Digital Communications and Networks 3 (2017) 129–139

HOSTED BY ELSEVIER

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.elsevier.com/locate/dcan

Optimal resource allocation solutions for heterogeneous cognitive radio networks☆

Babatunde Awoyemi[a,*], Bodhaswar Maharaj[a], Attahiru Alfa[a,b]

[a] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa
[b] Department of Electrical and Computer Engineering, University of Manitoba, Canada

ARTICLE INFO

Keywords:
Cognitive radio network
Heterogeneous system
Linear and non-linear programming
Resource allocation

ABSTRACT

Cognitive Radio Networks (CRN) are currently gaining immense recognition as the most-likely next-generation wireless communication paradigm, because of their enticing promise of mitigating the spectrum scarcity and/or underutilisation challenge. Indisputably, for this promise to ever materialise, CRN must of necessity devise appropriate mechanisms to judiciously allocate their rather scarce or limited resources (spectrum and others) among their numerous users. 'Resource Allocation (RA) in CRN', which essentially describes mechanisms that can effectively and optimally carry out such allocation, so as to achieve the utmost for the network, has therefore recently become an important research focus. However, in most research works on RA in CRN, a highly significant factor that describes a more realistic and practical consideration of CRN has been ignored (or only partially explored), i.e., the aspect of the heterogeneity of CRN. To address this important aspect, in this paper, RA models that incorporate the most essential concepts of heterogeneity, as applicable to CRN, are developed and the imports of such inclusion in the overall networking are investigated. Furthermore, to fully explore the relevance and implications of the various heterogeneous classifications to the RA formulations, weights are attached to the different classes and their effects on the network performance are studied. In solving the developed complex RA problems for heterogeneous CRN, a solution approach that examines and exploits the structure of the problem in achieving a less-complex reformulation, is extensively employed. This approach, as the results presented show, makes it possible to obtain optimal solutions to the rather difficult RA problems of heterogeneous CRN.

1. Introduction

It has been recently proposed that Cognitive Radio Networks (CRN), with Dynamic Spectrum Access (DSA) and usage capabilities, can significantly help in mitigating the spectrum scarcity and/or underutilisation challenge [1,2]. The preliminaries on CRN have been well established, and a fairly comprehensive overview can be found in References [3–5]. Importantly, from the detailed literature study presented in [6], the authors identified resource allocation (RA) as a key enabler for the realisation of the potentials and promises of CRN, and therefore, a sizeable amount of work is currently being carried out in this regard. However, there are still a few challenges with RA in CRN that are yet to be extensively addressed, and one such is the necessity of developing and studying RA problems in CRN with the more-realistic consideration of it being a heterogeneous system. In all fairness, introducing heterogeneity into CRN surely portends some intricacies in the RA problem formulations, either with the objectives to be realised or the constraints to be considered. These intricacies associated with such inclusion have made most authors, in their works on RA for CRN, simply ignore or only partially explore the consideration of heterogeneity. Still, because of its significance, it is imperative to study and develop RA models for CRN that incorporate relevant heterogeneous concepts, as well as to investigate the imports of such inclusion in the overall network realisation. This paper addresses that need. To achieve the goal, in the paper several associated heterogeneous considerations applicable to CRN are investigated and analysed, while an important approach for obtaining optimal solutions to the developed RA problems is established.

2. Heterogeneity in cognitive radio networks

As earlier observed, most works on RA in CRN have been carried out with the assumption that CRN are homogeneous systems. However, the practical and realistic CRN, in almost all certainty, would

Peer review under responsibility of Chongqing University of Posts and Telecommunication.
* Correspondence author.
E-mail address: awoyemibabatunde@gmail.com (B. Awoyemi).

http://dx.doi.org/10.1016/j.dcan.2016.11.003
Received 20 July 2016; Received in revised form 8 November 2016; Accepted 10 November 2016
Available online 20 November 2016

Table 1
Concepts/classification of heterogeneity, as applicable to CRN.

S/N	Heterogeneous consideration	Basis for user classification	Basis for user classification	Basis for user classification	Basis for user classification
1.	Heterogeneous networks	Different standards – GSM, EDGE, 3G, LTE, LTE-Advanced, etc.	Different cell sizes – macrocells, microcells, femtocells, picocells, etc.	Cooperative networking possibilities – direct communication, cooperative communication, relaying techniques, etc.	Communication technologies – wired, wireless, circuit-switched, packet-switched, etc.
2.	Heterogeneous users (QoS demand)	Rate demands – different minimum rates, different service rates, etc.	Priority – high priority (HP), low priority (LP) users, priority class (PC), best efforts (BE) users, etc.	Sensitivity – sensitive users (XU), general users (GU), etc.	Delay profile – delay sensitive (DS), delay insensitive (DI), delay tolerant (DT) users, etc.
3.	Heterogeneous channels	Different channel bands – channels and/or subchannels are on different slices of frequency bands	Different channel properties – different channels and/or subchannels may have different properties	Channel usage designs – a single user being able to use different channels and/or subchannels simultaneously	Channel usage examples – OFDM/OFDMA is a classic example of how heterogeneous channels can be applied in CRN.

be heterogeneous in nature. Therefore, in system modelling, describing CRN as heterogeneous is germane to achieving the desired near-accuracy in its interpretations and applications. This is because heterogeneity, when incorporated with CRN, certainly describes more appropriately the most realistic and very practical CRN scenarios. Investigating heterogeneity in CRN, especially in its RA designs, is therefore an imperative for achieving the desired level of network efficiency and productivity. Heterogeneity, as applicable to CRN, can generally be considered from three broad perspectives – heterogeneous networks, heterogeneous users (or user demands) and heterogeneous channels. Detailed explanations on these classifications are next provided.

The concept of heterogeneous networks, commonly referred to as HetNet, has gained attention in the research domain in recent times. As a result, several ideas on HetNet are currently and actively investigated. In simple terms, HetNet explains that near-future wireless communication paradigms must be built to work in such a way that they can accommodate simultaneously two or more network configurations, standards, radio access technologies, architectures, transmission solutions, base stations, user demands, etc., in order to expand the mobile network capacity [7]. A very good example of HetNet in recent wireless standards is the femtocells and/or picocells working alongside the more traditional macrocells, as currently being employed and deployed in the LTE-Advanced technologies. The authors in [8] have given a very coordinated analysis on both the concept as well as the major technical challenges associated with HetNet architecture. Importantly, a CRN needs to incorporate the relevant elements of HetNet into its RA problem formulation, so as to achieve a high level of accuracy in its design.

Heterogeneous users or user demands, as applicable to CRN, implies that different users may have different requirements or demands and each user or group of users must be treated based on such considerations [9,10]. Heterogeneous users or service demands can be further classified using the following yardsticks:

- *Rate requirements*: This classification is based on users' minimum rate that will guarantee acceptable Quality of Service (QoS). Users that do not have any rate requirements can be treated as best effort service users. An example of the use of this classification can be found in Reference [10].
- *Service type or traffic demands*: This classification is based on the kind of service being offered by the users, e.g. voice call, live-streaming, web surfing, background services like downloading, etc. This kind of classification was employed in Reference [11].
- *Service availability*: This classification is based on whether the demands are Real-Time (RT) or non-Real-Time (NRT). For example, authors in [12,13] classified their heterogeneous users as either RT or NRT, with RT users being given a higher priority and service provisioning over NRT users.
- *Waiting-time sensitivity*: This classification is based on whether the users are Delay-Sensitive (DS) or Delay-Tolerant (DT). An instance of the use of this classification is found in Reference [14] where the SUs were classified as either DS or DT, with DS users having a very short waiting time requirement while DT users have a longer waiting time demand.

Classifying channels and/or subchannels as heterogeneous is also very important in CRN. Actually, in practical CRN, channels will most likely be located on widely separated slices of frequency bands, and these different channels may have different properties. Therefore, a user should be capable of communicating with a heterogeneous set of neighbour users using different channels or channel combinations [15]. This description is further expatiated in Reference [16] where the authors explained that the channels in CRN may not all be identical; different channels would possibly have different propagation characteristics and may support different sets of transmission rates, as opposed to homogeneous channels where propagation characteristics and transmission rates may be similar for all channels, and interfaces can easily switch on all channels as they are usually identical. The very high possibility of having multiple channels for SUs in CRN therefore requires that the devices should be capable of using heterogeneous radios. The heterogeneity of channels and radios in CRN introduces a number of issues with their design that must be properly considered and studied. Classifying channels as heterogeneous in CRN and developing and analysing models that incorporate such an inclusion is imperative for the desired near-accurate representation of CRN. In order to cater for the possibility of frequency hopping and mobility, multi-carrier transmission techniques such as the orthogonal frequency division multiple access (OFDMA) and their variants have been tipped as the most likely technologies for CRN.

The above classifications of heterogeneity are the most prominent in the field and thus, the most applicable to CRN. Table 1 gives a summary of the classifications of heterogeneity applicable to CRN.

3. Related literature on heterogeneity in cognitive radio networks

There is already in the literature a fairly sizeable number of studies undertaken on RA in CRN. However, only a few of such works have incorporated heterogeneity in their problem formulation or analysis. Even among the few works that have developed their RA problems in CRN with a consideration for some forms of heterogeneity, most authors have either obtained suboptimal solutions or simply resorted to developing heuristic(s) to solve their formulated problem(s). The reason for this is the extreme difficulty in developing and analysing formulations that can be solved for optimal solutions when heterogeneity is incorporated into CRN. Few works on RA in CRN with some forms of heterogeneous considerations, as obtained in the course of this research, are briefly reviewed.

RA in heterogeneous CRN with imperfect spectrum sensing was studied in [9]. In the work, to reduce the complexity of the optimisation problem developed, the authors proposed a subchannel allocation scheme that removes the integer constraint in the channel allocation to the SUs, thus achieving suboptimal solutions. The authors in [10] developed a RA scheme for heterogeneous CRN on the assumption that only the estimates of the channel quality information of the network are available to the SUs. The SUBS carried out its RA to the SUs based on this imperfect channel information. The complexity in computation was reduced by first assuming that subchannel allocations were already known, and on that basis, power was optimally allocated to each SU. An algorithm based on the aggressive discrete stochastic approximation was also proposed to carry out both power and channel allocations for the SUs. Similarly, RA for heterogeneous CRN was studied in [17] while including a guaranteed QoS constraint in the optimisation problem. The complex problem developed was first relaxed and then a low-complexity suboptimal solution method, which separates the RA into two steps – subchannel assignment and power allocation – was employed to obtain solution. In general, the above-mentioned works and probably the few other similar ones in the literature have all identified that the optimisation problems in RA for heterogeneous CRN are extremely complex and difficult to solve. The use of heuristics such as greedy algorithms can help to reduce the complexity and obtain suboptimal solutions [18], and most authors have rather just resorted to using that approach. However, considering that heterogeneity of channels is also a reality in CRN, heuristics that employ suboptimal greedy algorithms may not be well suited for spectrum-sharing networks such as the OFDMA-based CRN because of the multiple constraints on transmit power, interference leakage and individual user data rates [19]. Again, with heuristics, it might be difficult to know how close (or distant) the solutions obtained are to the optimal. Moreover, obtaining solutions through heuristics alone make it improbable to know what the trade-off between optimality and complexity are. With these points raised, it can be implied that heuristics (alone) might not be the best bet for solving RA problems in heterogeneous CRN. It is therefore imperative to first seek to investigate possible means of obtaining optimal solutions that are both relevant and realistic, even if solutions from heuristics are to be later sought and employed.

In this paper, the various heterogeneous considerations summarised in Table 1 are incorporated into the RA problems and the resulting formulations analysed. In solving the developed complex RA problems for heterogeneous CRN, a solution approach that examines and exploits the structure of a problem in achieving a less-complex reformulation is extensively employed. With this approach, the RA problems, even though non-deterministic polynomial-time (NP)-hard in their original formulations, are smartly reformulated as integer linear programming (ILP) problems and optimal solutions are obtained for them. A clue is taken from the work in Reference [20] and exploited in achieving the reformulation. Thereafter, a special branch-and-bound (BnB) technique, called implicit enumeration [21], is employed to solve the ILP problems. Finally, the paper investigates the impacts that assigning weights to the various categories of SUs have on the overall performance of the network.

4. System model

The system model is shown in Fig. 1. The model is a centralised, underlay, heterogeneous CRN consisting of K heterogeneous SUs and L PUs, all located within the coverage range of the CRN. The underlay scenario means that both the PUs and SUs can transmit simultaneously but the SUs must operate within the interference limit of the PUs' network so that they cause no significant harm to the PUs. The network being centralised means that the SUBS is responsible for instructing the SUs on the resources (subchannels, data rate, transmit power, modulation scheme, etc.) that have been allocated or allotted to them. The model described in Fig. 1 incorporates the different heterogeneous classifications earlier discussed. For instance, network heterogeneity is captured by separating the SUs network from that of the PUs, with each being controlled by its own base station. The SUBS informs the SUs of the resources allocated to them for transmission while the PU base station does the same job for the PUs, both working at the same time. Channel heterogeneity is taken care of by the use of the OFDMA platform, making it possible to use different slices of the frequency band for different users at the same time. User heterogeneity is accounted for by classifying and serving the various users based on some predetermined criteria.

There are N OFDMA subchannels within the coverage region of the SUBS. The SUBS selects the subchannels for each SU and relays this

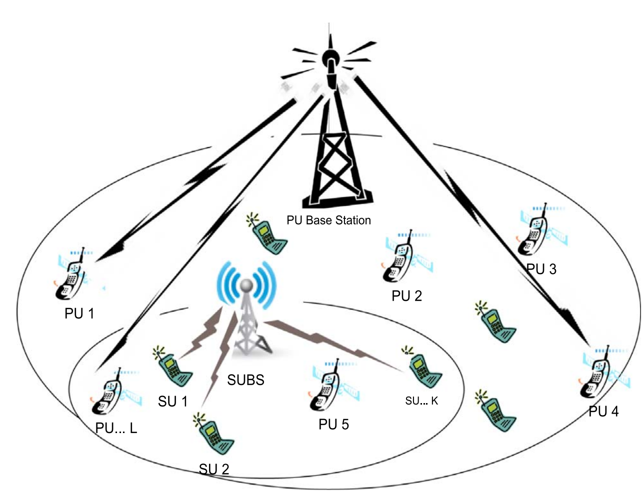

Fig. 1. System model for heterogeneous CRN.

decision to each SU through a separate control channel. The assumption is that the communication between SUs and the SUBS over the control channel is error-free and subchannels are in slow fading. Each subchannel data rate c is dependent on the modulation scheme assigned to the subchannel. Also, each category of SUs has a rate weight w ($w > 0$) associated with it. The modulation schemes considered are binary phase shift keying (BPSK), 4-quadrature amplitude modulation (QAM), 16-QAM and 64-QAM, which transmit $c = 1, 2, 4$ and 6 bits per OFDMA symbol respectively. To achieve a given bit error rate (BER) ρ value at the receiver, the minimum amount of power $P_r(c, \rho)$ required over any given subchannel for the modulation schemes can be determined easily from their power equations [20]. The minimum power for BPSK modulation is obtained from the equation $P(c, \rho) = N_\phi[c \times erfc^{-1}(2\rho)]^2$ (where $c = 1$), while for the M-ary QAM, the minimum power is given as $P(c, \rho) = \frac{2(2^c - 1)N_\phi}{3}\left[erfc^{-1}\left(\frac{c\rho\sqrt{2^c}}{2(\sqrt{2^c} - 1)}\right)\right]^2$ ($c = 2$, 4 or 6 for 4-QAM, 16-QAM and 64-QAM respectively) where, $erfc(x) = \left(\frac{1}{\sqrt{2\pi}}\right)\int_x^\infty e^{\frac{-t^2}{2}}dt$ is the complementary error function, $\pi = (22/7)$ and N_ϕ is the single-sided noise power spectral density, which is assumed to be the same for all subchannels.

For a given BER ρ value, the amount of power needed to achieve the QoS requirement generally increases (albeit non-linearly) as the number of bits (or modulation scheme) increases. The subchannel power gains matrix between the SUBS and the SUs is given as $\boldsymbol{H}^s \in R^{K\times N}$. The vector $\boldsymbol{H}^s_{k,n}$ therefore denotes the power gain between the SUBS and the kth SU at the nth subchannel. The minimum power $P_{k,n}(c_{k,n}, \rho)$ required at the kth SU over the nth subchannel to transmit $c_{k,n}$ bits is obtained by dividing the power $P(c_{k,n}, \rho)$ of that user k on the nth subchannel by the channel gain $H^s_{k,n}$ between the SUBS and the user k over that subchannel n. This is given as

$$P_{k,n}(c_{k,n}, \rho) = \frac{P(c_{k,n}, \rho)}{H^s_{k,n}}. \tag{1}$$

The power gain matrix between the SUBS and the PUs is given by $\boldsymbol{H}^p \in R^{L\times N}$. The vector $\boldsymbol{H}^p_{l,n}$ therefore denotes the subchannel power gain between the SUBS and the lth PU at the nth subchannel.

From the explanations so far presented, both network and channel heterogeneity have been effectively captured in the developed model. To capture the different classes of user heterogeneity (and their effects), the various classifications are developed and analysed one after another, following the categorisation given in the subsequent subsections. But first, for a clear understanding of the problem formulations, a general representation of the objective function and the constraints that capture the different classes of heterogeneous users is provided.

4.1. General representation of the RA formulation for heterogeneous CRN

Let the K heterogeneous SUs in a typical CRN be classified into v different categories, based on any given criterion of classification (as already identified in Table 1). The different categories of users are thus numbered $1, 2, 3, \ldots, v$ such that K_1 is the number of SUs in category 1, K_2 is the number of SUs in category 2 and so on. Let a weight w_i be attached to satisfying users in category $i \in v$. This implies therefore that w_1 is the weight attached to category 1 users, and w_v the weight attached to category v users. Given that the objective is to maximise the total data rate for all users in all categories of the network, the objective function can then be written as follows

$$\max z = \sum_{n=1}^{N}\left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=K_1+1}^{(K_1+K_2)} w_2 c_{k,n} + \sum_{k=K_1+K_2+1}^{(K_1+K_2+K_3)} w_3 c_{k,n} + \ldots + \sum_{k=K_1+\ldots+K_{v-1}+1}^{(K_1+K_2+\cdots+K_v)} w_v c_{k,n}\right) \tag{2}$$

$c_{k,n} \in \{0, 1, 2, 4, 6\}$.

Assume that these K heterogeneous SUs are classified based on their minimum data rate requirement. Let R_1 be the minimum rate that must be satisfied for users in category 1, R_2 the minimum rate that must be satisfied for users in category 2 and so on, so that R_v is the minimum rate requirement for category v SUs. The minimum rate constraints for the different categories of SUs can now be written as follows

$$\sum_{n=1}^{N} c_{k,n} \ge R_1, \quad k = 1, 2, \ldots, K_1 \tag{3}$$

$$\sum_{n=1}^{N} c_{k,n} \ge R_2, \quad k = K_1 + 1, K_1 + 2, \ldots, K_1 + K_2 \tag{4}$$

$$\sum_{n=1}^{N} c_{k,n} \ge R_3, \quad k = K_1 + K_2 + 1, K_1 + K_2 + 2, \ldots, K_1 + K_2 + K_3 \tag{5}$$

$$\vdots \sum_{n=1}^{N} c_{k,n} \ge R_v, \quad k = (K_1 + K_2 + \cdots + K_{v-1} + 1), (K_1 + K_2 + \cdots + K_{v-1} + 2), \ldots, (K_1 + K_2 + \cdots + K_v). \tag{6}$$

For completeness of the optimisation problem, other constraints such as the level of permissible interference to PUs, the maximum transmit power of the SU network, etc. are then included in the formulation to give a holistic representation. In the following subsections, the actual, complete RA formulations are presented one after another, based on the different heterogeneous user classifications provided in Table 1. In the considerations, the heterogeneous classes have been limited to two categories for each case. This is simply to make the model more manageable, and for the results to be easier to understand and compare. The models developed can however be easily extended to three, four or any given number of user categories, following the general formulation presented above, without a significant change in the results of the network.

4.2. Classification based on minimum rate requirement

In this subsection, the heterogeneous CRN are classified based on their minimum rate requirements. Hence, the K heterogeneous SUs are sub-divided into two categories and are differentiated as K_1: high-rate demand (HD) users, and $(K - K_1)$: low-rate demand (LD) users. The categories are differentiated in that they have different minimum data rate demands.

Using the representations already defined in the system model, the RA optimisation problem for heterogeneous CRN with the user heterogeneity based on the minimum rate demands of the different user categories is thus formulated as

$$\max z = \sum_{n=1}^{N}\left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=K_1+1}^{K} w_2 c_{k,n}\right), \quad c_{k,n} \in \{0, 1, 2, 4, 6\} \tag{7}$$

$$\text{subject to} \sum_{n=1}^{N} c_{k,n} \ge R_{\mathrm{I}}, \quad k = 1, 2, \ldots, K_1 \tag{8}$$

$$\sum_{n=1}^{N} c_{k,n} \ge R_{\mathrm{II}}, \quad k = K_1 + 1, K_1 + 2, \ldots, K \tag{9}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} P_{k,n} \le P_{\max} \tag{10}$$

$$\sum_{n=1}^{N} \Phi_n H^p_{l,n} \le \varepsilon_l, \quad l = 1, 2, \ldots, L \tag{11}$$

$$c_{k,n} = 0 \quad \text{if } c_{k',n} \ne 0, \ \forall\, k' \ne k; \ k = 1, 2, \ldots, K \tag{12}$$

where R_{I} is the minimum data rate that must be assigned to the kth SU in category one and R_{II} is the minimum data rate that must be assigned to the kth SU in category two, w_1 is the weight attached to the SUs in category one and w_2 is the weight attached to the SUs in category two, $\Phi_n = \sum_{k=1}^{K} P_{k,n}$ is the total power of the nth subchannel, $P_{k,n}$ is the transmit power of the kth SU over the nth subchannel, $H_{l,n}^{p}$ is the magnitude of the interference channel gain between the lth PU and the SUBS over the nth subchannel, ε_l is the threshold interference power to the lth PU from all the SUs in the network and $P_{\max}$ is the maximum transmit power at the SUBS.

The objective function in Eq. (7) gives the total weighted data rate achievable by all the SUs in the network. Constraints of Eqs. (8) and (9) show that the respective minimum data rate for category one and category two users must be met. The constraint in Eq. (10) explains that the total transmitting power of all the SUs cannot be greater than the maximum transmitting power of the SUBS. The constraint in Eq. (11) shows that the interference to PUs due to SUs' transmission must be within the acceptable interference limit. The constraint in Eq. (12) is the mutually exclusive constraint, which implies that no single subchannel can be assigned to two or more SUs. In other words, data rate in subchannel n must be 0 for user k if the subchannel n has been assigned to any other user k' that is not k.

The above formulation of the RA problem is non-linear because the power constraint in Eq. (10) is not a linear function. To make the problem solvable, after studying the problem structure, the non-linear optimisation problem is reformulated as an ILP problem. The reformulation is carried out next.

4.3. Integer linear programming reformulation of problem

By a careful study of the structure of the non-linear, complex NP-hard problem, two important facts are identified and used in achieving the ILP reformulation of the original problem. Firstly, it is observed that the bit allocation to the various subchannels is actually integer in nature. Secondly, the subchannels may either be allocated bit(s) to transmit (usually when their channel interference to PUs is within some acceptable limit) or they may not be assigned any bit to transmit (if their channel interference to PUs is too high). These facts are exploited in achieving a linear reformulation of the original problem. The ILP reformulation of the developed problem is carried out as follows: let $\boldsymbol{x_1}$ be a bit allocation vector for all the subchannels of category one users and $\boldsymbol{x_2}$ be a bit allocation vector for all the subchannels of category two users. $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are defined as:

$$\boldsymbol{x_1} = [(\boldsymbol{x}_{1,N}^{1})^T \ (\boldsymbol{x}_{1,N}^{2})^T \ \cdots \ (\boldsymbol{x}_{1,N}^{N})^T]^T \in \{0,1\}^{NK_1C\times 1} \tag{13}$$

$$\boldsymbol{x_2} = [(\boldsymbol{x}_{2,N}^{1})^T \ (\boldsymbol{x}_{2,N}^{2})^T \ \cdots \ (\boldsymbol{x}_{2,N}^{N})^T]^T \in \{0,1\}^{N(K-K_1)C\times 1} \tag{14}$$

where $\boldsymbol{x}_{1,N}^{n} = [x_{1,1,n}^T \ x_{1,2,n}^T \cdots x_{1,K,n}^T]^T \in \{0,1\}^{K_1C\times 1}$ shows that the nth subchannel is allocated with $\boldsymbol{x}_{1,k,n} = [x_{k,n,1} \ x_{k,n,2} \cdots x_{k,n,C}]^T \in \{0,1\}^{C\times 1}$; $n = 1,\ldots,N$; $k = 1,\ldots,K$; C is the number of modulation schemes considered and in this paper, C=4. This implies that, $\boldsymbol{x}_{1,k,n} = [x_{k,n,1} \ x_{k,n,2} \ x_{k,n,3} \ x_{k,n,4}]^T$. Similar explanations apply for $\boldsymbol{x_2}$. The combined bit allocation vector $\boldsymbol{x} = \boldsymbol{x_1} + \boldsymbol{x_2}$. Because of the mutually exclusive constraint, $\boldsymbol{x}_{1,N}^{n}$ and $\boldsymbol{x}_{2,N}^{n}$ can only take values from $\{[0\ 0\ \ldots\ 0]^T, [1\ 0\ \ldots\ 0]^T, [0\ 1\ \ldots\ 0]^T, \ldots, [0\ 0\ \ldots\ 1]^T\}$. This indicates that, at most one component in $\boldsymbol{x}_{1,N}^{n}$ is 1 and the other components are 0s (same applies for $\boldsymbol{x}_{2,N}^{n}$). $x_{k,n,c}$ being 1 means that the nth subchannel is assigned to the kth user, which transmits c number of bits per OFDMA symbol. If all the components of $\boldsymbol{x}_{1,N}^{n}$ (or $\boldsymbol{x}_{2,N}^{n}$) are 0s, the nth subchannel is not assigned to any user.

The modulation order vectors for the two categories of users $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ are defined as

$$\boldsymbol{b_1} = [(\boldsymbol{b}_{1,N}^{1})^T \ (\boldsymbol{b}_{1,N}^{2})^T \ \cdots \ (\boldsymbol{b}_{1,N}^{N})^T]^T \in \mathbb{Z}^{NK_1C\times 1} \tag{15}$$

$$\boldsymbol{b_2} = [(\boldsymbol{b}_{2,N}^{1})^T \ (\boldsymbol{b}_{2,N}^{2})^T \ \cdots \ (\boldsymbol{b}_{2,N}^{N})^T]^T \in \mathbb{Z}^{N(K-K_1)C\times 1} \tag{16}$$

where $\boldsymbol{b}_{1,N}^{n} = [b_{1,1,n}^T \ b_{1,2,n}^T \cdots b_{1,K,n}^T]^T \in \mathbb{Z}^{K_1C\times 1}$ and $\boldsymbol{b}_{1,k,n} = [b_{k,n,1} \ b_{k,n,2} \cdots b_{k,n,C}]^T \in \mathbb{Z}^{C\times 1}$. Similar explanations also apply for $\boldsymbol{b_2}$. Since only four modulation schemes (BPSK, 4-QAM, 16-QAM and 64-QAM) are considered, $b_{k,n} = [1\ 2\ 3\ 4]^T$. The data rate matrices for the two categories of SUs, $\boldsymbol{B_i} \in \mathbb{Z}^{K_1\times NK_1C}$ and $\boldsymbol{B_j} \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C}$ are defined respectively as

$$\boldsymbol{B_i} = \begin{bmatrix} b_1 & b_1 & \cdots & b_1 \\ b_2 & b_2 & \cdots & b_2 \\ \vdots & \vdots & & \vdots \\ b_{K_1} & b_{K_1} & \cdots & b_{K_1} \end{bmatrix}, \quad \boldsymbol{B_i} \in \mathbb{Z}^{K_1\times NK_1C} \tag{17}$$

$$\left\{\begin{matrix} b_1 = [b^T & 0_C^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times K_1C} \\ b_2 = [0_C^T & b^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times K_1C} \\ \vdots & \vdots & & \vdots \\ b_{K_1} = [0_C^T & 0_C^T & \cdots & b^T] \in \mathbb{Z}^{1\times K_1C} \end{matrix}\right\} \boldsymbol{B_j} = \begin{bmatrix} b_{K_1+1} & b_{K_1+1} & \cdots & b_{K_1+1} \\ b_{K_1+2} & b_{K_1+2} & \cdots & b_{K_1+2} \\ \vdots & \vdots & & \vdots \\ b_K & b_K & \cdots & b_K \end{bmatrix}, \tag{18}$$

$$\boldsymbol{B_j} \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C} \left\{\begin{matrix} b_{K_1+1} = [b^T & 0_C^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times (K-K_1)C} \\ b_{K_1+2} = [0_C^T & b^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times (K-K_1)C} \\ \vdots & \vdots & & \vdots \\ b_K = [0_C^T & 0_C^T & \cdots & b^T] \in \mathbb{Z}^{1\times (K-K_1)C} \end{matrix}\right\}.$$

Given that the rate weight for category one SUs is w_1 and the rate weight for category two SUs is w_2, the total data rate in the objective function (7) can thus be written as $\max_x[(w_1 \odot \boldsymbol{b_1})^T\boldsymbol{x_1} + (w_2 \odot \boldsymbol{b_2})^T\boldsymbol{x_2}]$, where $\odot$ is the Schur–Hadamard (or entry-wise) product. By defining $\boldsymbol{R_{\mathrm{I}}} \triangleq [R_1\ R_2 \cdots R_{K_1}]^T \in \mathbb{R}^{K_1\times 1}$ and $\boldsymbol{R_{\mathrm{II}}} \triangleq [R_{K_1+1}\ R_{K_1+2} \cdots R_K]^T \in \mathbb{R}^{(K-K_1)\times 1}$, the data rate per user constraint of Eq. (8) can be written as $\boldsymbol{B_i x_1} \geq \boldsymbol{R_I}$ while for Eq. (9), the data rate constraint can be written as $\boldsymbol{B_j x_2} \geq \boldsymbol{R_{\mathrm{II}}}$.

For the constraint in Eq. (10), the power transmission vector $\boldsymbol{p}$ is defined as:

$$\boldsymbol{p} = [(\boldsymbol{p}_N^1)^T \ (\boldsymbol{p}_N^2)^T \ \cdots \ (\boldsymbol{p}_N^N)^T]^T \in \mathbb{R}^{NKC\times 1} \tag{19}$$

where $\boldsymbol{p}_N^n = [\boldsymbol{p}_{1,n}^T \ \boldsymbol{p}_{2,n}^T \cdots \boldsymbol{p}_{K,n}^T]^T \in \mathbb{R}^{KC\times 1}$ and $\boldsymbol{p}_{k,n} = [p_{k,n,1}\ p_{k,n,2} \cdots p_{k,n,C}]^T \in \mathbb{R}^{C\times 1}$; $p_{k,n,c}$ is the required power to transmit c number of bits over the nth subchannel for the kth user. The power constraint in Eq. (10) can then be written as $\boldsymbol{p}^T\boldsymbol{x} \leq P_{\max}$.

In order to write the interference power constraint in Eq. (11) in terms of the vector $\boldsymbol{x}$, a matrix $\boldsymbol{A} \in \{0,1\}^{N\times NKC}$ is defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1_{KC}^T & 0_{KC}^T & \cdots & 0_{KC}^T \\ 0_{KC}^T & 1_{KC}^T & \cdots & 0_{KC}^T \\ \vdots & \vdots & & \vdots \\ 0_{KC}^T & 0_{KC}^T & \cdots & 1_{KC}^T \end{bmatrix}, \quad \boldsymbol{A} \in \{0,1\}^{N\times NKC}\ 1_{KC} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \{1\}^{KC\times 1}$$

$$0_{KC} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \{0\}^{KC\times 1}. \tag{20}$$

Given that $\boldsymbol{p} \odot \boldsymbol{x}$ is the Schur–Hadamard product of $\boldsymbol{p}$ and $\boldsymbol{x}$, $\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})$ is therefore an $N \times 1$ vector whose nth element characterises the total power used for the nth subchannel. Defining $\boldsymbol{\varepsilon_l} \triangleq [\varepsilon_1\ \varepsilon_2 \ldots \varepsilon_L]^T \in \mathbb{R}^{L\times 1}$, the interference power constraint in Eq. (11) can be written as

$$\boldsymbol{H^p}[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l. \tag{21}$$

The RA problem for heterogeneous CRN given in Eqs. (7)–(12) can now be rewritten in the ILP form as

$$z^* = \max_x[(w_1 \odot \boldsymbol{b_1})^T\boldsymbol{x_1} + (w_2 \odot \boldsymbol{b_2})^T\boldsymbol{x_2}] \tag{22}$$

$$\text{subject to } \boldsymbol{B_i x_1} \geq \boldsymbol{R}_{\mathrm{I}}, \quad k = 1, 2, \ldots, K_1 \tag{23}$$

$$\boldsymbol{B_j x_2} \geq \boldsymbol{R}_{\mathrm{II}}, \quad k = K_1 + 1, K_1 + 2, \ldots, K \tag{24}$$

$$\boldsymbol{p}^T\boldsymbol{x} \leq P_{\max} \tag{25}$$

$$\boldsymbol{H^p}[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon_l} \tag{26}$$

$$\mathbf{0}_N \leq \boldsymbol{Ax} \leq \mathbf{1}_N \tag{27}$$

$$\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x} \in \{0,1\} w_1, \quad w_2 \in \mathbb{R}^+. \tag{28}$$

The ILP problem in Eqs. (22)–(28) is a combinatorial linear programming problem which, in this paper, is solved by the branch-and-bound (BnB) method, a very adequate technique for solving such LP problems. To reduce the complexity in computation, the implicit enumeration method, which is a special case of BnB that solves binary integer LP problems, is employed [21]. Implicit enumeration makes use of the fact that each variable (in this case, the bit allocation vector $\boldsymbol{x}$) must be equal to 0 or 1 and uses this information to simplify both the branching and bounding components of the BnB process, and to determine efficiently when a node is infeasible. The implicit enumeration method thus helps to prevent possible prolonged branching by discreetly deciding what branches would not get any better result than the current 'best', and quickly eliminating such branches [22]. Thus, the implicit enumeration approach significantly reduces the overall computational complexity of the network.

4.4. Classification based on user priority or sensitivity

In this subsection, the heterogeneous classification of users is based on either the priority of the SUs or their sensitivity to changes within the network. In terms of priority, the SUs are categorised into two – high priority (HP) users and best effort service (BE) users. With this priority classification, category one HP SUs do have the higher priority and their demands are first met. The remaining resources are thereafter proportionally shared among the category two BE SUs based on a proportional rate constraint. In terms of sensitivity, the users are categorised as either sensitive users (XU) or general users (GU). The sensitivity in this classification is dependent on the data transfer rate requirement. Users in the XU category are indeed more sensitive in that they require guaranteed QoS, hence, a minimum transfer rate must be assigned to them to meet their demands at all times. The users in this category may have applications like audio and video communications that require constant data transfer at an acceptable rate for satisfactory QoS delivery. Users in the GU category are less sensitive and have less QoS requirement as compared to the XU users. GU users may be users that provide services like emails, short (text) messaging, web surfing or downloading, etc.

While these two classifications (i.e., in terms of priority or sensitivity) are slightly different from one another, the problem formulations and analyses for both classifications are however similar, hence, it is appropriate to group and study them together in this subsection. The K heterogeneous SUs in the two categories are differentiated as K_1: HP or XU users, and K_2: BE or GU users. In both considerations, the corresponding sets of the two categories of SUs are denoted as κ_A and κ_B respectively. The explanations of the system model given in the previous section are applicable in this consideration as well.

Let R_k be the minimum data rate that must be assigned to the kth SU in κ_A, γ_k be the predetermined value of the normalised proportional fairness factor for each SU in κ_B, data rate R_i indicates the rate for the element i in κ_B, let w_1 be the weight of the kth SU in κ_A and w_2 be the weight of the kth SU in κ_B. All other representations previously defined in the system model are equally applicable. The RA optimisation problem for heterogeneous CRN with priority or sensitivity considerations is thus formulated as:

$$\max z = \sum_{n=1}^{N} \left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=1}^{K_2} w_2 c_{k,n} \right), \quad c_{k,n} \in \{0, 1, 2, 4, 6\} \tag{29}$$

$$\text{subject to} \sum_{n=1}^{N} c_{k,n} \geq R_k, \quad \forall\, k \in \kappa_A \tag{30}$$

$$\frac{R_k}{\sum_{i \in \kappa_B} R_i} = \gamma_k, \quad \forall\, k \in \kappa_B \tag{31}$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} P_{k,n} \leq P_{\max} \tag{32}$$

$$\sum_{n=1}^{N} \Phi_n H_{l,n}^{p} \leq \varepsilon_l, \quad l = 1, 2, \ldots, L \tag{33}$$

$$c_{k,n} = 0 \quad \text{if } c_{k',n} \neq 0, \;\; \forall\, k' \neq k, \;\; k = 1, 2, \ldots, K. \tag{34}$$

The objective function (29) gives the total weighted data rate (throughput) achievable by all the SUs (in both categories) of the network. The constraint in Eq. (30) shows that the minimum data transfer rate for each HP or XU user of category one must be met. In Eq. (31), a proportional fairness factor is used to determine how much of the capacity left is assigned to each user in category two, the BE or GU user category. As earlier explained, Eq. (32) is the total transmitting power constraint for all the SUs, Eq. (33) is the maximum interference constraint and Eq. (34) is the mutually exclusive constraint. It is easy to show that Eq. (31) can be equivalently rewritten as:

$$R_k = \gamma_k \times \sum_{i \in \kappa_B} R_i$$

where $\sum_{i \in \kappa_B} R_i$ is the constant value of the sum of all the data rates of all category two users. Let the product $\gamma_k \times \sum_{i \in \kappa_B} R_i$ be represented as $\tilde{\gamma}_k$, then,

$$R_1 : R_2 : \ldots : R_{K_2} = \tilde{\gamma}_1 : \tilde{\gamma}_2 : \ldots : \tilde{\gamma}_{K_2} \quad \forall\, k \in \kappa_B. \tag{35}$$

Similar to the formulation in the previous heterogeneous consideration, the new formulation of the RA problem presented above is a non-linear programming problem because the power constraint in Eq. (32) is not a linear function. Again, just as in the previous subsection, to make the problem solvable, it is reformulated as an ILP problem and then solved using BnB. The ILP reformulation follows the same procedure as described in the previous subsection and it is therefore not necessary to repeat the process. The newly reformulated ILP problem of RA for heterogeneous CRN, given priority or sensitivity considerations, is therefore presented as:

$$z^* = \max_{x} [(w_1 \odot \boldsymbol{b_1})^T \boldsymbol{x_1} + (w_2 \odot \boldsymbol{b_2})^T \boldsymbol{x_2}] \tag{36}$$

$$\text{subject to } \boldsymbol{B_i x_1} \geq R_k, \quad \forall\, k \in \kappa_A \tag{37}$$

$$\boldsymbol{B_j x_2} = \tilde{\gamma}_k, \quad \forall\, k \in \kappa_B \tag{38}$$

$$\boldsymbol{p}^T \boldsymbol{x} \leq P_{\max} \tag{39}$$

$$\boldsymbol{H^p}[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l \tag{40}$$

$$\mathbf{0}_N \leq \boldsymbol{Ax} \leq \mathbf{1}_N \tag{41}$$

$$\boldsymbol{x_{1,2}} \in \{0, 1\}, \quad w_1, w_2 \in \mathbb{R}^+. \tag{42}$$

As established in the previous subsection, the ILP problem in Eqs. (36)–(42) is a combinatorial linear programming problem which is solved using the BnB method for solving ILP problems.

4.5. Classification based on delay tolerance

In this section, the SUs are classified based on their delay characteristics. The K heterogeneous SUs are differentiated as: K_1, representing the delay-sensitive (DS) users, and K_2, representing the delay-tolerant (DT) users. The corresponding sets of these two categories of SUs are also denoted as κ_A and κ_B respectively. The DS SUs in category one, because of their delay sensitivity, constantly have a minimum rate guarantee for their service to be acceptable. The DT SUs in category two could have a flexible data rate demand. Furthermore, the SUs in both categories might all have buffered data (i.e. data in a queue waiting to be transmitted), but the category two SUs, being DT, can accommodate a longer waiting period than the category one SUs. Users that fit into category one could be SUs that require services that need to be attended to urgently (for instance, in emergency service deliveries like hospital or fire-service ambulances, or service providers during disasters or crises). Such users would therefore prefer that their communications should not be initiated than be interrupted or delayed for a long duration before they can be completed. The traffic model of the SUs is described next.

For the DS SUs, their data buffer has a finite capacity. The arrival process of packets is modelled as a Poisson process [14]. The process has an independent arrival rate λ_k (packets/slot) $\forall\ k \in \kappa_A$, the set of DS SUs. For a user k that falls within this category of SUs, the sum of the average time that its data packets wait in the queue and the time required for service completion by the user gives the average delay duration of data packets, and is represented by the expectation value $\mathbb{E}[D_k]$. The data buffer for DT SUs is defined to be infinitely large such that at every given time, there will always be data packets for them to transmit. The available resources for these SUs are therefore shared proportionately, using a predetermined proportional fairness factor γ_k. Hence, for the set of DT SUs, data rate R_i indicates the rate for the element i in κ_B.

Let the maximum permissible delay duration for an acceptable QoS for each DS SU k (i.e., the delay constraint) be T_k. To meet this required QoS, the average delay during data packet transmission for the DS SU must therefore not exceed the delay constraint. Hence,

$$\mathbb{E}[D_k] \le T_k, \quad \forall\ k \in \kappa_B. \tag{43}$$

From the explanations given above, the optimisation problem of RA for heterogeneous CRN, having SUs with different delay characteristics, is presented as follows:

$$\max z = \sum_{n=1}^{N}\left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=1}^{K_2} w_2 c_{k,n}\right), \quad c_{k,n} \in \{0,\ 1,\ 2,\ 4,\ 6\} \tag{44}$$

$$\text{subject to} \sum_{n=1}^{N} c_{k,n} \ge R_k, \quad \forall\ k \in \kappa_A \tag{45}$$

$$\mathbb{E}[D_k] \le T_k, \quad \forall\ k \in \kappa_A \tag{46}$$

$$\frac{R_k}{\sum_{i\in\kappa_B} R_i} = \gamma_k, \quad \forall\ k \in \kappa_B \tag{47}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} P_{k,n} \le P_{\max} \tag{48}$$

$$\sum_{n=1}^{N} \Phi_n H_{l,n}^{p} \le \varepsilon_l, \quad l = 1, 2, \ldots, L \tag{49}$$

$$c_{k,n} = 0 \quad \text{if } c_{k',n} \ne 0,\ \forall\ k' \ne k;\ k = 1, 2, \ldots, K \tag{50}$$

where R_k is the minimum data rate that must be assigned to the kth SU of DS users, w_1 is the weight of the kth SU in κ_A and w_2 is the weight of the kth SU in κ_B. The other representations are as previously defined.

The objective function (44) gives the total data rate that the CRN can deliver. Eqs. (45) and (46) are specifical for the DS SUs. The constraint in Eq. (45) gives the minimum rate, while Eq. (46) is the permissible time delay constraint for the DS SUs. In Eq. (47), the proportional fairness factor is used to assign data rates to each user in the DT category of SUs. Similar to the previous cases considered, Eq. (48) shows that there is a total transmitting power constraint for all SUs, Eq. (49) gives the constraint on the permissible interference to PUs and Eq. (50) is the mutual exclusivity constraint.

Again, the formulated problem given in Eqs. (44)–(50) is a nonlinear optimisation problem since the power constraint in Eq. (48) is not linear. Similar to the other problems already discussed, to solve this problem, an ILP reformulation of the initial problem is realised. The reformulation follows the process already explained in the previous subsections. The ILP reformulated problem is thus given as:

$$z^* = \max_{x}[(w_1 \odot \boldsymbol{b_1})^T \boldsymbol{x_1} + (w_2 \odot \boldsymbol{b_2})^T \boldsymbol{x_2}] \tag{51}$$

$$\text{subject to } \boldsymbol{B_i x_1} \ge R_k, \quad \forall\ k \in \kappa_A \tag{52}$$

$$\mathbb{E}[D_k] \le T_k, \quad \forall\ k \in \kappa_A \tag{53}$$

$$\boldsymbol{B_j x_2} = \tilde{\gamma}_k, \quad \forall\ k \in \kappa_B \tag{54}$$

$$\boldsymbol{p}^T \boldsymbol{x} \le P_{\max} \tag{55}$$

$$\boldsymbol{H^p}[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \le \boldsymbol{\varepsilon}_l \tag{56}$$

$$\boldsymbol{0}_N \le \boldsymbol{Ax} \le \boldsymbol{1}_N \tag{57}$$

$$\boldsymbol{x}_{1,2} \in \{0,\ 1\}, \quad w_1,\ w_2 \in \mathbb{R}^+. \tag{58}$$

Similar to the earlier ones presented, the ILP problem in Eqs. (51)–(58) is a combinatorial ILP problem. Therefore, the BnB method is also employed in obtaining solutions, as used in solving previous problems.

5. Results and discussion

This section presents results for the RA solutions of all the heterogeneous CRN considerations analysed in this paper. The underlay, heterogeneous, OFDMA-based CRN described in the system model is simulated using the MATLAB software. The optimisation is carried out using the YALMIP toolbox developed for solving optimisation problems [23]. The general simulation parameters for all the results presented are the number of OFDMA subchannels $N = 64$, PUs $L = 4$ and SUs $K = 4$. The SUs, from the earlier classifications, are categorised as category one $K_1 = 2$ (representing the HD, HP, XU or DS SUs) and category two $K - K_1$ (or K_2) =2 (representing the LD, BE, GU or DT SUs). The choice of the number of PUs, SUs and other parameters used in the simulation is informed by the need to compare results obtained in this paper with similar works in the literature so as to validate the results. For all simulation results presented in this paper, random multipath fading channels of length six were generated for the PUs and SUs using statistically independent Gaussian random variables. The average channel gain between SUBS and PUs was set at 0.1 while the gain between the SUBS and SUs was set at 1. The maximum interference limit to PUs was set as 0.001 mW while the interference caused by the PUs, considered as noise by the SUs, had a power spectral density of (0.01/64)mW/subchannel. All the simulation results were obtained using 100 randomly generated channel pairs H^s and H^p. The required BER ρ has a value of 0.01 for all SUs. A weight of unity for all SU categories is considered, except in the final results where the effects of weight are explored. The results are discussed in subsequent subsections based on the various classifications carried out in the previous section, and in the order of their presentation.

5.1. Results based on minimum data rate classification

For the results discussed in this subsection, the minimum data rate for the HD category one SUs is 64 bits/user and for the LD category two SUs, it is 32 bits/user. Generally, because they require a higher data rate, the category one SUs might be the users who are billed higher or there might be some other criteria by which they are charged to pay for the better QoS being provided for them.

The results presented in Fig. 2 are similar to and validated by the ones obtained in [20,24]. The data rate (bits) allocated to each of the SUs over each subchannel is shown in Fig. 2(a). To explain the allocation in the figure, an '×' at a bit allocation of 6 for subchannel 9 means that subchannel 9 has been allocated to SU 3 to transmit 6 bits. It is significant to note that the bit allocation is done with careful consideration of the interference gains to the PUs. At high interference gains (which signifies low or less fading), the subchannels are allocated

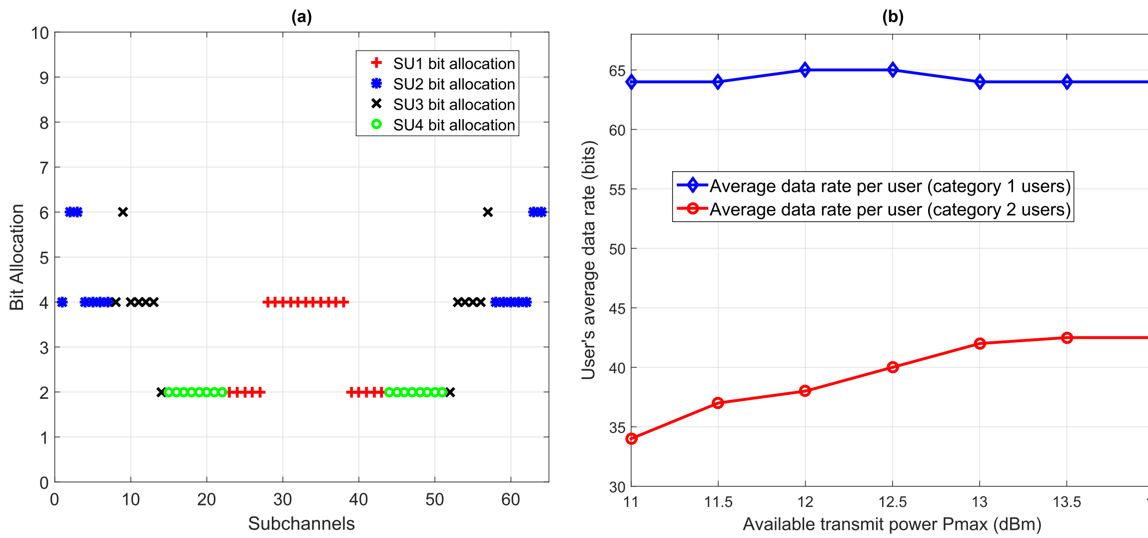

Fig. 2. (a) SUBS bit allocation for each of the SUs. Results obtained are comparable to those presented in [20], (b) Average data rate as a function of available transmit power at the SUBS for the two categories of SUs.

low data rates to avoid potential high interference power to the PUs. Conversely, at low interference gains (signifying high or deep fading) subchannels are allocated high data rates, as this will likely cause minimal interference to the PUs. The allocation algorithm developed uses this 'smartness' in its RA procedure in order to achieve optimality for the heterogeneous CRN. Examples of this smart exploitation can be seen in subchannels 2, 3, 9, 57, 63 and 64 of Fig. 2(a) where a high data rate has been allocated. The combined interference to the PUs on those subchannels is lower than the combined interference on the other subchannels. On subchannels 14–27 and 39–52, the combined interference to PUs is quite high and the subchannels have been allocated low data rates to transmit. This is the basic principle by which the bit allocation is carried out to obtain optimal results on the overall utility (average data rates, total data rates, etc.) of the network.

The average data rate of each SU against the maximum transmitting power at the SUBS is shown in Fig. 2(b) for the two categories of SUs considered. To obtain an accurate result, the interference channel gain between the PUs and the SUBS was kept constant as the transmitting power of the SUBS was varied. In the plot, the minimum data rate requirement for each category of SUs must at least be satisfied for the optimisation problem to be feasible. The plot also shows that the average data rate increases gradually as the transmitting power of the SUBS increases until it gets to a saturation point. After that point, an increase in the transmitting power at the SUBS does not cause any further increase in the average data rate of the users. This is because, the other constraints (e.g. the maximum amount of interference power leaked to the PUs) also come into play in the optimisation problem, thereby making it impossible for the SUs' data rates to keep increasing indefinitely with an increase in SUBS transmitting power.

5.2. Results based on priority and sensitivity classifications

For the results presented in this section, the HP or XU category one SUs K_1 have a minimum data transfer rate requirement of 64 bits/user while BE or GU category two SUs K_2 have the remaining resources proportionately distributed between them with a normalised proportional fairness factor $\gamma_k = 1$.

The average user data rate achieved for each category of SUs over a varying interference power to the PUs is shown in Fig. 3(a). The maximum acceptable interference power to each PU, i.e. ε_l, was varied between 20 dBm and 30 dBm with the available SUBS power set at 12 dBm, and then later increased to 30 dBm. It is important to first note that below 20 dBm interference the problem becomes infeasible. Also, it can be observed that, when the problem is feasible, the minimum data rate requirement for category one SUs is achieved at all points. Furthermore, the plot shows that the algorithm achieves a similar trend (continuous improvement) until about 22 dBm of maximum interference power. Beyond this limit, the average rate for users in both categories begins to stabilise when the SUBS maximum power is at 12 dBm. However, the average rate for users in category two SUs increases further and further when the SUBS maximum power is at 30 dBm. The reason for this is that, with a higher power at the SUBS, the average data rate of the users is greatly improved if all the other constraints do not change. It is also very significant to observe that the algorithm would rather increase the average rate of the category of SUs with BE or GU demand when it has a slightly higher resource than it would have with the category of SUs with a HP or XU demand. This signifies that it is easier to slightly (or even significantly) improve resource allocations to the category of SUs that have the most flexibility (such as the BE or GU SUs) because their demands are a lot easier to satisfy than the demands of the more rigid HP or XU SUs.

In Fig. 3(b), the total data rate or throughput of the system against varying values of interference power to the PUs is presented. The PUs' maximum interference power is varied between 20 dBm and 30 dBm for values of SUBS power at 12 dBm and 30 dBm. The result clearly shows that the CRN will generally achieve a better QoS in terms of throughput as the amount of permissible interference power to the PUs is relaxed (i.e. when the permissible interference power to PUs assumes higher values). Also, it can be seen that, for a higher SUBS power (30 dBm), the throughput keeps improving, unlike its lower SUBS power (12 dBm) counterpart where the throughput quickly stabilises, even with an increasing interference limit.

The outage probability is the probability that the formulated problem will be infeasible, given the prevalent and/or immediate constraints and conditions under consideration. In Fig. 4(a), the outage probability over a varying amount of interference power to the PUs is shown for different values of SUBS power. From the plot, it can be depicted that the outage probability decreases with an increasing interference power limit to the PUs. It can also easily be observed that the outage probability generally improves (by achieving lower values) with an increasing SUBS power ($P_{\max}$). This implies that, for a given value of interference power to PUs, the outage probability would be better at a higher SUBS power than it would be at a lower SUBS transmitting power.

Fig. 4(b) describes the total data rate of the CRN against the maximum transmitting power at the SUBS when the number of

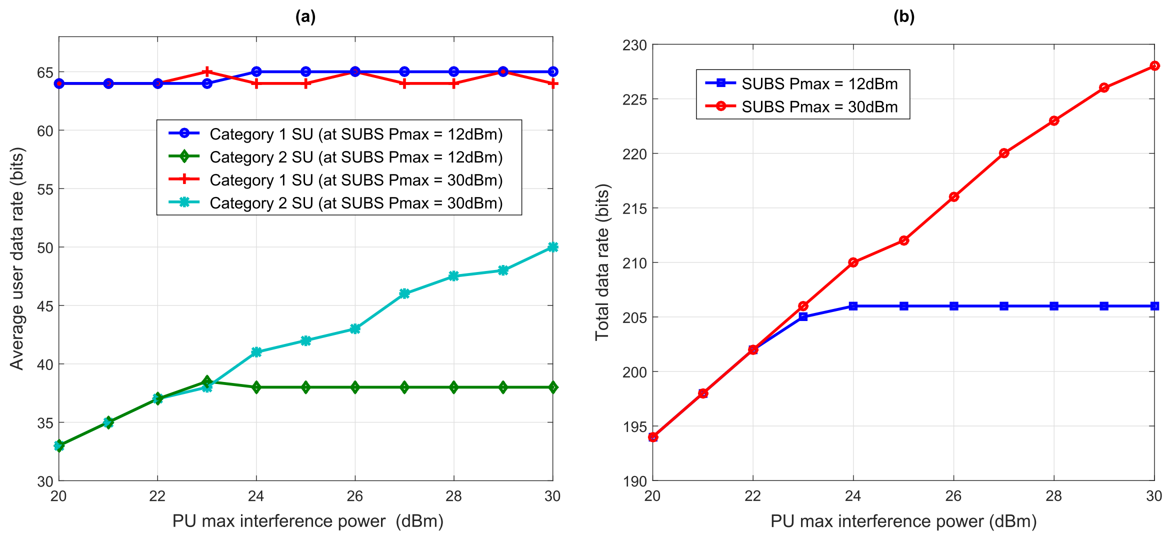

Fig. 3. (a) Average data rate of user versus maximum interference power to PUs at different SUBS power for the categories of SUs, (b) Total data rate or throughput versus maximum interference power to PUs at different SUBS power.

available SUs in the various categories is differently combined. The maximum permissible interference to PUs has been pegged at 50 dBm. From the plot, it can be observed that at an increasing SUBS power the total data rate of the CRN increases steadily until it saturates. The reason for this is that at a larger value of SUBS power a higher modulation rate (and hence, a larger data rate) is achieved for the SUs. However, the total data rate does not increase indefinitely because at some point other constraints such as the maximum interference to PUs, which are also not to be violated, come into play. The results show further that the more the number of category two users in the network (in comparison with the category one users), the better the overall throughput of the system. This can be seen in that at $\frac{K_1}{K_2} = 3$ the overall best throughput is achieved. The reason for this is that it is easier to satisfy category two users because of the flexibility in their demand, as compared to the category one users whose rate expectations are higher and quite static.

5.3. Results based on delay tolerance classification

The simulation is carried out with the number of category one DS SUs $K_1 = 2$ and given that their minimum data rate requirement is 64 bits/user, while the maximum permissible delay time $T_k = 20$ ms. The number of category two DT SUs $K_2 = 2$ and the remaining resources are proportionally distributed among them.

Fig. 5(a) gives the average data rate of each SU against the maximum transmit power at the SUBS for the two categories of SUs considered. The results for the delay tolerance classification are compared with those obtained using the minimum data rate classification already presented in Fig. 2(b). The plot shows that it takes a higher transmit power for the delay tolerant classification to become feasible, as the problem only begins to be solvable at about 12 dBm SUBS transmit power. Furthermore, the performance of the system with delay tolerance classification was constantly below comparative results obtained from the minimum rate classification. The reason that can be given for these observations is that, for the delay tolerant consideration, a further constraint in terms of the maximum permissible delay duration for the DS SUs is also incorporated into the problem formulation and its effect is what makes the overall performance of the network to be slightly degraded, as compared to only when the minimum rate requirement is considered.

Fig. 5(b) gives the total data rate of each SU against the maximum transmitting power at the SUBS for the two categories of SUs considered. The results for the delay tolerance classification are also compared with those obtained using the minimum data rate classifica-

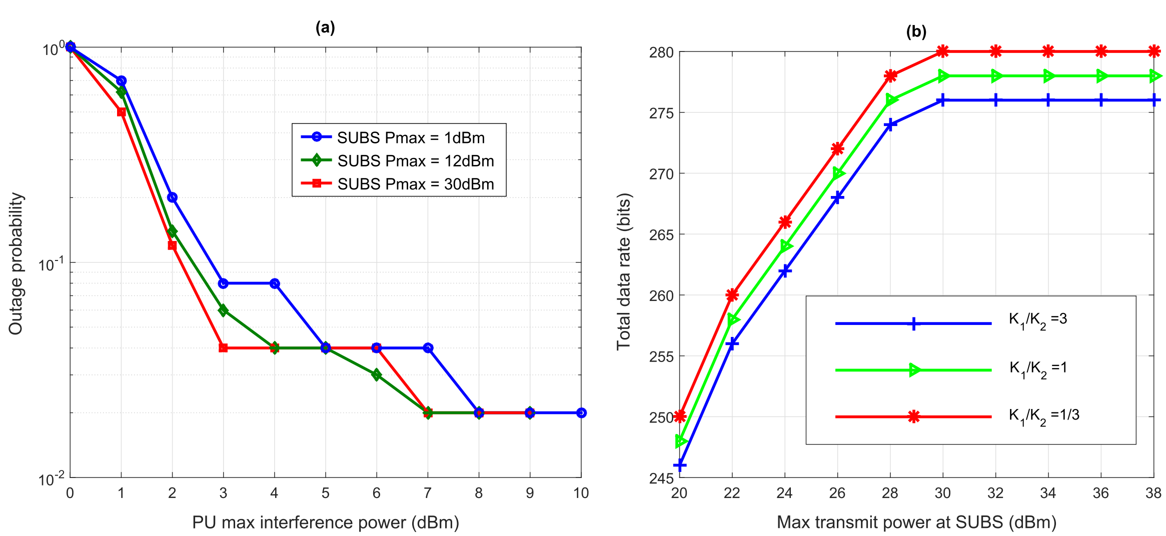

Fig. 4. (a) Outage probability versus maximum interference power to PUs at different SUBS power, (b) Total data rate of CRN against the maximum transmitting power at the SUBS for different possible combinations of categories of SUs. Maximum permissible interference to PUs is set at 50 dBm.

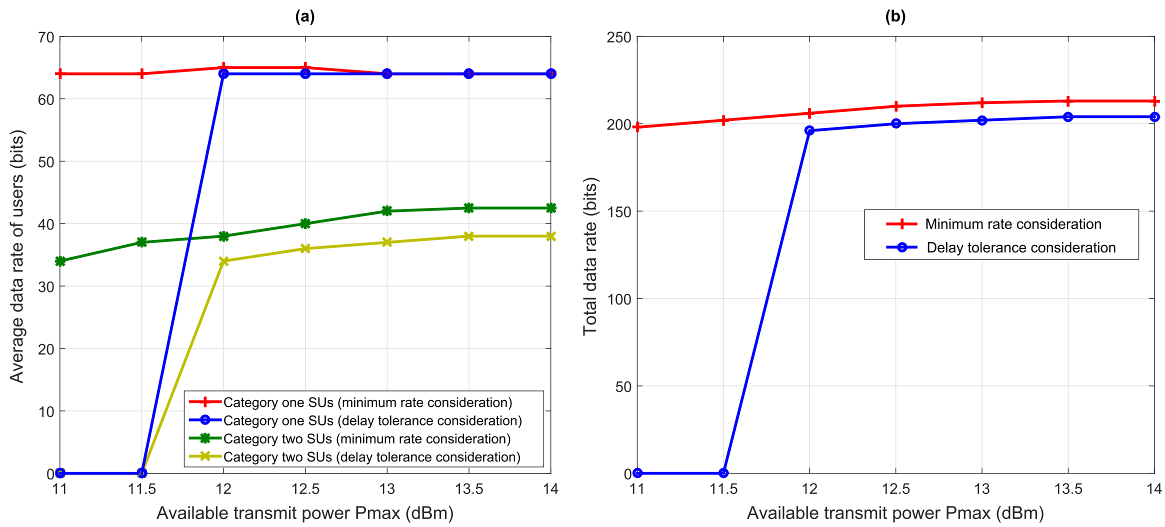

Fig. 5. (a) Average data rate against maximum transmit power at the SUBS for the delay tolerant consideration, (b) Total data rate against maximum transmit power at the SUBS for the delay tolerant consideration.

tion. The results and reasoning for the observations are similar to those given in Fig. 5(a).

5.4. Effects of weight on RA in heterogeneous CRN

Weight is an important factor in the allocation of resources to various user categories in heterogeneous CRN. This is because weight can be used effectively in a number of ways to influence the decision of the allocation algorithm to favour some user categories over other categories. Weight can therefore be used as a powerful bias mechanism in the RA decision making for CRN to provide options for further improvement that would not have been available should the user categories not have been given such weight considerations.

In Fig. 6(a), the average data rate is plotted against the weight ratio to demonstrate the importance of weight on the data rate achieved by the different categories of users. The minimum data rate classification has been employed (results can thus be compared with the ones in Fig. 2(b)), while the weight ratio between the two user categories has been steadily increased from unity to some higher values. It can be observed that, for larger values of weight ratio, the average data rate for category one users increases while the average data rate for category two users decreases. This implies therefore that, contrary to the results presented in Fig. 2(b), a higher weight in this case compels the algorithm to give a higher data rate (or resources) to users with the higher demand (the category one SUs). Indisputably, users in category one are the most valuable, since they, in some way, pay a higher price in order to get a better service. It therefore becomes meaningful to give them preference when their is a slight improvement in the quantity of resources available for use and this is achieved by the impact of the weight. The minimum data rate requirement for each category of users is, however, still satisfied in all cases, otherwise the problem becomes infeasible.

As a final consideration, Fig. 6(b) gives a comparison of the performance of different weight distributions. The authors in [25] used weights randomly chosen between 0 and 1 and normalised so that the sum of all user weights equalled 1. In this plot, as a significant improvement, three different weight distributions – uniform, normal and exponential distributions – are compared for the SUs. From the result, it can be observed that the normal weight distribution outperforms the exponential and uniform distributions, with the uniform distribution performing the least. This would imply therefore that the performance of CRN with heterogeneous users could be slightly influenced by the choice of the weight distributions employed for the network.

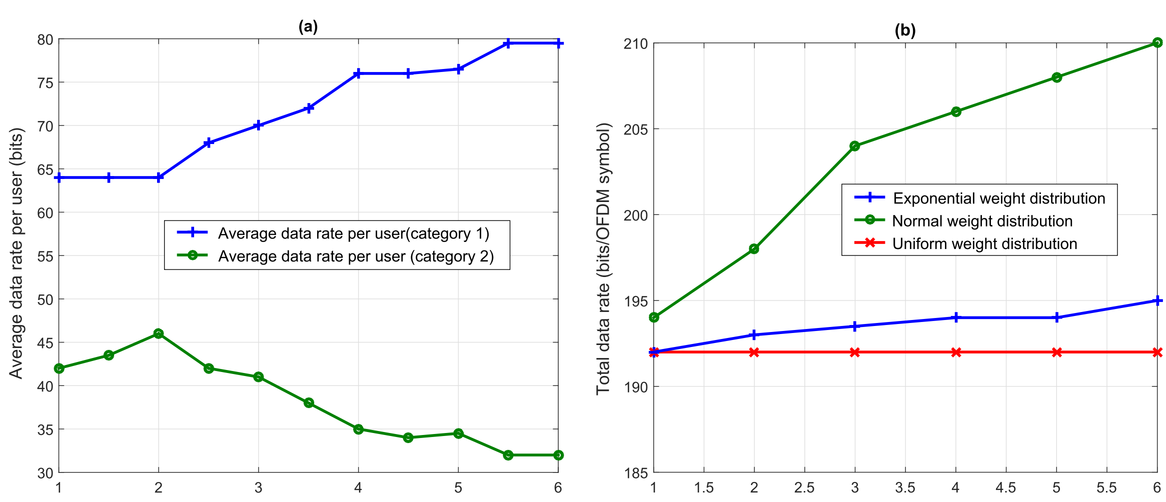

Fig. 6. (a) Average data rate at different weight ratios for the two categories of users, (b) Total data rate performance for different weight distributions.

6. Conclusion

CRN, being an emerging next-generation wireless communication paradigm, must be capable of delivering optimal productivity with the limited resources at its disposal to a wide variety of user categories. Heterogeneous CRN, which incorporates various concepts of heterogeneity as applicable to CRN, is therefore the more realistic CRN consideration. In this paper, appropriate RA models that capture the various heterogeneous considerations for CRN are developed and analysed. The models are such that heterogeneous SUs in each classification are adequately served within the limits of the network's available resources. The optimisation problems developed from the RA formulations are all NP-hard and obtaining optimal solutions to such problems are, in reality, very difficult to achieve. In the paper, however, an extensive investigation into how to solve RA problems is conducted. In the developed solution models, by carefully studying the problems' structure, easier-to-solve ILP reformulations of the original problems are realised. The BnB approach for solving ILP problems is then used to determine optimal solutions for all the classifications of heterogeneity considered. The optimal results of the average data rate, throughput, outage probability, the impact of the number of available users in each category, and the effect of weight on the overall performance of the network that were obtained were extensively discussed. A great future work will be done to develop heuristics that can achieve even less computational complexity, especially for larger networks.

References

[1] L.E. Doyle, Essentials of Cognitive Radio, The Cambridge Wireless Essentials Series, Cambridge University Press, New York, USA, 2009.
[2] B.A. Fette, Cognitive Radio Technology, Communications Engineering Series, Newness (Elsevier) Publications, Burlington, MA, USA, 2006.
[3] Y.-C. Liang, K.-C. Chen, G. Li, P. Mahonen, Cognitive radio networking and communications: an overview, IEEE Trans. Veh. Technol. 60 (7) (2011) 3386–3407. http://dx.doi.org/10.1109/TVT.2011.2158673.
[4] B. Wang, K. Liu, Advances in cognitive radio networks: a survey, IEEE J. Sel. Top. Signal Process. 5 (1) (2011) 5–23. http://dx.doi.org/10.1109/JSTSP.2010.2093210.
[5] B. Fette, Fourteen years of cognitive radio development, in: Proceedings of the IEEE MILCOM, 2013, pp. 1166–1175. http://dx.doi.org/10.1109/MILCOM.2013.200.
[6] X. Liu, Y. Zhang, Y. Li, Z. Zhang, K. Long, A survey of cognitive radio technologies and their optimization approaches, in: Proceedings of the 8th International Conference on CHINACOM, 2013, pp. 973–978. http://dx.doi.org/10.1109/ChinaCom.2013.6694736.
[7] S. Landstrom, A. Furuskar, K. Johansson, L. Falconetti, F. Kronestedt, Heterogeneous networks (HetNets) – an approach to increasing cellular capacity and coverage, in: Proceedings of the 15th International Symposium on WPMC, 2012, pp. 108–112.
[8] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T.Q.S. Quek, J. Zhang, Enhanced intercell interference coordination challenges in heterogeneous networks, IEEE Trans. Wirel. Commun. 18 (3) (2011) 22–30. http://dx.doi.org/10.1109/MWC.2011.5876497.
[9] S. Wang, Z.-H. Zhou, M. Ge, C. Wang, Resource allocation for heterogeneous cognitive radio networks with imperfect spectrum sensing, IEEE J. Sel. Areas Commun. 31 (3) (2013) 464–475. http://dx.doi.org/10.1109/JSAC.2013.130312.
[10] R. Xie, F. Yu, H. Ji, Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing, IEEE Trans. Veh. Technol. 61 (2) (2012) 770–780. http://dx.doi.org/10.1109/TVT.2011.2181966.
[11] M. Kaplan, F. Buzluca, A dynamic spectrum decision scheme for heterogeneous cognitive radio networks, in: Proceedings of the 24th International Symposium on ISCIS, 2009, pp. 697–702. http://dx.doi.org/10.1109/ISCIS.2009.5291908.
[12] S. Wang, M. Ge, C. Wang, Efficient resource allocation for cognitive radio networks with cooperative relays, IEEE J. Sel. Areas Commun. 31 (11) (2013) 2432–2441. http://dx.doi.org/10.1109/JSAC.2013.131128.
[13] A. Alshamrani, X. Shen, L.-L. Xie, QoS provisioning for heterogeneous services in cooperative cognitive radio networks, IEEE J. Sel. Areas Commun. 29 (4) (2011) 819–830. http://dx.doi.org/10.1109/JSAC.2011.110413.
[14] C. Shi, Y. Wang, P. Zhang, Joint spectrum sensing and resource allocation for multi-band cognitive radio systems with heterogeneous services, in: Proceedings of the IEEE GLOBECOM, 2012, pp. 1180–1185. http://dx.doi.org/10.1109/GLOCOM.2012.6503273.
[15] M. Ma, D.H.K. Tsang, Impact of channel heterogeneity on spectrum sharing in cognitive radio networks, in: Proceedings of the IEEE ICC, 2008, pp. 2377–2382. http://dx.doi.org/10.1109/ICC.2008.452.
[16] V. Bhandari, N.H. Vaidya, Heterogeneous multi-channel wireless networks: routing and link layer protocols, SIGMOBILE Mob. Comput. Commun. Rev. 12 (1) (2008) 43–45 URL ⟨http://doi.acm.org/10.1145/1374512.1374526⟩.
[17] F. Chen, W. Xu, Y. Guo, J. Lin, M. Chen, Resource allocation in OFDM-based heterogeneous cognitive radio networks with imperfect spectrum sensing and guaranteed QoS, in: Proceedings of the 8th International Conference on CHINACOM, 2013, pp. 46–51. http://dx.doi.org/10.1109/ChinaCom.2013.6694563.
[18] W. Guo, X. Huang, Maximizing throughput for overlaid cognitive radio networks, in: Proceedings of the IEEE MILCOM, 2009, pp. 1–7. http://dx.doi.org/10.1109/MILCOM.2009.5380005.
[19] Y. Rahulamathavan, S. Lambotharan, C. Toker, A. Gershman, Suboptimal recursive optimisation framework for adaptive resource allocation in spectrum-sharing networks, IET Signal Process. 6 (1) (2012) 27–33. http://dx.doi.org/10.1049/iet-spr.2011.0005.
[20] Y. Rahulamathavan, K. Cumanan, L. Musavian, S. Lambotharan, Optimal subcarrier and bit allocation techniques for cognitive radio networks using integer linear programming, in: Proceedings of the 15th IEEE Workshop on SSP, 2009, pp. 293–296. http://dx.doi.org/10.1109/SSP.2009.5278582.
[21] W.L. Winston, M. Venkataramanan, Introduction to Mathematical Programming, 4th edition, Thompson Brooks/Cole, Pacific Grove, CA, London, 2003.
[22] R.E. Davis, D.A. Kendrick, M. Weitzman, A branch-and-bound algorithm for zero-one mixed integer programming problems, Oper. Res. 19 (4) (1971) 1036–1044. http://dx.doi.org/10.1287/opre.19.4.1036.
[23] J. Lofberg, YALMIP : a toolbox for modeling and optimization in MATLAB, in: Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design, 2004, pp. 284–289. http://dx.doi.org/10.1109/CACSD.2004.1393890.
[24] B.S. Awoyemi, B.T. Maharaj, A.S. Alfa, Resource allocation for heterogeneous cognitive radio networks, in: Proceedings of the IEEE WCNC, 2015, pp. 1759–1763. http://dx.doi.org/10.1109/WCNC.2015.7127734.
[25] P. Cheng, Z. Zhang, H. Huang, P. Qiu, A distributed algorithm for optimal resource allocation in cognitive OFDMA systems, in: Proceedings of the IEEE ICC, 2008, pp. 4718–4723. http://dx.doi.org/10.1109/ICC.2008.884.