag{9}$$

where t_j represents the traffic type demanded by user j (t_j is equal to 1, 2 or 3 for email, web and real-time video services respectively).

In scenarios where it is not possible to achieve equal utility values for all active users due to the scarcity of available resources and the users' high radio resources demands, users are served based on the following service priorities: real-time H.263 video (higher priority), web, and email. Among real-time video users, users demanding higher video bit rates are served first. The user priority criterion is represented by $\succ$ ($k \succ j$ indicates that user k is higher priority than user j). In this context, the relation between the utility value of a given user k and a lower priority user j is given by (10):

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} U_k(r,s) \cdot y_k^{r,s} \geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}$$
$$\text{if } k \succ j, \forall k, j \in \{1,...,N\} \text{ and } k \neq j \tag{10}$$

If the lowest priority user (j) is a video user who obtained radio resources in the previous JRRM distribution round, the condition in (9) has to be satisfied, and the real-time video user will obtain at least a utility value $U_j(r_{\min,j}, s_{\min,j})$ corresponding to the assignment of the $s_{min,j}$ radio resources from RAT $r_{min,j}$ needed to guarantee its minimum QoS level. However, (9) and (10) might not be simultaneously satisfied in some scenarios with radio resources shortage. If this is the case, the condition (9) is prioritized, and the lower priority user (j) will maintain the $s_{min,j}$ radio resources from RAT $r_{min,j}$ that are needed to guarantee its minimum QoS level. When such minimum level is achieved, the lowest priority user will not be assigned additional resources until the

highest priority user (k) surpasses its utility value ($U_j(r_{\min,j}, s_{\min,j})$). This situation is expressed as:

$$U_j(r_{\min,j}, s_{\min,j}) \geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s} \tag{11}$$

$$\text{if } u_k < U_j(r_{\min,j}, s_{\min,j}) \text{ and } k \succ j$$

If (r_a, s_a) is defined as the subset of the total possible radio resource assignments that verify $U_k(r,s) < U_j(r_{\min,j}, s_{\min,j})$, the following statement holds:

$$\sum_{r_a}\sum_{s_a} y_k^{r,s} = 1 \quad \text{if } u_k < U_j(r_{\min,j}, s_{\min,j}),$$

$$(r_a, s_a) = \{r \in \{1,2,3\}, \tag{12}$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) < U_j(r_{\min,j}, s_{\min,j})\}$$

Following the expression (12), (11) can also be expressed as:

$$\sum_{r_a}\sum_{s_a} U_j(r_{\min,j}, s_{\min,j}) \cdot y_k^{r,s} \geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}$$

$$\text{if } u_k < U_j(r_{\min,j}, s_{\min,j}) \text{ and } k \succ j,$$

$$(r_a, s_a) = \{r \in \{1,2,3\}, \tag{13}$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) < U_j(r_{\min,j}, s_{\min,j})\}$$

When there are enough radio resources to provide the higher priority user with a utility value higher than $U_j(r_{\min,j}, s_{\min,j})$, i.e., $u_k \geq U_j(r_{\min,j}, s_{\min,j})$, the prioritization condition (10) has to be satisfied. If (r_b, s_b) is defined as the subset of the total possible radio resource assignments verifying $U_k(r,s) \geq U_j(r_{\min,j}, s_{\min,j})$, the following expression is also satisfied:

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} U_k(r,s) \cdot y_k^{r,s} = \sum_{r_b}\sum_{s_b} U_k(r,s) \cdot y_k^{r,s}$$

$$\text{if } u_k \geq U_j(r_{\min,j}, s_{\min,j}),$$

$$(r_b, s_b) = \{r \in \{1,2,3\}, \tag{14}$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) \geq U_j(r_{\min,j}, s_{\min,j})\}$$

It is important to note that (r_a, s_a) and (r_b, s_b) are a partition of $\{r \in \{1,2,3\}, s \in \{1,...,c^r\}\}$, and verify the following expressions:

$$(r_a, s_a) \cup (r_b, s_b) = \{r \in \{1,2,3\}, s \in \{1,...,c^r\}\} \tag{15}$$

$$(r_a, s_a) \cap (r_b, s_b) = \varnothing \tag{16}$$

Following the expression (14), the prioritization condition in (10) can be expressed as:

$$\sum_{r_b}\sum_{s_b} U_k(r,s) \cdot y_k^{r,s} \geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}$$

$$\text{if } u_k \geq U_j(r_{\min,j}, s_{\min,j}) \text{ and } k \succ j, \tag{17}$$

$$(r_b, s_b) = \{r \in \{1,2,3\},$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) \geq U_j(r_{\min,j}, s_{\min,j})\}$$

The expressions (13) and (17) can then jointly be expressed as:

$$\sum_{r_a}\sum_{s_a} U_j(r_{\min,j}, s_{\min,j}) \cdot y_k^{r,s} + \sum_{r_b}\sum_{s_b} U_k(r,s) \cdot y_k^{r,s}$$

$$\geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}$$

$$\text{if } k \succ j, \forall k, j \in \{1,...,N\} \text{ and } k \neq j,$$

$$(r_a, s_a) = \{r \in \{1,2,3\}, \tag{18}$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) < U_j(r_{\min,j}, s_{\min,j})\},$$

$$(r_b, s_b) = \{r \in \{1,2,3\},$$

$$s \in \{1,...,c^r\} \,\big|\, U_k(r,s) \geq U_j(r_{\min,j}, s_{\min,j})\}$$

Condition (18) is only applied when the priority of user k is higher than that of user j. Finally, the JRRM solution is subject to the constraints (4), (6), (7), (8), (9), and (18). Following (18), if a solution cannot be obtained, and the linear objective function does not have a solution that satisfies all the services and system constraints, users with the lowest priority will be eliminated from the JRRM distribution round until the present users and their respective QoS demands allow for an integer linear programming JRRM solution to be achieved.

To summarise, the JRRM problem can then be modelled as follows:

$$\max \quad z$$

s.t.:

$$z \leq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}, \quad \forall j \in \{1,...,N\}$$

$$y_j^{r,s} \leq K \cdot U_j(r,s), \quad \forall j \in \{1,...,N\}, \forall r \in \{1,2,3\}, \forall s \in \{1,...,c^r\}$$

$$\sum_{j=1}^{N}\sum_{s=1}^{c^r} s \cdot y_j^{r,s} \leq c^r, \quad \forall r \in \{1,2,3\}$$

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} y_j^{r,s} = 1, \quad \forall j \in \{1,...,N\}$$

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s} \geq U_j(r_{\min,j}, s_{\min,j}) \quad \forall j \in \{1,...,N \mid t_j = 3\}$$

$$\sum_{r_a}\sum_{s_a} U_j(r_{\min,j}, s_{\min,j}) \cdot y_k^{r,s} + \sum_{r_b}\sum_{s_b} U_k(r,s) \cdot y_k^{r,s}$$
$$\geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_j^{r,s}$$

if $k \succ j, \forall k, j \in \{1,...,N\}$ and $k \neq j$,

$$(r_a, s_a) = \{r \in \{1,2,3\}, s \in \{1,...,c^r\} \mid U_k(r,s) < U_j(r_{\min,j}, s_{\min,j})\},$$

$$(r_b, s_b) = \{r \in \{1,2,3\}, s \in \{1,...,c^r\} \mid U_k(r,s) \geq U_j(r_{\min,j}, s_{\min,j})\}$$

$$z \in \Re$$

$$y_j^{r,s} \in \{0,1\} \quad \forall j \in \{1,...,N\}, \forall r \in \{1,2,3\}, \forall s \in \{1,...,c^r\}$$

3. PERFORMANCE ANALYSIS AND OPTIMIZATION

To analyse the performance of the proposed JRRM MAXILOU algorithm, a multi-RAT and multimedia wireless platform emulating GPRS, EDGE and HSDPA has been implemented in C++. The implemented simulator is not aimed at accurately modelling radio transmissions, but at measuring the efficiency of the resource distribution and optimizing it according to each RAT's operational characteristics and specific system constraints. To this aim, a single cell with equal GPRS, EDGE and HSDPA coverage has been modelled where email, web and real-time video users request resources. The users are equally distributed among the three service classes, and two different scenarios are differentiated based on the mean bit rates of the video users. While in the first scenario (S1) users demand real-time H.263 video transmissions at 16, 64 and 128kbps mean bit rates, in the second one (S2), users request 64, 256 and 512kbps video bit rates. In both scenarios, real-time video users are also equally distributed among the three simulated video bit rates. For both scenarios, cell loads of 10, 20 and 30 users have been simulated with one frequency carrier per RAT (i.e. eight timeslots for

GPRS and EDGE, and 14 HSDPA codes). It is important to note that other RATs could be considered to evaluate the proposed algorithm. To this aim, the utility functions should be redefined to include the possible radio resource assignments from these RATs.

The MAXILOU policy models the JRRM objective function and the system and service constraints by means of several linear equations with unknown binary integer variables $y_j^{r,s}$ and a real variable z. The type of optimization problems that consider both integer and real variables is referred as Mixed Integer Programming (MIP) problems in the operational research field [20]. To solve the integer lineal programming optimization problems associated to the JRRM radio resources distribution dilemma, the linear optimization software CPLEX [21] has been used. CPLEX is a powerful commercial software that incorporates state of the art algorithms and techniques, as well as proprietary solutions to solve difficult MIP problems with high computational requirements. Among the different approaches proposed to solve MIP problems, the one employed in this work is the well-known Branch and Cut method which is based on the Simplex linear programming method. The reader is referred to [20] for a detailed and analytical explanation of these linear programming methods. The CPLEX software provides an interface that allows its easy interconnection with our C++ simulation platform.

3.1. MAXILOU Performance

The performance of MAXILOU is reported in Table III. This table shows the percentage of users per service class that achieved the minimum, mean, and maximum QoS levels defined in Fig. 1. The results show that MAXILOU achieved its initial objectives. First, the highest possible QoS satisfaction level has been obtained for the user with the lowest utility value. This results in more homogeneous satisfaction levels for all users in the system. If the radio resources demand does not significantly exceed the system capacity, the maximum QoS level is achieved for most of the users (S1 scenario with 10 users per cell). However, when the radio resources demand increases, as a result of an increase of the system load or service requirements, the possibility of achieving homogeneous QoS levels for all service classes decreases, and the effect of the service priorities are most notable. Under system loads of 20 or 30 users, a certain percentage of low priority users do not even receive resources to satisfy their minimum QoS level, while some of the most demanding users (real-time

TABLE III
PERCENTAGE OF USERS ACHIEVING THEIR QoS DEMAND PER
SERVICE CLASS USING MAXILOU

	S1 - 10 users			S2 - 10 users		
	min	mean	max	min	mean	max
email	100	99.99	87.55	99.73	97.58	59.57
web	100	100	88.06	99.90	98.92	68.49
lower-rate video	100	100	99.44	100	97.08	88.06
mean-rate video	100	100	99.55	99.95	99.02	99.02
higher-rate video	100	100	100	100	96.56	96.56
	S1 - 20 users			**S2 - 20 users**		
	min	mean	max	min	mean	max
email	88.32	22.53	0.12	59.55	10.18	0.07
web	99.01	87.51	1.58	90.97	67.54	1.12
lower-rate video	100	99.29	77.10	99.81	79.30	64.89
mean-rate video	100	98.69	97.06	99.65	94.83	94.83
higher-rate video	100	100	100	99.90	88.64	88.64
	S1 - 30 users			**S2 - 30 users**		
	min	mean	max	min	mean	max
email	7.74	0.0	0.0	4.39	0.0	0.0
web	75.69	53.28	0.005	58.66	29.86	0.003
lower-rate video	100	94.25	55.73	99.04	55.36	46.78
mean-rate video	99.99	91.34	91.19	98.62	85.51	85.51
higher-rate video	100	100	100	99.60	76.15	76.15

Fig. 2. Radio resource assignments for 64kbps real-time video users.

video users) achieve their maximum QoS satisfaction level. This is due to the fact that if low priority users receive some of the discrete resources initially assigned to higher priority users to achieve their minimum QoS level, they will obtain a utility value (or QoS satisfaction) higher than their minimum QoS demand and that achieved by higher priority users, which fails to satisfy the service prioritization constraint. This effect, that would not be observed if the discrete nature of radio resources was not taken into account, results from the fact that certain radio resources offer transmission capabilities significantly higher than needed by certain low priority users. In this context, the obtained results show that MAXILOU serves the maximum possible number of users satisfying the system and service constraints, and provides the highest possible satisfaction level to the user with the lowest utility value in the system.

The results depicted in Table III show that MAXILOU adapts its solution to the current system and service conditions. To better illustrate this capability, it is interesting to analyze the JRRM radio resources distribution as the load or service QoS requirements change. To this aim, Fig. 2 depicts the radio resource assignments received by the 64kbps video users in the different simulated scenarios. This figure shows that when the radio resources demand was low, the 64kbps video users received in most of the distribution rounds the 2H assignment satisfying their maximum QoS level (see Table II). However, in scenarios with higher load or more demanding services, the 64kbps video users received radio resource assignments that provide them with lower
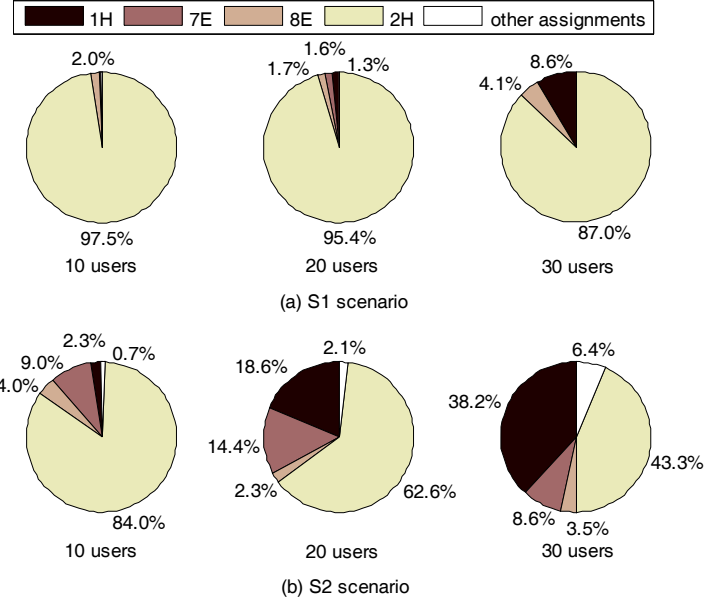
QoS levels. For example, under a load of 30 users per cell, 87% of 64kbps video users received the 2H assignment in the S1 scenario, while in the S2 scenario only 43.3% of the 64kbps video users received this assignment. On the other hand, the percentage of 64kbps video transmissions that received the 1H assignment increased from 8.6% to 38.2%. These figures demonstrate the capacity of the proposed JRRM algorithm to adapt its resource distribution decisions to the specific operating conditions.

In addition to analyzing the user perceived QoS, it is also important to investigate the usage of the available radio resources. Fig. 3 shows the percentage of JRRM distribution rounds where radio resources were left unassigned, and the percentage of distribution rounds where the minimum QoS demand of some of the users who didn't receive any resources could have been satisfied with the remaining radio resources. It is possible to observe that in a high percentage of the MAXILOU distribution rounds, some resources were left unassigned. This fact can be due to several reasons. In low load scenarios, the main reason of the high percentage of distribution rounds with unassigned resources is that all users satisfy their maximum QoS demands with the assigned radio resources, and consequently do not need additional radio resources. Other reasons arise from the definition of the distribution process. If the minimum QoS level of a given user cannot be satisfied with the available radio resources, the user is eliminated from the JRRM distribution process so that a solution can be found.

(a) Percentage of distributions rounds where radio resource were left unassigned

(b) Percentage of distributions rounds where users could have obtained their minimum QoS demand with the unassigned resources.
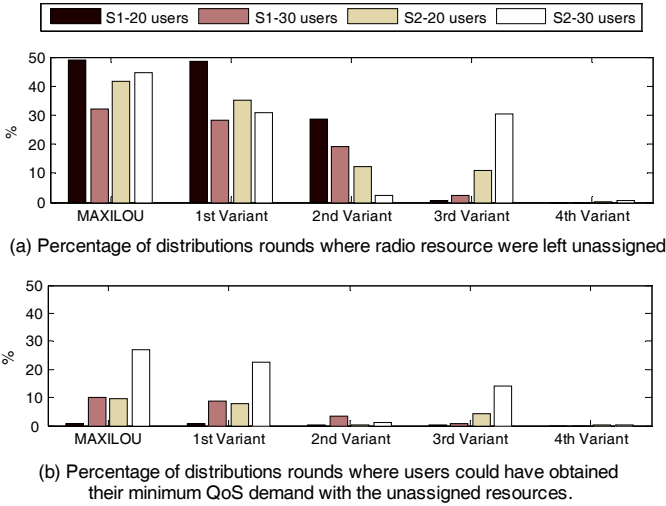
Fig. 3. Efficiency of the MAXILOU distribution process.

Following the service prioritization constraint, lower priority users are the first to be eliminated, even if users with higher QoS and resource demands are those blocking the possibility of finding a JRRM solution. In this case, it is possible that low priority users who could have been satisfied with resources that are left unassigned do not receive any radio resources. The final reason explaining the unassigned radio resources is the fact that when the maximum of the z variable is obtained, the JRRM policy stops the radio resources distribution, and users with higher utility values stop competing for additional resources. Once the z variable has been maximized, a solution assigning the highest possible satisfaction levels to the remaining users would be preferred. However, the established MAXILOU objective function prevents the proposed algorithm from discriminating among different distribution solutions providing the same value for the minimum utility value assigned to any user in the system. To overcome these limitations, and carry out a more efficient use of the available radio resources, additional variants of the original JRRM MAXILOU algorithm are proposed and described in the next section.

3.2. *MAXILOU Optimization*

The first JRRM MAXILOU variant tries to overcome the inefficiency that takes place when it is not possible to satisfy the minimum QoS demands of all users with the radio resources available in the system. In these situations, the original JRRM MAXILOU policy sequentially eliminates from the distribution process the lower priority users until a feasible solution can be achieved. Users are eliminated following the service prioritization policy, and due to

the established service QoS demands, the users who are first eliminated are generally the ones with the lower resource requirements. As a result, the users who are last eliminated before reaching a feasible solution are the ones demanding more resources, and those who were generally at the cause of why the integer linear programming techniques could not reach a feasible solution. Consequently, certain low priority users who could have been served with the available resources are not assigned any. In this context, the first JRRM variant tries to identify and eliminate only the users who block the possibility to find a feasible distribution solution with the original MAXILOU policy. Once these users are eliminated from the distribution process, the original MAXILOU algorithm is performed considering only the remaining users. To identify the users who prevented achieving a feasible solution, this variant first performs the original MAXILOU distribution process considering all the active users. If resources are left unassigned while some users did not received any, this variant tries to include again the maximum number of eliminated users to the distribution process; the last eliminated user is automatically excluded given that it was at the origin of not being able to originally achieve a feasible solution. To this end, the eliminated users are introduced one by one in the distribution process in the inverse service priority order. If a feasible integer linear programming solution can be achieved, the user remains in the distribution process. If not, the user is permanently eliminated. As a result, this first MAXILOU variant serves the highest possible number of users who better adjusts to the system and service constraints with the available resources. In addition, it is important to note that the service prioritization constraint given by (18) is satisfied by all users who receive radio resources in each distribution round. The performance obtained using the first JRRM variant is shown in Table IV and Fig. 3. The comparison of Tables III and IV shows that the first MAXILOU variant slightly increases the number of users with assigned radio resources as well as their QoS, in particular for the lower priority services. This improvement is due to a more efficient allocation of radio resources compared to MAXILOU, and does not result in a decrease of performance for real-time video users. In fact, the results depicted in Fig. 3 show that the first variant is capable of reducing the percentage of distribution rounds over which radio resources were left unassigned. Such reduction is particularly important in high load scenarios, whereas no significant differences are observed under low load

TABLE IV
PERCENTAGE OF USERS ACHIEVING THEIR QoS DEMAND PER
SERVICE CLASS USING MAXILOU'S FIRST VARIANT

	S1 - 20 users			S2 - 20 users		
	min	mean	max	min	mean	max
email	88.40	22.53	0.12	60.79	10.18	0.07
web	99.01	87.51	1.58	91.00	67.54	1.11
lower-rate video	100	99.29	77.10	99.91	79.35	62.69
mean-rate video	100	98.69	96.81	99.65	94.83	94.83
higher-rate video	100	100	100	99.90	88.59	88.59
	S1 - 30 users			S2 - 30 users		
	min	mean	max	min	mean	max
email	8.24	0.0	0.0	6.21	0.0	0.0
web	75.69	53.28	0.001	58.73	29.86	0.003
lower-rate video	100	94.28	55.68	99.42	55.35	43.30
mean-rate video	100	91.35	90.60	98.62	85.48	85.48
higher-rate video	99.98	99.98	99.98	99.60	75.68	75.68

conditions. Fig. 4 represents the average percentage of assigned radio resources per RAT, and Fig. 5 depicts the average number of email and web users who received radio resources per RAT (results for real-time services are not shown since smaller differences were observed with each MAXILOU variant). The results depicted in Fig. 4 and 5 show that the first MAXILOU variant increases the average number of email users who receive radio resources to satisfy their minimum QoS demand thanks to a more efficient use of EDGE radio resources.

The results depicted in Fig. 3 show that despite the improvements obtained with the first variant, there is still a relatively high percentage of distribution rounds with unassigned resources that could have satisfied the minimum QoS requirements for some users who did not receive any. These resources were left unassigned because the service prioritization constraint would have been broken. To overcome this limitation, a second variant is proposed. This variant initially executes the first variant distribution process, and the obtained radio resources distribution solution is used to establish a new constraint expressed as follows:

$$\sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_{2,j}^{r,s} \geq \sum_{r=1}^{3}\sum_{s=1}^{c^r} U_j(r,s) \cdot y_{1,j}^{r,s},$$
$$\forall j \in \{1,...,N\} \tag{19}$$

The distribution solution obtained with the first variant ($y_{1,j}^{r,s}$, $\forall j \in \{1,...,N\}$, $\forall r \in \{1,2,3\}$, $\forall s \in \{1,...,c^r\}$) is established as the minimum QoS levels that users will receive with the distribution solution achieved by the second variant ($y_{2,j}^{r,s}$, $\forall j \in \{1,...,N\}$, $\forall r \in \{1,2,3\}$, $\forall s \in \{1,...,c^r\}$). A new distribution process is then conducted considering all active users, but eliminating the service prioritization constraint.

As a result, the second MAXILOU variant tries to improve the best possible distribution solution that satisfies all service and system constraints, i.e. the solution initially obtained with the first MAXILOU variant. The results obtained using MAXILOU's second variant are reported in Table V. The comparison of Table V with Tables III and IV shows that the second variant improves the satisfaction levels achieved by all service classes, with the improvement being smaller for the real-time video service than for the lower priority services. In addition, the second
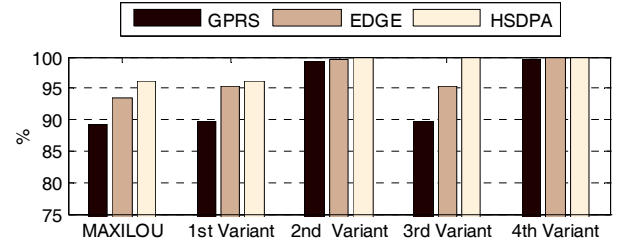

Fig. 4. Percentage of radio resources assigned per RAT (S2 scenario with 30 active users).

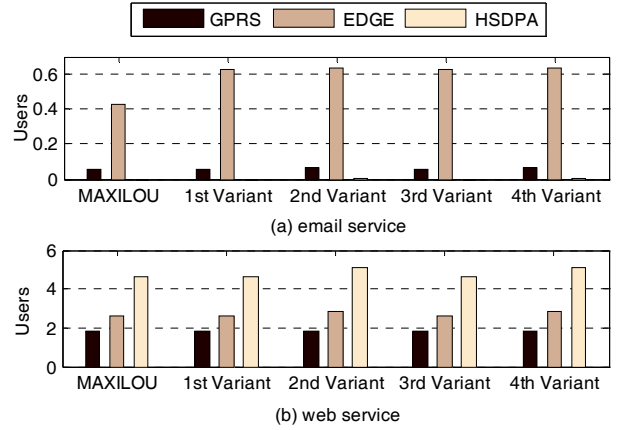(a) email service

(b) web service

Fig. 5. Number of users served by each RAT for non-real time services in the S2 scenario with 30 active users.

TABLE V
PERCENTAGE OF USERS ACHIEVING THEIR QoS DEMAND PER
SERVICE CLASS USING MAXILOU'S SECOND VARIANT

	S1 - 20 users			S2 - 20 users		
	min	mean	max	min	mean	max
email	88.98	37.17	6.78	63.20	20.22	4.91
web	99.08	87.87	1.60	93.16	69.72	1.12
lower-rate video	100	99.36	78.59	99.97	79.47	62.95
mean-rate video	100	98.69	96.85	99.72	94.90	94.90
higher-rate video	100	100	100	99.90	88.59	88.59
	S1 - 30 users			S2 - 30 users		
	min	mean	max	min	mean	max
email	8.77	0.06	0.06	6.36	0.05	0.05
web	76.77	54.35	0.001	63.42	33.10	0.01
lower-rate video	100	94.41	55.76	99.85	55.68	43.61
mean-rate video	100	91.38	90.62	98.94	85.41	85.41
higher-rate video	99.98	99.98	99.98	99.57	75.52	75.52

variant improves the percentage of users who obtain their minimum QoS demand. These improvements are due to a more efficient distribution of radio resources as shown in Fig. 3 and Fig. 4. In fact, the second MAXILOU variant considerably reduces the percentage of distribution rounds where radio resources were left unassigned, and nearly assigns all radio resources available at each RAT. By eliminating the service prioritization constraint in the last distribution process, the second MAXILOU variant assigns lower priority users high performance radio resources that were left unassigned with the first variant, which increases the number of email and web users served by EDGE and HSDPA (see Fig. 5 and Fig. 6). For example, in the S2 scenario with 20 users per cell, only 0.07% of email users received HSDPA resources with the first MAXILOU variant, while the second MAXILOU variant assigned HSDPA resources to 4.9% of email transmissions (see Fig. 6). It is important to note that these improvements are obtained without decreasing the performance of the first variant that maintained the established service priorities.

Despite the reported improvements, the second variant still leaves some unassigned resources (Fig. 3). This is due to the fact that when the lowest utility value is maximized, users achieving higher utility values stop competing for additional resources that could further improve their QoS satisfaction levels. To overcome this situation, a third MAXILOU variant is proposed. This third variant first executes the same procedure as the first JRRM variant in order to identify the maximum number of users who can be served satisfying all the system and service constraints. Next, a new distribution is executed, but only considering the users served with the first JRRM variant, and using a modified objective function expressed as follows:

$$\max \left(z + 0.001 \cdot \sum_{j=1}^{N} u_j \right) \qquad (20)$$

This modified objective function incorporates a second term that represents the sum of the utility values achieved by all users participating in the JRRM distribution process. This term provides an incentive to assign maximum possible utility values (QoS satisfaction levels) to all users participating in the distribution process, even though the lowest utility value z is already maximised. The sum of the utility values is pondered by a low factor to guarantee that its presence in (20) does not influence the original MAXILOU distribution results until the maximum z value is obtained.

Table VI reports the performance of the third MAXILOU variant. By comparing Tables IV and VI, it is interesting to note that the third MAXILOU variant serves the same number of users as the first variant, but improves the percentage of users who achieve the mean and maximum QoS levels. The third variant mainly benefits the higher priority services, with such benefits increasing with the system load. In terms of distribution of radio resources, Fig. 3 shows that the percentage of distribution rounds where resources were left unassigned decreases considerably with the third variant, except for the scenario with the higher demand for radio resources, i.e. S2 scenario with 30 users per cell. In this scenario, the percentage of distribution rounds over which radio resources were left unassigned is equal to that experienced with the first JRRM variant, and only the use of HSDPA radio resources is improved (Fig. 4). This is due to the fact that the remaining resources cannot further increase the users' utility values while guaranteeing the service prioritization constraint.

The second and third MAXILOU variants improve differently the performance achieved with the first variant. While the second MAXILOU variant increases the number of served users, the third one increases the percentage of users who achieve their maximum QoS satisfaction levels. Fig. 3 shows that in scenarios where the demand for radio resources is not high, the unassigned radio resources are a consequence of MAXILOU's objective function that results in users not competing for radio resources once the lowest
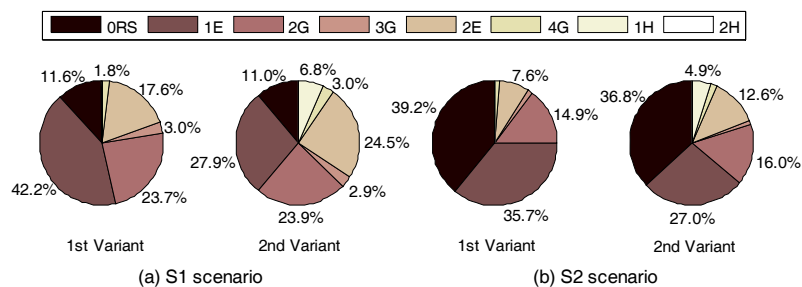

Fig. 6. Radio resource assignments for email users in the S1 and S2 scenarios with 20 users per cell.

TABLE VI
PERCENTAGE OF USERS ACHIEVING THEIR QoS DEMAND PER
SERVICE CLASS USING MAXILOU'S THIRD VARIANT

	S1 - 20 users			S2 - 20 users		
	min	mean	max	min	mean	max
email	88.35	26.47	0.13	60.79	11.56	0.07
web	99.01	88.72	6.02	91.00	68.22	4.18
lower-rate video	100	99.36	87.36	99.91	80.32	63.41
mean-rate video	100	98.36	97.41	99.65	94.93	94.93
higher-rate video	100	100	100	99.90	94.27	94.27
	S1 - 30 users			S2 - 30 users		
	min	mean	max	min	mean	max
email	8.25	0.07	0.0	6.21	0.0	0.0
web	75.69	52.64	0.03	58.73	29.86	0.04
lower-rate video	100	94.79	83.18	99.42	57.23	45.18
mean-rate video	100	95.04	94.46	98.64	86.34	86.34
higher-rate video	99.98	99.98	99.98	99.60	85.16	85.16

TABLE VII
PERCENTAGE OF USERS ACHIEVING THEIR QoS DEMAND PER
SERVICE CLASS USING MAXILOU'S FOURTH VARIANT

	S1 - 20 users			S2 - 20 users		
	min	mean	max	min	mean	max
email	89.02	41.07	7.27	63.20	22.50	5.14
web	99.08	87.80	1.58	93.16	69.64	1.12
lower-rate video	100	99.50	84.51	99.97	79.56	63.01
mean-rate video	100	98.53	96.63	99.72	94.95	94.95
higher-rate video	100	100	100	99.90	88.59	88.59
	S1 - 30 users			S2 - 30 users		
	min	mean	max	min	mean	max
email	8.78	0.08	0.06	6.36	0.05	0.05
web	76.80	54.41	0.001	63.43	33.15	0.01
lower-rate video	10	94.84	68.89	99.85	55.87	43.81
mean-rate video	100	91.26	90.47	98.92	85.66	85.66
higher-rate video	99.98	99.98	99.98	99.57	75.52	75.52

utility value is maximized. On the other hand, in scenarios characterized by a high demand of radio resources, the service prioritization constraint is the main restrictive factor. In this context, this paper also proposes a fourth MAXILOU variant aimed at exploiting the advantages exhibited by the three previous variants. The fourth variant executes the same distribution process as the second JRRM variant, but modifies the objective function according to (20) in the last distribution process where the service priority constraint is not maintained. The results reported in Table VII show that the fourth variant results in that a lower number of high priority video users achieve their maximum QoS levels compared to the third variant. However, the fourth variant guarantees the minimum QoS demand for the same number of users as the second variant (see Fig. 5 for non-real time services), while it improves the number of users obtaining their mean and maximum QoS levels for various services. Finally, Fig. 3 and Fig. 4 show that the fourth variant efficiently distributes the radio resources, and results in the lower number of unassigned radio resources compared to the three previous variants.

Once the performance of MAXILOU has been presented and optimized, it is then necessary to compare it with that achieved by other well established algorithms reported in the literature. In particular, the following JRRM algorithms are used for comparison:

- Service based RAT selection, SeRS [3]. SeRS is based on pre-established service-to-RAT assignments. For each service, a prioritized list of RATs is maintained. When a new user requests access to the system, the system tries to allocate the user to the first RAT in its corresponding list. If no capacity is available in the primary RAT for the new user, the following RAT in the list is selected instead, and so on.

- Load balancing based RAT selection, LBRS [3]. LBRS assigns each user requesting access to the system to the RAT having the lowest load in that moment. This process is followed to try balancing the load, and thereby the interference experienced in each RAT. The load metric is calculated as the ratio of utilized capacity to the total available capacity in each RAT.
- Satisfaction based RAT selection, SaRS [22]. Each time a new user requests access to the system, SaRS evaluates all users in each RAT and determines those who are satisfied in that moment. Then, SaRS assigns the new user to the RAT with a higher proportion of satisfied users.

The reference JRRM algorithms have been evaluated using the same multi-RAT and multimedia wireless platform employed to evaluate MAXILOU and its variants. Fig. 7 depicts the QoS satisfaction levels achieved by each algorithm in the S2 scenario under loads of 10 and 20 users per cell. The results show that the fourth MAXILOU variant outperforms LBRS and SaRS in all the simulated scenarios. Only in the scenario with 20 users per cell, LBRS achieves higher QoS levels for the lowest priority users, but this is done at the expense of not satisfying the minimum QoS level for a large percentage of real-time video users. The results depicted in Fig. 7 also demonstrate that MAXILOU outperforms SeRS in low load scenarios, and manages more efficiently the available radio resources since it provides higher QoS levels to all services. SeRS serves more satisfactorily the lower priority users under higher loads, but this is done at the expense of significantly degrading the mean and maximum QoS levels for the rest of services. This is due to an inefficient resource distribution with SeRS that resulted in low priority or background users being assigned resources with transmission capabilities
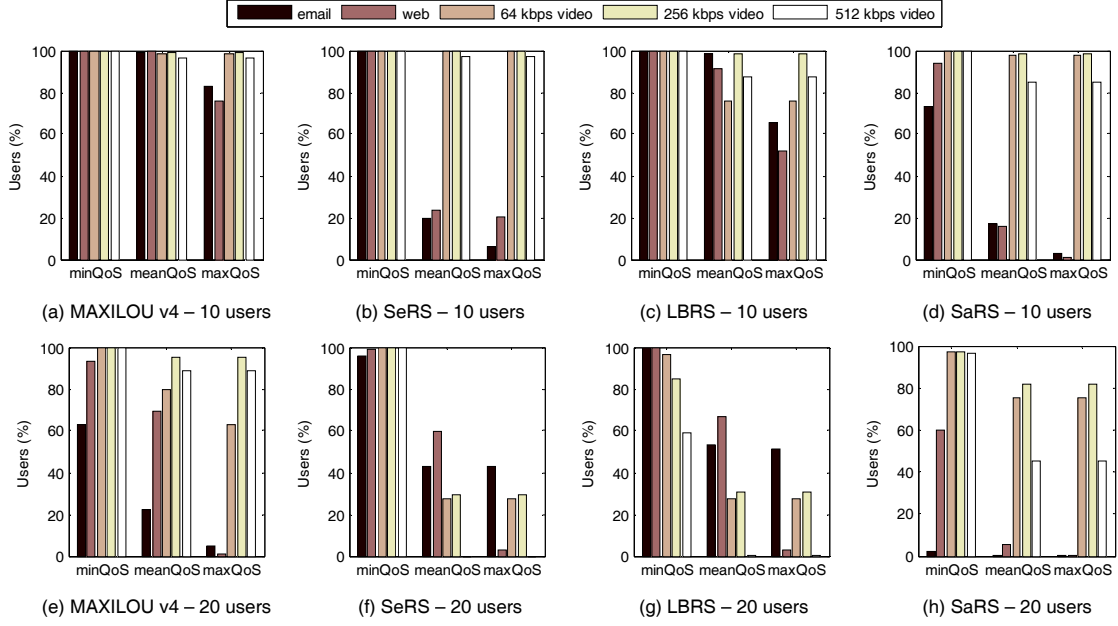
Fig. 7. QoS performance for the fourth MAXILOU variant and other reference algorithms.

exceeding their QoS demands. As a result, they achieved higher QoS satisfaction levels than real-time video users. For example, while approximately 40% of email users achieve their maximum QoS levels with SeRS, real-time video users do not even receive resources to satisfy their mean QoS level. These results highlight MAXILOU's benefits as well as its capacity to adapt and satisfy the established system conditions and QoS objectives under varying operating conditions.

4. REAL-TIME IMPLEMENTATION COST

This section is aimed at evaluating the potential application of the proposed JRRM algorithms in real systems. In particular, the section analyses two main aspects, the delay introduced by vertical handovers, and the computational cost and capability of the JRRM algorithms to obtain a solution to the distribution process in real-time.

Vertical handovers (VHO) can increase the system and user performance. The 3GPP defines the vertical handover process and the signalling message exchange for both Circuit Switched (CS) and Packet Switched (PS) services with the aim of providing a seamless inter-system handover process [23][24]. As reported in [25], VHO delays should be in the order of a few hundred milliseconds. However, vertical handovers must be carefully controlled in particular for delay sensitive real-time services. To this aim, the

MAXILOU algorithm and its variants guarantee that in each distribution round, active real-time users will maintain at least their minimum QoS level using resources from the RAT they were previously assigned (9). These users will only change RATs if they can obtain higher QoS levels using resources available from other RATs. This approach has been adopted to achieve a balance between QoS and cost of switching RATs. For non real-time services, vertical handovers are permitted without any restrictions. Fig. 8 shows the percentage of transmissions that didn't perform a vertical handover when applying MAXILOU's fourth variant (very similar results were obtained for MAXILOU and the remaining variants). The results depicted in Fig. 8 confirm that VHOs are limited in the case of real-time services, and that VHOs mainly occur under high load conditions. In this case, the vertical handovers are justified by the QoS improvements obtained by real-time video users as they switch RATs.

Since JRRM decisions are based on an increasing number of variables and data, the JRRM processing time might become an important factor that can compromise the implementation feasibility of advanced JRRM policies. To evaluate the real-time computational performance of the proposed JRRM algorithm, the Code Composer Studio (CCStudio) software [26] has been used. The CCStudio software emulates the real-time behavior of most of Texas Instrument's DSPs. Specifically, the TMS320C6455
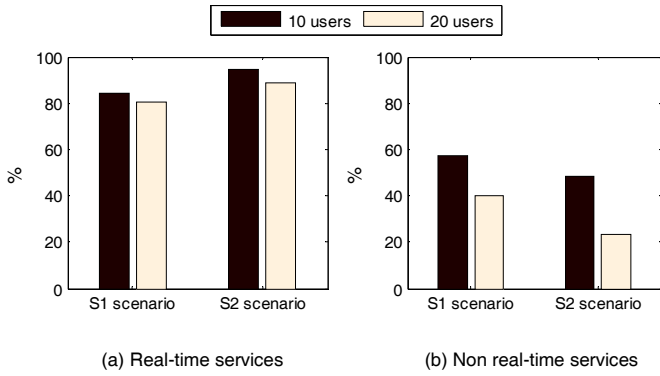
(a) Real-time services   (b) Non real-time services

Fig. 8. Percentage of transmissions that ended up without switching RATs using MAXILOU's fourth variant.

Fig. 9. Time needed by MAXILOU to find a JRRM optimum solution using the TMS320C6455 and LP_SOLVE.

[27] has been chosen since it is one of the highest performance fixed-point DSPs in the TMS320C60000TM DSP family, and is a DSP commonly used in 3G base stations where high capacities and processing data rates are required. The TMS320C6455 incorporates eight functional units to achieve maximum parallelism in processing 3G algorithms. Each functional unit works with an internal clock frequency of 1200MHz, and is capable of executing one instruction every clock cycle. As a result, the TMS320C6455 can perform at up to 9600 MIPS. To measure the real computational cost of MAXILOU, the algorithms and the linear programming tools used to solve them had to be implemented in the DSP emulator software. Given that CPLEX does not provide access to its source code, the open source LP_SOLVE linear programming solver [28] has been used instead.

Fig. 9 depicts the computational time needed by the TMS320C6455 DSP to resolve the JRRM problem using the MAXILOU algorithm. The figure shows the average time and the 95 percentile for the simulated JRRM distribution rounds. The depicted results correspond to the S1 scenario under loads of 5, 8 and 10 users per cell actively participating in the distribution process. The obtained results show that MAXILOU required an average time of 0.20s to find a JRRM distribution solution when 5 users are participating in the process. However, the average time increases up to 4.52s and 8.25s in the scenarios with 8 and 10 active users per cell. These results are derived from the number of elapsed clock cycles provided by the CCStudio DSP simulator. The CCStudio DSP simulator also provides the number of executed instructions. The analyses of the number of executed instructions shows that, although the TMS320C6455 is able to perform up to 8 instructions per clock cycle, the DSP only executed in average one instruction every 5 or 6 clock cycles depending on the number of users
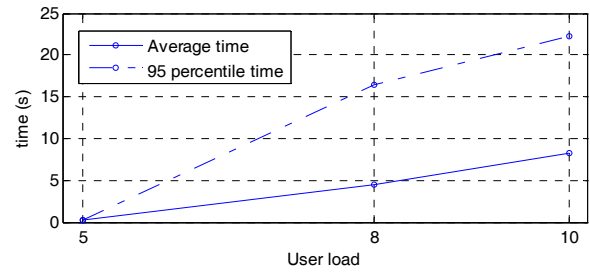
participating in the distribution problem. This low number of executed instructions per clock cycle is due to the fact that the source code of the JRRM algorithm and the linear programming tools have not been optimized for the DSP platform. This non-optimized code results in a high number of clock cycles spent without executing instructions due to cache penalties and/or memory wait states required by the physical device to access memory and read data. The linear programming solver is usually developed for computer simulation uses, and applied to analyse and solve non-real time optimization situations where the time required to access memory is not a critical issue. In this context, a computational improvement factor up to 40 or 48 could be achieved with an optimized code that makes the most of the eight functional units available in the TMS320C6455.

Additional computational improvements can also be obtained with a higher performance MIP solver such as CPLEX [29]. To this aim, the authors have analyzed the computational improvements that could be achieved using CPLEX instead of LP_SOLVE by implementing and testing both solvers in a PC. The obtained results showed that the use of CPLEX significantly reduced the JRRM computational times. For example, the MAXILOU implementation using CPLEX reduced the 95 percentile of the time needed to find a JRRM solution by a factor of 5 and 10 when 8 and 10 active users demanded resources respectively. If we apply the CPLEX and optimized code improvement factors (a factor of just 10 in the case of the optimized source code), the 95 percentile of the time needed by the DSP platform to find a JRRM solution when 10 active users request resources could be reduced below 250ms.

The previous results have demonstrated the feasibility to implement the MAXILOU algorithm in real hardware systems when current and high performance DSP platforms and MIP solvers are considered. Since all the MAXILOU variants introduce additional processes, it is also necessary to evaluate their computational performance. To do so,

their computational performance is compared against that of the original MAXILOU algorithm by analyzing their CPU execution time[1]. Fig. 10 depicts the CDF of the CPU time spent by each JRRM algorithm to achieve an optimum solution[2]. The results depicted in Fig. 10 show that in all simulated scenarios, only the third MAXILOU variant increases significantly the time required to achieve an optimum solution to the JRRM problem in comparison with the original MAXILOU algorithm. Since the previous results demonstrated the feasibility to implement and operate in real-time the MAXILOU proposal, the results in Fig. 10 confirm that the MAXILOU variants exhibiting better results (e.g., the fourth variant) can also be implemented and operated with low computational cost.
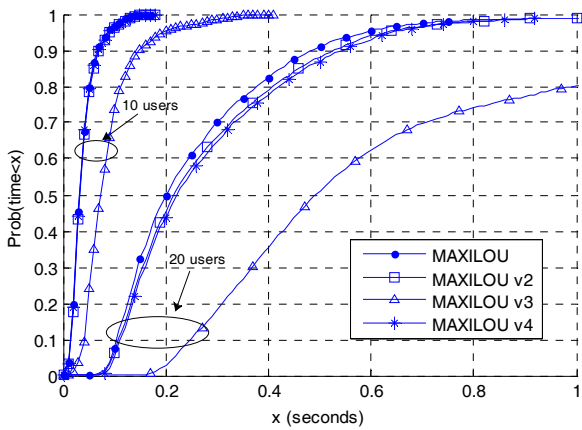

Fig. 10. CDF of the CPU time (seconds) required by MAXILOU and its variants.

5. CONCLUSIONS

This paper has presented a novel JRRM policy and several variants designed to jointly manage the radio resources of various RATs present in heterogeneous wireless networks. The proposed JRRM algorithm considers the discrete nature of mobile radio resources, and is based on integer linear programming optimization techniques that try to optimize the distribution of radio resources by maximizing their use and providing users with satisfactory QoS levels. The conducted study has demonstrated the potential of integer linear programming-based algorithms to efficiently solve JRRM problems, providing high QoS levels and efficiently using the available radio resources. The obtained results have also shown the capacity of the JRRM proposals to satisfy the established resource managements policies and requirements. The JRRM proposal could easily be adapted to other policies and requirements by modifying its system and service constraints. The paper has also highlighted the possibility to optimize the operation and resource management of the integer linear programming JRRM proposal with some simple heuristics improvements that result in a more efficient use of the radio resources. Finally, this work has analysed the computational cost of the JRRM proposal in hardware emulators, and has demonstrated its capacity to achieve distribution solutions in real-time, thereby increasing its future implementation perspectives.

REFERENCES

[1] 3GPP, Improvement of Radio Resource Management (RRM) across RNS and RNS/BSS, TR 25.881, v5.0.0, 2002.

[2] 3GPP, Improvement of Radio Resource Management (RRM) across RNS and RNS/BSS post-Rel-5, TR 25.891, v0.3.0, 2003.

[3] X. Gelabert, J. Pérez-Romero, O. Sallent, R. Agustí, On the Suitability of Load Balancing Principles in Heterogeneous Wireless Access Networks, in: Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC'05), Aalborg (Denmark), September 2005.

[4] O. Sallent, J. Pérez-Romero, R. Ljung, P. Karlsson, A. Barbaresi, Operator's RAT Selection Policies based on the Fittingness Factor Concept, in: Proc. 16th IST Mobile and Wireless Communications Summit 2007, pp. 1-5, Budapest (Hungary), July 2007.

[5] O. Yilmaz, A. Furuskär, J. Pettersson, and A. Simonsson, Access selection in WCDMA and WLAN multi-access networks, in: Proc. IEEE 61st Vehicular Technology Conference (VTC 2005), vol. 4, pp. 2220–2224, Stockholm (Sweden), April 2005.

[6] S. J. Lincke, Vertical handover policies for common radio resource management, International Journal of Communication Systems 18 (2005) 527–543.

[7] A. Tolli, P. Hakalin, and H. Holma, Performance evaluation of common radio resource management (CRRM), in: Proc. IEEE International Conference on

---

[1] Simulations have been conducted using a PC with a 2.6GHz AMD Opteron processor, 1MB of cache and 3GB of RAM.

[2] The data for the first MAXILOU variant is not shown since this variant does not provide a significant QoS performance improvement.

Communications (ICC 2002), vol. 5, pp. 3429–3433, New York (USA), April 2002.

[8] L. Pirmez, J.C. Carvalho Jr., F.C. Delicato, F. Protti, L.F.R.C. Carmo, P. F. Pires, M. Pirmez, SUTIL - Network selection based on utility function and integer linear programming, Computer Networks 54 (13) (2010) 2117-2136.

[9] Q.T. Nguyen-Vuong, N. Agoulmine, Y. Ghamri-Doudane, A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless environments, Computer Networks 52 (18) (2008) 3358-3372.

[10] O.E. Falowo, H.A. Chan, Joint Call Admission Control Algorithm for Fair Radio Resource Allocation in Heterogeneous Wireless Networks Supporting Heterogeneous Mobile Terminals, in: Proc. 7th IEEE Consumer Communications and Networking Conference (CCNC), pp. 1-5, Las Vegas (USA), January 2010.

[11] L. Giupponi, R. Agusti, J. Pérez-Romero, O. Sallent, A Novel Approach for Joint Radio Resource Management based on Fuzzy Neural Methodology, IEEE Transactions on Vehicular Technology 57 (2008) 1789-1805.

[12] D. Calabuig, J.F. Monserrat, D. Martín-Sacristán, N. Cardona, Joint Dynamic Resource Allocation for QoS Provisioning in Multi-Access and Multi-Service Wireless Systems, Mobile Networks and Applications (2009).

[13] 3GPP, Services and service capabilities, 3GPP TS 22.105, version 6.3.0, 2005.

[14] P. Barford, M. Crovella, Generating representative web workloads for network and server performance evaluation, in: Proc. International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/PERFORMANCE 1998), pp. 151–160, New York, (USA), June 1998.

[15] J. Ho, Y. Zhu, S. Madhavapeddy, Throughput and buffer analysis for GSM general packet radio service (GPRS), in: Proc. IEEE Wireless Communications and Networking Conference (WCNC 1999), pp. 1427–1431, New Orleans (USA), September 1999.

[16] O. Lázaro, D. Girma, J. Dunlop, H.263 video traffic modelling for low bit rate wireless communications, in: Proc .of the IEEE Personal, Indoor and Mobile Radio Communications (PIMRC 2004), pp. 2124-2128, Barcelona (Spain), September, 2004.

[17] L. Hanzo, P. Cherriman, J. Streit, Wireless video communications: second to third generation systems and beyond, IEEE Press, 2001.

[18] ETSI, Digital Cellular Telecommunications System (Phase 2+); General Packet Radio Service (GPRS); Overall Description of the GPRS Radio Interface; Stage 2, 3GPP TS 03.64, version 8.10.0, 1999.

[19] 3GPP, Technical Specification Group Radio Access Network; Physical layer procedures (FDD), 3GPP TS 25.214, version 7.3.0, 2006.

[20] F. S. Hillier, G. J. Lieberman, Introduction to operations research, 7th edition, McGraw-Hill, 2001.

[21] IBM ILOG CPLEX Optimizer, URL: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[22] A.P. da Silva, F.R.P. Cavalcanti, R.A. de O. Neto, VoIP capacity analysis of wireless multi-access networks using access selection schemes, in: Proc. of the IEEE Personal, Indoor and Mobile Radio Communications (PIMRC 2007), pp. 1-5, Athens (Greece), September, 2007.

[23] 3GPP, *Radio Resource Control (RRC); Protocol specification*, TS 25.331, v10.2.0, December 2010.

[24] 3GPP, *Packet-switched handover for GERAN A/Gb mode*, 3GPP TS 43.129, version 6.10.0, 2006.

[25] A. Mohammed, H. Kamal, S. AbdelWahab, 2G/3G Inter-RAT Handover Performance Analysis, in: Proc. 2nd European Conference on Antennas and Propagation, (EuCAP 2007), pp. 1-8, Edinburgh (UK), November 2007.

[26] Code Composer Studio (CCStudio) Integrated Development Environment (IDE) v4.x, URL: http://focus.ti.com/docs/toolsw/folders/print/ccstudio.html.

[27] TMS320C6455 Data Sheet, Fixed-Point Digital Signal Processor, Texas Instruments, SPRS276I, May 2005, revised April 2009.

[28] LP_Solve, URL: http://lpsolve.sourceforge.net/5.5/.

[29] H. Mittelmann, Benchmarks for Optimization Software, MILP Benchmark - serial codes, Arizona State University, February 2011. Available: http://plato.asu.edu/bench.html.