IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 4, 2021, accepted October 21, 2021, date of publication October 26, 2021, date of current version October 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3122882

Time-Efficient Sub-Optimal Solutions for Dynamic Spectrum Allocation in CRN With User Fairness

AHMED ABOULFOTOUH[1] AND SAMY S. SOLIMAN[2,3], (Member, IEEE)
[1]Department of Electrical and Computer Engineering, Lakehead University, Thunder Bay, ON P7B 5E1, Canada
[2]Faculty of Engineering, Cairo University, Giza 12613, Egypt
[3]CIE Program, University of Science and Technology, Zewail City 12578, Egypt

Corresponding author: Samy S. Soliman (samy.soliman@cu.edu.eg)

ABSTRACT The incredibly increasing demand for higher rates in the last decade as well as the introduction of many new applications that require wireless connectivity necessitate proper and efficient utilization of the frequency spectrum. During the last decade, the concept of cognitive radio has been introduced and extensively researched as a possible solution for telecommunication operators to increase the use of the spectrum, which is usually under-utilized. In this paper, the problem of dynamic spectrum allocation is discussed. A new mathematical formulation is proposed for the dynamic spectrum allocation problem. The new formulation defines a bi-objective function that considers the maximization of both the total system throughput and the number of active users. The proposed formulation is solved optimally using the branch and bound algorithm with a linear programming solver at its core. In addition, two novel heuristic algorithms are proposed for use instead to alleviate the time complexity of the branch and bound algorithm. Simulations show that although sub-optimal, the solutions obtained by the proposed algorithms are at least 80% of the optimum solution obtained by the branch and bound algorithm, with the advantage of significantly shorter time.

INDEX TERMS Cognitive radio, binary linear program, optimization, spectrum allocation, throughput, user fairness.

I. INTRODUCTION

The electromagnetic spectrum is a scarce natural resource and its use is typically licensed by governments and regulatory authorities. Static spectrum assignment is the dominant policy through which the spectrum is assigned. Under this policy, system operators are granted the privilege to access their licensed frequency band, while others' access to that band is prohibited. This scheme has led to the radio spectrum below 6 GHz becoming crowded. Although unallocated resources are currently limited in the frequency bands of interest, it has been reported that the actual utilization of licensed spectrum is considered low. The Federal Communications Commission (FCC) revealed that the utilization of the spectrum below 3 GHz varies vastly where the occupancy ranges between 15% to 85%. This is known as the spectrum under-utilization problem [1]. Cognitive radio networks (CRNs) have been proposed as an approach to increase the efficient utilization of licensed but under-utilized spectrum.

The associate editor coordinating the review of this manuscript and approving it for publication was Pietro Savazzi.

While there are multiple schemes of cognitive radio, this paper focuses on interweave cognitive radio. In an interweave cognitive radio network, referred to as the secondary network, the licensed spectrum is continuously monitored, and when available, spectrum holes are opportunistically accessed by the secondary network users with no interference to the active users of the primary network, which owns licensed rights of the spectrum [2]–[5].

The main idea in interweave cognitive radio networks is to utilize the spectrum as efficiently as possible while avoiding the interference with the primary network. Co-existence of the secondary user with the primary user is not allowed in this sharing mechanism. To accomplish that, the secondary network must be capable of four fundamental tasks [2]–[5], summarized in Fig. 1, as follows:

- Identify the holes in the spectrum at a particular instant of time and geographical location. This is commonly known as spectrum sensing.
- Estimate the characteristics of the spectrum holes to adapt the transmission parameters accordingly. This is known as spectrum analysis.

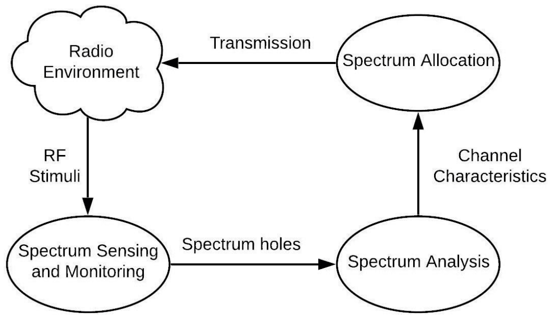


FIGURE 1. Functional model of interweave cognitive radio network.

- Allocate each secondary user (SU) an appropriate sub-channel (hole) according to its quality-of-service (QoS) requirements. This is known as spectrum allocation. The characteristics of the spectrum vacancy with respect to the SU is estimated to determine whether the available vacancy is appropriate for the SU or not.
- While the SU is active, it monitors the frequency band to detect the reappearance of the primary user (PU). This is to referred to as spectrum monitoring. In case the PU reappears, the SU halts transmission and vacates the sub-channel. Traditionally, spectrum monitoring is based on the spectrum sensing techniques.

The main contributions of the paper can be summarized as follows:

- Formulation of the dynamic spectrum allocation problem as a quality-of-service (QoS) constrained binary linear program, with a weighted objective function of the total system throughput and the number of active users. The new problem is optimally solvable using the branch and bound algorithm with a linear programming solver at its core.
- Proposing the Fair Channel Allocation (FCA) algorithm, a centralized heuristic algorithm that is capable of realizing fair sub-channel allocations among SUs.
- Proposing the Greedy Rate Allocation (GRA) algorithm, a distributed heuristic algorithm that is capable of realizing high throughput allocations per user.

The paper is organized as follows. In Section II, an introduction to CRN and a review of the related work on dynamic spectrum allocation are presented. In Section III, the system model of typical CRN is explained. The new mathematical formulation of the dynamic spectrum allocation problem is then presented in Section IV. In Section V, two heuristics algorithms are proposed for solving the dynamic spectrum allocation problem. The Fair Channel Allocation (FCA) algorithm is introduced first and then, the Greedy Rate Allocation (GRA) algorithm is discussed. The performance of the proposed algorithms are examined in Section VI in terms of the optimality of the solution, the time complexity, the average achieved total throughput, the average achieved throughout per active user and the average number of active users. Finally, conclusions are drawn in Section VII.

II. LITERATURE REVIEW AND RELATED WORK

Cognitive Radio (CR) technology, proposed by Mitola [6], Mitola and Maguire [7], Mitola [8], evolved from the concept of software-defined-radio (SDR) which liberates radio devices from hard-wired characteristics. SDRs are programmable and flexible because significant percentage of signal processing is done on general-purpose processors rather than special-purpose hardware. A CR is an evolved SDR that is aware of the radio environment and capable of adapting its communication parameters to meet the network and user demands. Moreover, it learns from the past and uses this knowledge to improve its decisions in the future. An introduction to the fundamentals of cognitive radio systems can be found in [2]–[5]. Regulatory authorities have considered allowing unlicensed users in licensed bands if there is no interference to the licensed users which lead to focusing cognitive radio research on dynamic spectrum access.

The dynamic spectrum allocation problem has been subject to extensive study in the literature [9]–[11]. As followed in [9], the first step in the solution procedure is to specify the target criteria for allocation such as minimizing interference, maximizing spectrum utilization, minimizing delay, maximizing throughput, fairness or maximizing energy efficiency. The various optimality criteria are not usually mutually exclusive and the objective of the solution can include more than one criterion. The second step is to model the spectrum allocation problem in a way that fits the target objectives. The third step is to define an algorithm to solve the problem. The execution mode of the solution can either be centralized, distributed or cluster mode.

There are numerous studies in the literature focusing on spectrum allocation in underlay-based spectrum sharing and [11] is an extensive survey of such problem. In the underlay scheme, SUs and PUs are allowed to co-exist while ensuring that SUs operate underneath a predefined interference constraint causing little-to-no degradation in the performance to the primary network. The problem of power allocation with the objective of maximizing the overall throughput while maintaining the interference levels below certain thresholds has been subject to extensive study such as in [12] and [13]. In [12], a framework for multi-hop CRNs in a fading environment with interference constraints is developed as a non-convex non-linear optimization problem and successive convex approximation is used to obtain the optimal solution. Furthermore, a practical distributed heuristic algorithm is proposed.

In [13], the problem of maximizing the throughput for all SUs under interference and received SINR constraints is modeled as a mixed integer non-linear program and transformed into a binary linear program using simplifying assumptions which is solvable in polynomial time. Throughput maximization can lead to starvation of some SUs, so some research studies include fairness in the target objectives.

In [14], the problem of fair bandwidth allocation among SUs is studied and solutions using linear programming and heuristic techniques are proposed.

In [15], the authors suggest that vehicle-to-vehicle (V2V) communication co-exist with vehicle-to-infrastructure (V2I) communication which preoccupies the spectrum along with interference management to meet the requirements for both V2V and V2I communications. The objective is to maximize the capacity of V2I links and improve the reliability of V2V links. The resource sharing problem is modeled as a multi-agent reinforcement learning problem and a solution using fingerprint-based deep Q-network is proposed for implementation in a distributed manner.

The problem of dynamic spectrum allocation in interweave spectrum sharing has been also of significant interest to numerous studies. In [16], a bijective optimization problem is designed to minimize the transmission power and maximize the rate while ensuring the satisfaction of the QoS requirements of SUs. In addition, a distributed algorithm is proposed to find the optimal solution. It is worth noting that the formulation in [16] allows the co-existence of SUs in the same sub-channel.

In [17], the channel allocation problem is reduced to a variant of the graph-coloring problem and approximate centralized and distributed solutions to achieve fairness and high throughput are presented. In [18], optimal control policies are developed to maximize the secondary network throughput while satisfying a constraint on the number of collisions with the primary network using the Lyapunov optimization technique. In [19], the problem of minimizing the delay through optimizing the routing decisions in a multi-hop CRN is studied and a distributed multi-agent learning algorithm based on adaptive fictitious play is proposed. In [20], the authors propose a multi-channel contention graph to model the interference among secondary users in a multi-hop CRN. Furthermore, an optimal maximum throughput solution and an optimal fair solution -to prevent starvation of some SUs- are presented.

In [21], a distributed multi-agent reinforcement learning approach to realize collision-free sub-channel allocation among SUs is developed and numerical results for the cases of 2 frequency sub-channels and 2 SUs, as well as 3 frequency sub-channels and 3 SUs, are presented. In [22], the authors use the spectrum access model developed in [17] and provide solutions based on genetic algorithm, quantum genetic algorithm and particle swarm optimization. The algorithms have been shown to outperform color-sensitive graph-coloring approaches.

In [23], a channel assignment problem is formulated according to Jain's fairness criterion which is classified as a quadratic integer program. A fair distributed multi-channel assignment algorithm that can realize a good trade-off between network throughput and fairness is proposed. In the proposed scheme, each SU can utilize more than one channel. In [24], a channel allocation scheme based on a greedy algorithm is proposed to maximize the network throughput for cognitive vehicular networks in low-load scenarios under the constraints of total transmission time and number of assigned channels to each vehicle.

A recent approach to spectrum allocation is based on non-orthogonal multiple access (NOMA). In Power-Domain NOMA, multiple users can use the same radio resource. Superposition coding at the transmitter and successive-interference-cancellation (SIC) at the receiver are utilized to recover the desired signal. For SIC to decode the signals correctly, the transmit power of all users must be optimized taking into account their individual channel-gains [25]. In [26], a cluster-based cognitive industrial IoT is proposed where cooperative spectrum sensing is done to improve sensing performance, and nodes transmit using NOMA. A joint optimization problem of sensing time, nodes transmit power, and the number of clusters is formulated, and the objective is to maximize the average total throughput under the constraints of minimal rate for each node, maximum total power, and cooperative detection probability. The problem is solved using power and sensing time optimization. In [27], a multi-beam satellite in the Ka-Band which uses NOMA to improve beam transmission rate is proposed for industrial IoT. The transmission rate of each beam is maximized by optimizing the transmit power of each node under the constraints of total beam power, and minimal rate for each node.

In order to improve the throughput of the secondary network, a hybrid overlay-underlay mode was proposed in [28]–[30]. In this mode, secondary users are allowed to access the spectrum both when the PU is absent, and when the PU is present, with power constraints in the latter case. In [29], the throughput in cases of perfect transmission, false alarm transmission, spectrum sharing transmission, and interference transmission were analyzed. A joint optimization problem of sub-channel transmission power and spectrum sensing time is formulated for spectrum allocation subject to interference, power, and detection probability constraints. An alternating direction optimization algorithm is proposed to solve the optimization problem. In [30], a general resource allocation problem for a secondary network utility function was proposed. Since the proposed optimization problem was non-convex, the problem was reformulated as a convex problem using the quadratic transform. Two fair resource allocation approaches were proposed and compared with respect to performance measures such as throughput and energy efficiency.

In the following, the cognitive radio system model is explained, followed by a formulation for the dynamic resource allocation problem. The proposed formulation considers both a data rate-based objective as well as fair allocation objective.

III. SYSTEM MODEL

In this section, the system model of a typical CRN is presented. A primary network is assumed to be assigned/licensed a fixed range of the spectrum for the exclusive use. In addition, a secondary network of N SUs can access the spectrum assigned to the primary network while ensuring no interference following the interweave principle. It is assumed that the whole frequency band assigned to the primary network is divided into sub-channels which are allocated to each primary and secondary users. From the point-of-view of the secondary network, the spectrum is divided into K fixed frequency sub-channels. Each SU senses the spectrum, resulting in a network status vector

$$\boldsymbol{s_i} = \left[s_{i1}, s_{i2}, \cdots, s_{iK}\right]^T \tag{1}$$

where $i \in [1, N]$ refers to the i^{th} user and s_{ij} refers to the status of sub-channel j, for $j \in [1, K]$, as seen by user i. Accordingly,

$$s_{ij} = \begin{cases} 0, & \text{if subchannel j is vacant} \\ 1, & \text{if subchannel j is occupied} \end{cases} \tag{2}$$

After sensing, each SU sends a network status packet (NS) which contains its network status vector $\boldsymbol{s_i}$ on a shared broadcast channel. The SUs can access the broadcast channel using TDMA or a contention-based protocol. In the simplest scenario, all SUs will be able to reach each other using direct transmission. However, in case some of the SUs are not within the same transmission range, sharing information can be accomplished through multi-hop transmission. All the NS packets are combined at each SU which ensures that all the SUs come to the same decision about the vacant sub-channels. The most conservative combining mechanism is to declare a sub-channel busy if one SU decided it is busy and declare a sub-channel vacant only if all the SUs agree that it is vacant. Note that it is not necessary for all the SUs to undergo spectrum sensing, it is enough if only some of the SUs do the sensing task and distribute the information.

The spectrum analysis phase then follows, in which each SU analyses the characteristics of the vacant sub-channels. Due to the difference in the geographical location of the SUs, the channel characteristics differ from a user to another.

There are two parameters which characterize each sub-channel:

1. The power spectral density (PSD) of the additive-white Gaussian noise (AWGN), denoted by σ^2.
2. The magnitude of the channel attenuation factor (reciprocal of the magnitude of the channel gain), denoted by g, which includes the path loss and the fading effects. The analysis in this paper assumes operation over narrow-band fading channels.

With the assumption of $M \le K$ spectrum holes, *i.e.* vacant sub-channels, the output of the spectrum analysis phase can be expressed in two matrices; $\boldsymbol{\sigma}$ and $\boldsymbol{g}$, which can be written as follows:

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MN} \end{bmatrix} \tag{3}$$

and

$$\boldsymbol{g} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1N} \\ g_{21} & g_{22} & \cdots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1} & g_{M2} & \cdots & g_{MN} \end{bmatrix} \tag{4}$$

where σ_{ij} refers to the square-root of the noise power of sub-channel i as seen by user j and g_{ij} refers to the attenuation factor of sub-channel i experienced by user j.

Additionally, quality-of-service (QoS) requirements are determined for each secondary user. In general, each communication session has unique requirements depending on its nature. In order to help in the problem formulation, it is assumed that the QoS requirements are determined by the minimum acceptable data rate vector, $\boldsymbol{R}$, and the maximum tolerable bit-error rate vector, $\boldsymbol{E}$, expressed, respectively, as:

$$\boldsymbol{R} = \left[R_1 \; R_2 \; \cdots \; R_N\right]^T \tag{5}$$

and

$$\boldsymbol{E} = \left[E_1 \; E_2 \; \cdots \; E_N\right]^T \tag{6}$$

where R_i is the minimum acceptable rate for the i^{th} SU, and where E_i is the maximum acceptable rate for the i^{th} SU.

The data rate and the bit-error rate are mutually entangled due to their dependence on the transmit power, channel attenuation factor and noise power. Consequently, each SU, knowing its maximum available transmit power, computes the maximum noise variance and channel attenuation such that its rate and bit-error rate requirements are satisfied. According to this computation, the acceptable performance can be related to $\boldsymbol{\sigma^*}$ and $\boldsymbol{g^*}$, defined as

$$\boldsymbol{\sigma^*} = \left[\sigma_1^* \; \sigma_2^* \; \cdots \; \sigma_N^*\right]^T \tag{7}$$

and

$$\boldsymbol{g^*} = \left[g_1^* \; g_2^* \; \cdots \; g_N^*\right]^T \tag{8}$$

where σ_i^* and g_i^* are, respectively, the square-root of the maximum tolerable noise PSD and the maximum acceptable sub-channel attenuation factor for the i^{th} SU.

IV. PROBLEM FORMULATION

In this section, a mathematical formulation of the dynamic spectrum allocation problem is developed as an optimization problem. The formal definition of any optimization problem comprises an objective function, with the purpose being either to maximize or minimize it, and constraints which must be satisfied by the solution.

Let R_{ij} denotes the achievable data rate for sub-channel i if assigned to user j. Assuming unit transmit energy and a unit bandwidth, R_{ij} can be calculated as:

$$R_{ij} = \log_2\left(1 + \frac{1}{|g_{ij}\sigma_{ij}|^2}\right) \tag{9}$$

Moreover, let $\bar{R}$ be the average achievable data rate for all possible sub-channel assignments. This can be calculated as.

$$\bar{R} = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\log_2\left(1+\frac{1}{|g_{ij}\sigma_{ij}|^2}\right) \tag{10}$$

As discussed in the literature review, an objective (*i.e.* rate, fairness, delay, and energy efficiency) must be chosen as the target of optimization. While in the majority of the related literature, the optimization is done with respect to a single objective, in here, a parameterized objective is proposed to take into account the total achievable throughput in addition to fairness among users. In the following context, fairness refers to the capability of a larger number of SUs to operate using the available spectrum holes. Accordingly, the proposed optimization problem can be formulated as:

$$\max\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\alpha+(1-\alpha)\frac{R_{ij}}{\bar{R}}\right)y_{ij} \tag{11}$$

subject to

$$\sum_{j=1}^{N} y_{ij} \leq 1, \forall i \in \{1,2,\cdots,M\} \tag{12}$$

$$\sum_{i=1}^{M} y_{ij} \leq 1, \forall j \in \{1,2,\cdots,N\} \tag{13}$$

$$\sum_{j=1}^{N}(\sigma_{ij}-\sigma_i^*)y_{ij} \leq 0, \forall i \in \{1,2,\cdots,N\} \tag{14}$$

$$\sum_{j=1}^{N}(g_{ij}-g_i^*)y_{ij} \leq 0, \forall i \in \{1,2,\cdots,N\} \tag{15}$$

The solution of the problem can be written as a binary matrix as follows.

$$\boldsymbol{y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M1} & y_{M2} & \cdots & y_{MN} \end{bmatrix} \tag{16}$$

where y_{ij} is defined as:

$$y_{ij} = \begin{cases} 1, & \text{sub-channel } j \text{ is assigned to user } i \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

The objective function in (11) can be viewed as a hybrid of two objectives, the first objective is the number of allocated sub-channels represented by the term $\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}$ and the second objective is the total achievable rate of the allocated sub-channels normalized with respect to $\bar{R}$ represented by the term $\sum_{i=1}^{N}\sum_{j=1}^{M}(\frac{R_{ij}}{\bar{R}})y_{ij}$. The two objectives are weighted by the variable α which is a tunable parameter that represents how much the first objective is favored compared to the second objective. At $\alpha = 1$, the objective function is equivalent to maximizing the number of allocated sub-channels. While at $\alpha = 0$, the objective function is equivalent to maximizing the total achievable rate of the allocated sub-channels. Introducing the parameter α makes the problem formulation flexible such that it accommodates applications that require high total throughput allocations and applications that require allocating as many sub-channels as possible (fair allocations). It can be tuned for each setting such that the solution of the optimization problem fits the target objective which is either maximizing the total throughput, maximizing the number of active users, or a weighted combination of both. The constraints in (12) ensure that no user is assigned more than one sub-channel. This constraint can be relaxed in future for the cases when the number of the available resources are larger than the number of SU, in which case, the SU can benefit from higher data rates. The constraints in (13) ensure that no sub-channel is assigned to more than a single SU. The constraints in (14) and (15) ensure that the assigned sub-channels have a noise power and a sub-channel attenuation factor less than or equal to the user demands; respectively. Including the quality-of-service constraints ensures that the users will be able to initiate communication sessions with their desired quality.

The problem, as defined in (11), can be classified as a binary linear program which can be solved optimally using the branch and bound algorithm. The branch and bound algorithm uses a linear programming solver at its core and partitions the feasible solution space into smaller subsets of solutions. The smaller subsets are systematically evaluated until the optimal solution is reached. The approach has an exponential time complexity in the problem size which can be a huge issue in some real time applications. Thus, heuristic polynomial time algorithms to find practical sub-optimal solutions are developed in the next section.

V. PROPOSED HEURISTIC ALGORITHMS

In this section, two channel allocation algorithms are proposed, the first of which is a centralized algorithm, while the second can be implemented in a distributed manner. As discussed in section IV, the optimization problem defined in equations (11) to (16) can be solved optimally using Branch and Bound which will be illustrated in section VI. The worst-case time complexity of Branch and Bound is exponential in the problem size which can problematic in delay-sensitive, and real time applications. As well, it is desirable that the allocation phase is completed in minimal time, so that the SUs can exploit the vacant sub-channels before the reappearance of PUs. Thus, heuristic polynomial-time algorithms that are capable of realizing practical solutions in a significantly shorter period of time compared to that needed by the branch-and-bound algorithm are proposed in the following. The proposed heuristics are energy-efficient as only simple computations are required, which is important for possible use cases of CRN such as in wireless sensor networks.

A. FAIR CHANNEL ALLOCATION ALGORITHM (FCA)

The optimization problem defined in equations (11) to (16) can be considered as a constraint satisfaction problem (CSP) [31]. A constraint satisfaction problem consists of three components:

- $\boldsymbol{X}$ is a set of *variables* $\{X_1, X_2, \cdots, X_n\}$.
- $\boldsymbol{D}$ is a set of *domains* $\{D_1, D_2, \cdots, D_n\}$, one for each variable. Each domain D_i consists of the set of allowable values for each variable X_i.
- $\boldsymbol{C}$ is a set of *constraints* that specify the allowable combination of values

Mapping to (11), the variables are the secondary users, and the domain of each SU is the set of indices of the sub-channels that satisfy the quality-of-service constraints for the this user, as specified by (14) and (15). With such mapping, the FCA algorithm is proposed as a variant of the backtracking algorithm that is used for solving CSPs [31].

The proposed FCA algorithm works as follows. In each allocation step, the algorithm has two choices to make. 1) It chooses the SU to which it will assign a sub-channel. 2) It chooses which sub-channel to assign to the chosen SU. The algorithm makes these decisions according to the following two heuristics:

- *Minimum Remaining Values:*
  A sub-channel is assigned to the SU with the minimum remaining values in its domain, *i.e.* the SU that has the smallest number of feasible sub-channels.
- *Least Constraining Value:*
  The sub-channel that leads to the lowest reduction in the domains of the other users is assigned to the SU. This is the sub-channel in the user's domain which is least common among all the other users' domains.

The rationale behind using the minimum remaining values heuristic is to start with the user that has the least number of options. Then, the choice of the sub-channel is made using the least constraining value heuristic to keep the options open for the subsequent allocation steps of the other SUs. The described policy avoids starvation of SUs with high requirements, and hence less options, by allocating such SUs early on. Moreover, other SUs with low requirements, and hence more options, are more likely to be allocated a feasible spectral resource as the algorithm proceeds. Accordingly, the algorithm will practically maintain equal opportunities among all users. As such, the FCA algorithm is executed in a centralized manner. The channel characteristics and the QoS requirements for all the secondary users are collected in a central node which executes the allocation algorithm. Afterwards, it broadcasts a network allocation vector with the required information for the secondary users to start transmission. The flowchart of the FCA algorithm is provided in Fig. 2.

Algorithm 1 FCA Algorithm (at a Centralized node)

```
Input: Channel characteristics σ, and g defined in (3)
       and (4), in addition to QoS requirements σ*, and
       g* defined in (7) and (8)
Output: Assignment //variable to keep
        track of channel assignment
M ← number of vacant sub-channels
N ← number of secondary users
Domains ← indices of all the sub-channels which
satisfy the SUs QoS constraints
while M > 0 and N > 0 do
    M ← M − 1
    N ← N − 1
    i ← index of the secondary user with the smallest
    domain
    j ← index of the sub-channel that is least common
    among the domains of all SUs
    Assignment[i] ← j //Assign sub-channel j
    to user i
    Update Domains such that sub-channel j is removed
    from all the SUs' domains
end
return Assignment
```

Algorithm 2 GRA Algorithm (at Each SU)

```
Input: Channel characteristics σ, and g defined in (3)
       and (4), in addition to QoS requirements σ*, and
       g* defined in (7) and (8)
Domain ← indices of all the sub-channels which satisfy
the SU QoS constraints
while Domain is not empty do
    i ← index of the sub-channel that the SU can use to
    transmit with the highest possible rate
    noBetter ← True
    for user in all other SUs do
        if user can transmit using sub-channel i at a
        higher rate then
            Remove i from Domains
            noBetter ← False
            Break for loop
        end
    end
    if noBetter is True then
        Start transmission on sub-channel i
        Terminate the algorithm
    else
        Continue while loop
    end
end
```

B. GREEDY RATE ALLOCATION ALGORITHM (GRA)

The FCA algorithm has the drawback that it is better executed in a fully centralized manner. In a distributed allocation algorithm, each SU should execute the algorithm to compute only its own allocation. However, in case of using the FCA, each SU will need to compute the allocation of other SUs as well (all SUs in the worst case) which is a waste of

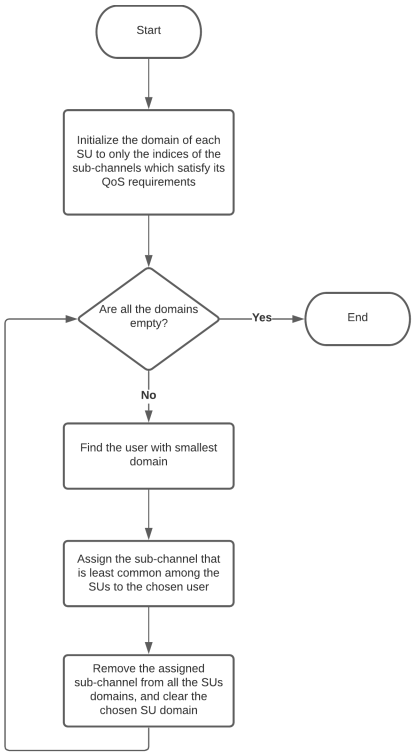


FIGURE 2. Flowchart of the FCA algorithm.

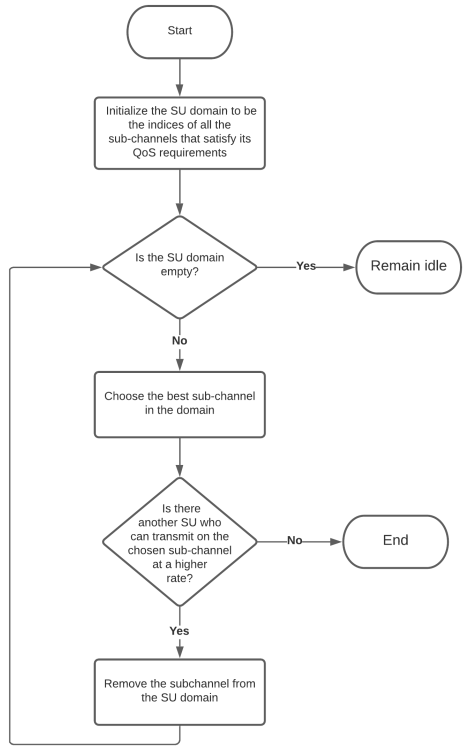


FIGURE 3. Flowchart of the GRA algorithm.

resources. In order to overcome such a challenge, the GRA is proposed. To achieve this, each SU chooses the sub-channel it operates on according to a greedy, yet cooperative policy.

The flowchart of the GRA algorithm in Fig. 3 shows the following. Each SU starts with the best sub-channel in its domain. To assess the candidacy of a sub-channel, the SU checks if there is any other SU who can transmit on this sub-channel at a higher rate. This is done by calculating the capacity of the sub-channel as seen by other SUs using the knowledge of the channel characteristics defined in eqs. (3) and (4). Note that the SU only needs to calculate the capacity as seen by SUs whose QoS requirements, defined in eqs. (7) and (8), are satisfied by the sub-channel of interest. The SU only chooses the sub-channel if there is no other SU who can use it to communicate at a higher rate. Otherwise, the SU discards the sub-channel, and checks the other sub-channels using the same policy.

It should be noted that a conflict might arise in the case of two (or more) SUs being able to use a given sub-channel to communicate at the same transmission rate. If the two SUs have the highest transmission rate over such sub-channel relative to all other SUs, the two SUs may choose the same sub-channel leading to a collision (duplicate allocation). One way to avoid such collision is that both SUs should discard this sub-channel to avoid conflicts. The drawback of this approach is that it wastes spectrum resources. The other way is that an SU, facing such a case, starts transmission, taking the chance that the other SU will end up choosing another sub-channel that achieves a higher throughput for it. In the latter approach, it is acceptable that this might lead to a collision. However, the rareness of such a scenario justifies using either approach with the inherent drawbacks.

Upon executing the GRA algorithm, each SU either finds a sub-channel, on which it has the highest rate of transmission among all SUs, this is the greedy part, or it remains idle. As such also, the GRA scheme ensures no collisions on all sub-channels without the need for communication, this is the cooperative part.

VI. RESULTS AND DISCUSSION

In order to assess the performance of the proposed algorithms, several experiments are conducted. The results of using the branch and bound algorithm are used as a benchmark for comparison.

Each of the proposed algorithms is tested for its capability to find the optimal solution at different values of α. Moreover, the average achievable total throughput, and the throughput per active user are compared across both algorithms. Both metrics are included as greedy approaches can, in general, achieve high throughput per active user without achieving a high overall total throughput, because other SUs are left without feasible sub-channels to use. Furthermore, the average number of active users, which is used as a metric for fairness, the runtime, and the time complexity are all investigated for each of the proposed algorithms.

A. SIMULATION SETUP

In the following simulation examples, a network of N secondary users and M vacant frequency sub-channels is considered. Note that the focus of the following simulation examples is to assess the dynamic resource allocation, hence the step of spectrum sensing is not included. One objective of the following simulations is to emphasize the effect of the weighting factor α on the optimal solution of the problem. Hence, it is necessarily to have a simulation setup in which there is a trade-off between the number of allocated sub-channels and the total achievable throughput using the allocated sub-channels. In other words, the simulation setup needs to have cases in which allocating as many sub-channels as possible comes at the expense of the total achievable throughput. In order to have this, the quality-of-service requirements are set to be equal to a fraction of the mean of the actual channel characteristics defined in (3) and (4), $\bar{\sigma}$ and $\bar{g}$, *i.e.*

$$\sigma_i^* = \frac{1}{k}\bar{\sigma}, \quad \text{and} \quad g_i^* = \frac{1}{k}\bar{g} \quad \forall\, i = 1, 2, \cdots, N \tag{18}$$

where $k \in [1, \infty]$ is an arbitrary factor. Such setup tightens the feasible solution space by ensuring that all users compete on the limited number of sub-channels which satisfy their QoS requirements.

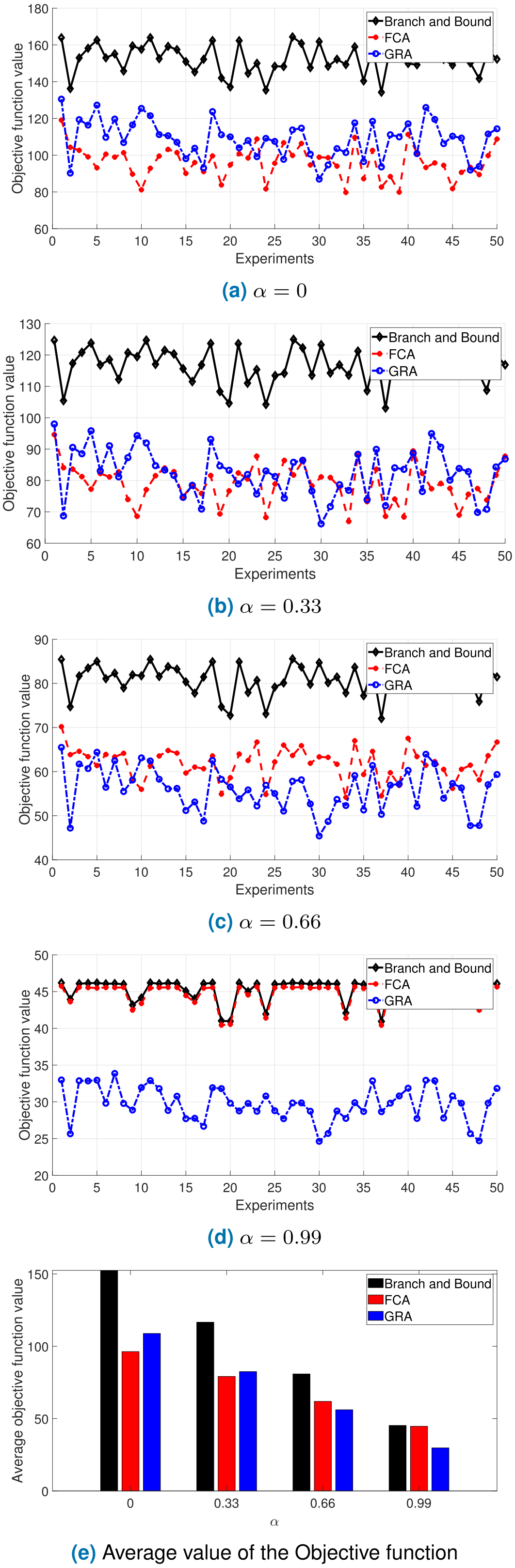
B. NUMERICAL RESULTS

This first set of simulations are used to study the optimality of the branch and bound algorithm as well as the proposed heuristic algorithms. A total of 50 experiments are generated, in which the channel characteristics are sampled from a half-normal distribution where both σ_i^* and g_i^* follow (18) with $k = 2$. The number of users is assumed to be $N = 50$, and the number of vacant sub-channels is assumed to be $M = 48$.

Fig. 4 shows the optimal solution, obtained by applying the branch and bound algorithm, as well as the solution achieved by the heuristic algorithms for values of α equal to 0, 0.33, 0.66 and 0.99, in Figs. 4a to 4d, respectively.

It can be observed that at $\alpha = 0$, where the objective function completely favors the maximization of the total achievable throughput, the GRA algorithm outperforms the FCA algorithm. Moreover, the GRA algorithm achieves about 85% of the optimal value, obtained by the branch and bound algorithm at $\alpha = 0$. For example, in the 20^{th} experiment, while the branch and bound results in a value of 137.1 for the objective function, and the GRA results in a value of 110. The average values of the objective function, over the conducted experiments, for the three algorithms are shown in Fig. 4e for different values of α. It can be seen that for $\alpha = 0$, while the branch and bound algorithm results in an average value of 152.5, and GRA algorithm results in an average value of 108.8.

As the value of α increases, the objective function becomes more inclined towards maximization of the number of active



(a) $\alpha = 0$
(b) $\alpha = 0.33$
(c) $\alpha = 0.66$
(d) $\alpha = 0.99$
(e) Average value of the Objective function

FIGURE 4. Comparison of the objective function value achieved by the algorithms for 50 experiments at (a) $\alpha = 0$, (b) $\alpha = 0.33$, (c) $\alpha = 0.66$, (d) $\alpha = 0.99$, (e) Average value of the objective function.

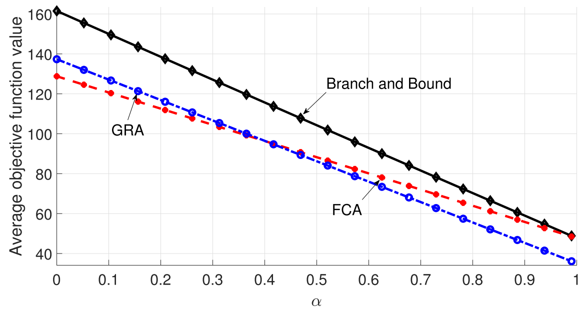
users. Consequently, the performance of the FCA algorithm starts to improve compared to that of the GRA algorithm and it achieves near optimal behavior at values of α closer to 1. For example, if can be seen from Fig. 4e that the FCA achieves 98.76% of the optimal value at $\alpha = 0.99$. These observations are justified by fact that the FCA algorithm makes cautious allocations to maintain equal opportunities among all users. This can waste the chances of allocating high-throughput sub-channels. On the other hand, the GRA algorithm prioritizes users that can transmit with higher throughput and depletes the resources by prioritizing high throughput transmissions.

While the branch and bound results in the optimal solution, it will be shown later that it has a complexity and runtime disadvantage. With the proposed heuristic algorithms, sub-optimum solutions can be obtained with far less complexity. Moreover, from the previous examples, it can be concluded that, for the applications where the maximization of the number of active users is favored, the FCA algorithm performs better and the resulting solutions are near-optimal. On the other hand, for the applications where the maximization of the total throughput is required, irrespective the number of operating users, the GRA algorithm is recommended.

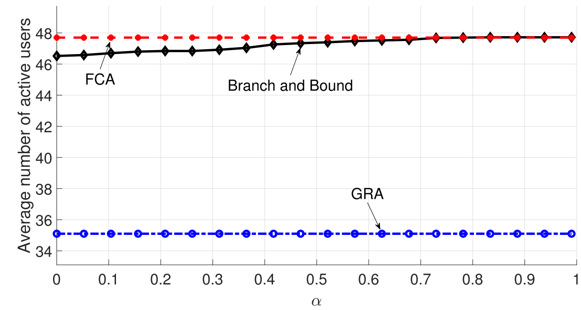
The next set of simulations are used to investigate the effect of the weighting factor α on the performance of the algorithms under study, rather than just the value of the optimization function. Similar to the previous setup, a total of 50 experiments are generated. The average objective function value, the average total achieved throughput, the average throughput per user and the average number of active users at values of α ranging from 0 to 1. The results are displayed in Figs. 5a to 5d. It is important to note that the solutions obtained through the proposed heuristic algorithms are insensitive to the value of α, because the value of α is neither involved nor affects the allocation algorithm in both. This can be seen clearly from Figs. 5a to 5d. Even though the value of the objective function itself changes, for the proposed heuristic algorithms FCA and GRA, with α, the numbers of the active users, the average total throughput and the average throughput per user do not change with α.

As depicted in Fig. 5a, the performance of the FCA algorithm steadily improves compared to that of the GRA algorithm, and it has a near-optimal behavior at $\alpha = 1$. The steady decline in the overall value of the objective function is due to the fact that each sub-channel allocation has a weight of α, while the throughput achieved by the allocated sub-channel contributes with a weight of $(1-\alpha)R_{ij}/\bar{R}$. It is understood that the value of the objective function itself is of marginal interest compared to the performance indicators, namely; number of active SUs and the total throughput, resulting from the solution itself.

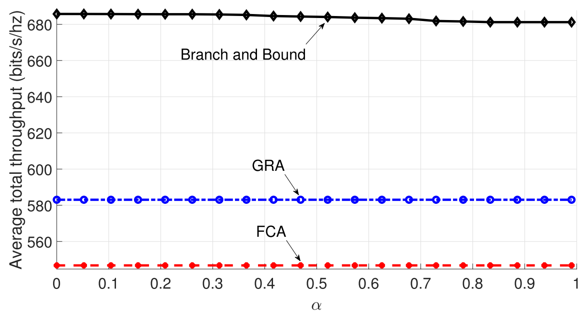
Fig. 5b shows the average number of active users resulting in from each of the three algorithms at different values of α. It can be seen that the FCA algorithm results in the highest average number of active users. It can be seen also that the average number of active users achieved by the branch and bound algorithm increases as the value of the weighting factor α increases. This is expected because the higher the value of α, the more the optimization functions favors the objective of increasing the number of active users. It can be also noted that using the GRA results in the lowest number of active users. In the example, an average of only 35 out of



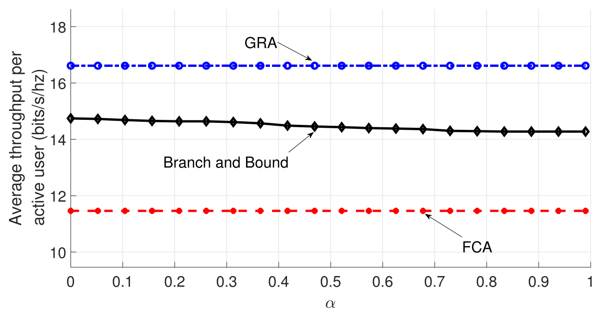
(a) The average value of the objective function.



(b) The average number of active users.



(c) The average total throughput.



(d) The average throughput per active user.

FIGURE 5. The average effect of α, over 50 experiments, on the performance of the allocation algorithms.

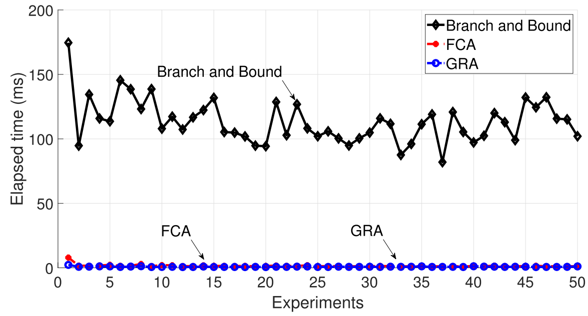
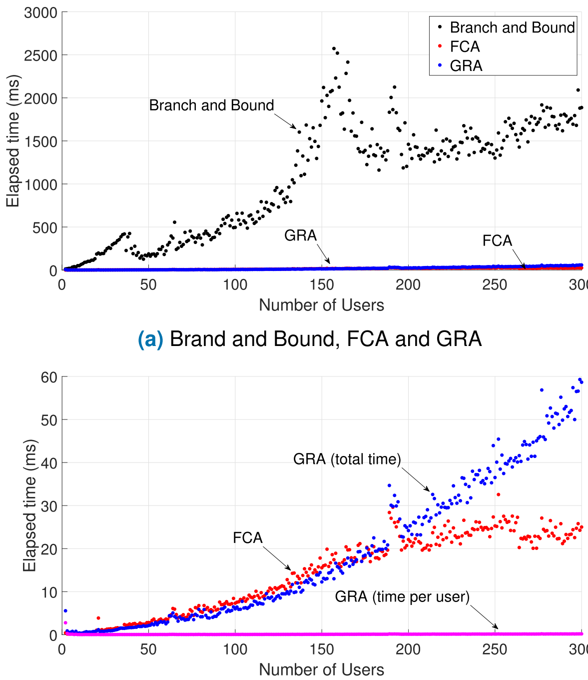


FIGURE 6. Comparison of the elapsed time by the three algorithms for 50 experiments.



(a) Brand and Bound, FCA and GRA

(b) FCA and GRA

FIGURE 7. Comparison of the computation time versus the number of users, N.

the total $N = 50$ users are allocated feasible resources. This is again because of the greedy approach that each SU uses the most rewarding resource maximizing its own throughput irrespective the other users. So, in this setup, while almost all of the $M = 48$ resources are fairly allocated to SUs when the FCA algorithm is used, only ~ 35 of them are allocated when the GRA is used.

The average total achievable throughput is assessed in Fig. 5c. It can be seen that the branch and bound algorithm results in the highest total throughout, followed by the GRA algorithm. It can be observed that while the total throughput resulting from the FCA algorithm is about 80% of the optimal value achieved by the branch and bound algorithm, the results of the GRA algorithm is approximately 85% of such optimum value. It can be seen also that the total throughput achieved by the branch and bound algorithm decreases for higher values of α as expected.

TABLE 1. Summary of the comparison points between the proposed heuristic algorithms.

Point of Comparison	FCA	GRA
Average rate per active user	Capable of achieving high average rate per active user due to the greedy allocation approach	Low average rate per active user due to the cautious allocation approach
Fairness	Maintains nearly equal chances among all secondary users	Prioritizes users who can initiate sessions with high transmission rates
Time complexity	$O(N^2M)$	$O(N^2M)$
Execution mode	Centralized	Supports both centralized and distributed modes
Applications	Suitable to address-centric networks whose priority is to provide service to all users while achieving minimum requirements	Suitable to data-centric networks whose nodes collaborate on achieving a certain task; however, the identity of the nodes is irrelevant such as some applications of wireless sensor networks.

Finally, Fig. 5d shows the average throughput per active user. As seen, the solution obtained by the GRA algorithm has the highest average rate per user. Combined with the result of a lower number of active users, this is considered an expected result due to the GRA's greedy allocation approach. It can be concluded from these observations that while the GRA algorithm is in general sub-optimal compared to the branch and bound algorithm, it is recommended for use when not only the total throughput but also the throughout per active user are sought for maximization.

Although the proposed FCA and GRA algorithms result in sub-optimal solutions, their main advantage is in their reduced complexity and fast processing time compared to the branch and bound algorithm. The complexities of the FCA and the GRA algorithms are both $O(N^2M)$ in the worst case. This is an immense speedup compared to branch and bound whose complexity is exponential in both N and M. In order to show such low complexity advantage, the three algorithms are compared to each other from the time complexity point of view, and results are shown in Figs. 6, 7a and 7b.

In these figures, a total of 50 experiments are conducted for a network setup of $N = 150$ secondary users and $M = 100$ vacant frequency sub-channels. The times needed by each of the three algorithms to obtain the solution are recorded and are shown in Fig. 6. It can be seen that for any of the conducted experiments, the proposed algorithms requires significantly shorter time to reach the solution

Fig. 7a shows the time taken by each algorithm to reach the solution as the number of SUs, N, increases. The number of sub-channels is assumed fixed irrespective the number of SUs. It can be observed that the increase in the time needed by the proposed algorithms is negligible compared to that needed by the branch and bound algorithm. Fig. 7b emphasizes on the time needed by the FCA and the GRA algorithms in order to clearly compare them. It can be observed that the FCA algorithm requires less computational time that the total time needed by all users to perform the GRA algorithm. However, taking into consideration that the GRA is executed in a distributed manner, with all users processing simultaneously, it can be observed that the time needed by each user to run the GRA algorithm is nearly constant irrespective the number of users. This is a considerable advantage for the GRA algorithm over both the FCA and the branch and bound algorithms. A comprehensive comparison between the FCA algorithm and the GRA algorithm is shown in Table 1

VII. CONCLUSION

In this paper, a new formulation of the dynamic spectrum allocation problem in cognitive radio networks was proposed. The new formulation defines a weighted objective function that takes into consideration the maximization of both the total throughout and the number of users allocated resources to achieve fairness among users. The problem is formulated as a binary linear program that can be solved optimally using the branch and bound algorithm with a linear programming solver at its core. Furthermore, two sub-optimal heuristic algorithms are proposed, the fair channel allocation (FCA) algorithm and the greedy rate allocation (GRA) algorithm. The proposed algorithms have significantly shorter computation time compared to the branch and bound algorithm. The resource allocations are guaranteed to be collision-free and they also satisfy the QoS requirements of the secondary users. Simulations were conducted, and it was shown that the new formulation achieves balance between both its objectives. It was shown also that the proposed heuristic solutions can achieve near optimal solutions with in significantly shorter times. Moreover, it was shown that the proposed algorithms are sub-optimal regarding the combined objective functions, the FCA algorithm always results in a higher number of active users, and the GRA results in a higher average achievable throughput per user when compared to the time consuming branch and bound algorithm.

REFERENCES

[1] *Spectrum Policy Task Force Report*, Federal Communication Commission, Atlanta, GA, USA, vol. 2155, Nov. 2002.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, pp. 2127–2159, Sep. 2006.

[4] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[5] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1277–1304, 2nd Quart., 2016.

[6] J. I. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio dissertation," *Diss. Royal Inst. Technol. Sweden*, vol. 294, no. 3, pp. 66–73, 2000.

[7] J. Mitola and G. Q. Maguire, Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Apr. 1999.

[8] J. Mitola, "The software radio architecture," *IEEE Commun. Mag.*, vol. 33, no. 5, pp. 26–38, May 1995.

[9] E. Z. Tragos, S. Zeadally, A. G. Fragkiadakis, and V. A. Siris, "Spectrum assignment in cognitive radio networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1108–1135, 3rd Quart., 2013.

[10] G. I. Tsiropoulos, O. A. Dobre, M. H. Ahmed, and K. E. Baddour, "Radio resource allocation techniques for efficient spectrum access in cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 824–847, 1st Quart., 2014.

[11] M. E. Tanab and W. Hamouda, "Resource allocation for underlay cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1249–1276, 2nd Quart., 2017.

[12] M. V. Nguyen, T. Q. Duong, C. S. Hong, S. Lee, and Y. Zhang, "Optimal and sub-optimal resource allocation in multi-hop cognitive radio networks with primary user outage constraint," *IET Netw.*, vol. 1, no. 2, p. 47, 2012.

[13] H. A. Bany Salameh, "Throughput-oriented channel assignment for opportunistic spectrum access networks," *Math. Comput. Model.*, vol. 53, nos. 11–12, pp. 2108–2118, Jun. 2011.

[14] J. Tang, R. Hincapie, G. Xue, W. Zhang, and R. Bustamante, "Fair bandwidth allocation in wireless mesh networks with cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1487–1496, Mar. 2010.

[15] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2282–2292, Oct. 2019.

[16] L. Akter and B. Natarajan, "Distributed approach for power and rate allocation to secondary users in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1526–1538, May 2011.

[17] C. Peng, H. Zheng, and B. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Netw. Appl.*, vol. 11, no. 4, pp. 555–576, May 2006.

[18] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 766–777, Jun. 2009.

[19] H.-P. Shiang and M. van der Schaar, "Distributed resource management in multihop cognitive radio networks for delay-sensitive transmission," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 941–953, Feb. 2009.

[20] J. Tang, S. Misra, and G. Xue, "Joint spectrum allocation and scheduling for fair spectrum sharing in cognitive radio wireless networks," *Comput. Netw. J.*, vol. 52, no. 11, pp. 2148–2158, Aug. 2008.

[21] A. Ahmed, G. Amel, S. Anis, and M. Abdellatif, "Resource allocation for multi-user cognitive radio systems using multi-agent Q-learning," *Proc. Comput. Sci.*, vol. 10, pp. 46–53, Jun. 2012.

[22] Z. Zhao, Z. Peng, S. Zheng, and J. Shang, "Cognitive radio spectrum allocation using evolutionary algorithms," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4421–4425, Sep. 2009.

[23] Z.-H. Wei and B.-J. Hu, "A fair multi-channel assignment algorithm with practical implementation in distributed cognitive radio networks," *IEEE Access*, vol. 6, pp. 14255–14267, 2018.

[24] R. Li and P. Zhu, "Spectrum allocation strategies based on QoS in cognitive vehicle networks," *IEEE Access*, vol. 8, pp. 99922–99933, 2020.

[25] R. C. Kizilirmak and H. K. Bizaki, *Towards 5G Wireless Networks—A Physical Layer Perspective: Non-Orthogonal Multiple Access (NOMA) for 5G Networks*. London, U.K.: IntechOpen, 2016. [Online]. Available: https://www.intechopen.com/chapters/52822

[26] X. Liu and X. Y. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5379–5388, Aug. 2020.

[27] X. Liu, X. B. Zhai, W. Lu, and C. Wu, "QoS-guarantee resource allocation for multibeam satellite industrial Internet of Things with NOMA," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2052–2061, Mar. 2021.

[28] K. Zheng, X.-Y. Liu, X. Liu, and Y. Zhu, "Hybrid overlay-underlay cognitive radio networks with energy harvesting," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4669–4682, Jul. 2019.

[29] X. Liu and M. Jia, "Intelligent spectrum resource allocation based on joint optimization in heterogeneous cognitive radio," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 1, pp. 5–12, Feb. 2020.

[30] N. S. Moayedian, S. Salehi, and M. Khabbazian, "Fair resource allocation in cooperative cognitive radio iot networks," *IEEE Access*, vol. 8, pp. 191067–191079, 2020.

[31] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. London, U.K.: Pearson, 2021.

AHMED ABOULFOTOUH received the B.Sc. degree (Hons.) in communications and information engineering from the University of Science and Technology, Zewail City, Egypt, in 2020. He is currently pursuing the M.Sc. degree in artificial intelligence. After graduating, he worked as a Teaching Assistant with the Communications and Information Engineering Program, Zewail City. His research interests include applying artificial intelligence and machine learning in communication systems.

SAMY S. SOLIMAN (Member, IEEE) received the B.Sc. and M.Sc. (Hons.) degrees from the Electronics and Electrical Communications Engineering Department, Faculty of Engineering, Cairo University, Egypt, in 2007 and 2009, respectively, and the Ph.D. degree from the Electrical and Computer Engineering Department, University of Alberta, Canada, in 2014.

In 2009, he joined the AITF/iCORE Wireless Communications Laboratory (iWCL) as a Research Assistant and worked as a Post-Secondary Teaching Assistant at the Electrical and Computer Engineering Department, University of Alberta. Then, he held a Postdoctoral Research Fellowship at the University of British Columbia, Canada, in 2014. In 2015, he was appointed as an Assistant Professor at the Electronics and Electrical Communications Department, Cairo University. Since 2016, he has been with the University of Science and Technology, Zewail City, Egypt. His research interests include 5G communication systems, wireless sensor networks and IoT applications, wireless cooperative networks, optical and ultra-wide bandwidth wireless systems, and the use of machine intelligence in future communication systems. In 2008, he was awarded the Cairo University Prize of Engineer Nabil El-Gebaly as the Highest Standing Student in Telecommunications and the Cairo University Prize of Engineer Reda Hamza as the Highest Standing Student in Electronics. He was also awarded the GSA Professional Development Award from the University of Alberta, in 2012 and 2013.