IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 28, 2022, accepted May 28, 2022, date of publication June 8, 2022, date of current version June 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3180753

Optimal Resource Allocation for GAA Users in Spectrum Access System Using Q-Learning Algorithm

WASEEM ABBASS[1], RIAZ HUSSAIN[1], JAROSLAV FRNDA[2,3], (Senior Member, IEEE), IRFAN LATIF KHAN[1], MUHAMMAD AWAIS JAVED[1], (Senior Member, IEEE), AND SHAHZAD A. MALIK[1]

[1]Department of Electrical and Computer Engineering, COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
[2]Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communication, University of Žilina, 01026 Zilina, Slovakia
[3]Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, 70800 Ostrava, Czech Republic

Corresponding authors: Waseem Abbass (waseem.abbas@comsats.edu.pk) and Muhammad Awais Javed (awais.javed@comsats.edu.pk)

This work was supported in part by the Institutional Research of the Faculty of Operation and Economics of Transport and Communications, University of Zilina, under Grant 2/KE/2021; and in part by the Ministry of Education, Youth and Sports of the Czech Republic conducted by the VSB-Technical University of Ostrava, Czechia, under Grant SP2022/5.

ABSTRACT Spectrum access system (SAS) is a three-tier layered spectrum sharing architecture proposed by the Federal Communications Commission (FCC) for Citizens Broadband Radio Service (CBRS) 3.5 GHz band. The available 150 MHz spectrum is dynamically shared among Incumbent Access (IA), Primary Access Licensees (PAL) and General Authorized Access (GAA) users. IA users are the highest priority federal military users, PAL users are the licensed users and the GAA users are the least priority unlicensed users. In this scenario, PAL operators are willing to give access to their idle spectrum to GAA users to generate extra revenue. SAS will ensure to protect IA users and PAL users from interference caused by lower-tier users. It is the responsibility of SAS to allocate resources to GAA users but the method to do so is left open. In this article, a novel auction algorithm based on Q-learning for dynamic spectrum access (SAS-QLA) is proposed. In SAS-QLA, multiple GAA users dynamically and intelligently bid using Q-learning to access PAL reserved idle channels. SAS will decide to allocate the channels to GAA users with maximum bidding offers. GAA users have their own quality of service (QoS) demands i.e., transmission rate, packet loss, bidding efficiency, and maintain the preference of available PAL reserved idle channels based on Q-learning considering the available QoS. The proposed scenario is also modeled as a knapsack NP-hard problem and solved using dynamic programming and distributed relaxation method. Numerical results demonstrate the effectiveness of the SAS-QLA algorithm in improving the bidding efficiency, maximizing the data rate per unit cost and spectrum utilization.

INDEX TERMS Auction algorithm, CBRS-SAS, GAA bidding, Q-learning.

I. INTRODUCTION

Over the decade, the demand for internet-based applications, the internet of things (IoT), and machine-to-machine communications is increasing exponentially. To meet the requirements, flexible and rapid access to the radio spectrum is needed. Fifth-generation (5G) cellular communication systems provide efficient data connectivity with high speed [1].

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

The radio spectrum is a limited resource and the efficient use of the radio spectrum is the main concern of cellular network operators, industry, and government. The federal communications commission (FCC) proposed to share the radio spectrum held by the government with commercialized access in the USA [2].

The citizens broadband radio service band (CBRS) 3.5 GHz (3550 MHz – 3700 MHz) band was proposed to be shared with licensed and unlicensed users [3]. A centralized framework based on the spectrum access system (SAS) was

introduced to share the radio spectrum held by the government with licensed and unlicensed users. Moreover, three types of users are introduced in the SAS framework [4]. Incumbent access (IA) users are the highest priority users such as navy radars and fixed earth stations. Primary access licensee (PAL) users are the second priority, commercial license holders. PAL users pay for the license to get a guaranteed quality of service (QoS). General authorized access (GAA) users were introduced as unlicensed users that opportunistically access the available radio spectrum. In SAS based CBRS framework 150 MHz band is distributed between PAL and GAA users. 70 MHz band (3550 MHz – 3620 MHz) is dedicated to PAL users and the 80 MHz (3620 MHz-3700MHz) band is reserved for unlicensed GAA users. PAL operators can acquire the radio spectrum through competitive bidding for a region for up to three years. Moreover, there can be a maximum of four PAL operators and each PAL operator can get a maximum of four radio bands of 10 MHz each from 70 MHz dedicated radio spectrum for PAL users. The 80 MHz (3620 MHz - 3700 MHz) radio spectrum band is reserved for GAA users and can be used for unlicensed services. QoS is not guaranteed to GAA users but the users can opportunistically access the PAL reserved radio spectrum to get the required QoS if the quality of PAL users is not compromised. Furthermore, PAL users cannot use GAA reserved radio spectrum [5]. SAS is the central entity in the CBRS-SAS framework to authorize spectrum sharing between PAL and GAA users. It is also the responsibility of SAS to protect IA users from interference caused by PAL and GAA users [6].

The SAS based CBRS architecture is different from conventional cellular networks and a lot of research has been done in the domains of spectrum access, adaptive resource allocation for multiple tiers of users, spectrum pricing, maintaining QoS, protocols, and standards, operational security and interference management [7], [8]. In the recent three years, most of the work focuses on spectrum sharing and spectrum trading between PAL and GAA users. Spectrum trading is an efficient method for PAL operators to sell the licensed PAL reserved idle radio spectrum to GAA users [9]. FCC proposed a set of rules to dynamically share the available spectrum among IU, PAL, and GAA users [10]. In SAS based CBRS framework spectrum trading is allowed where idle spectrum held by PAL operators can be leased to GAA users for financial gains and to use the available spectrum efficiently while satisfying the rules of spectrum sharing. However, the methods to trade and allocate the spectrum to GAA users are left open in the current release of CBRS alliance technical specifications [11].

In this paper, a novel auction algorithm SAS-QLA based on Q-learning for dynamic spectrum access in the SAS-based CBRS framework is proposed. The proposed algorithm aims to improve spectrum access for GAA users. In the SAS-QLA algorithm, reinforcement learning is used that allows the GAA users to improve their bidding strategy based on their past payoff knowledge to compete for the available idle channel. The GAA users can consider factors like transmission opportunity, environmental factors, and current state to make their bid independently. Meanwhile, PAL operators ensure to generate profits and lease radio spectrum to the GAA users with a maximum bid. Moreover, a mathematical model to allocate the available idle PAL reserved channel to the best bidder i.e., GAA user, is proposed that is based on the classical Knapsack problem. The formulated NP-hard problem is solved using the dynamic programming and distributed relaxation method.

In the defined scenario of SAS-based architecture where GAA users have to select a channel from the list of available idle PAL reserved channels with the best data rate and minimum cost. This problem can be modeled as a knapsack problem as the GAA users have the only option to accept the channel for transmission or reject the channel. In the scenario of a greedy knapsack problem, a portion of the item can also be sacked. In our case, it is not feasible as GAA users cannot take a portion of the channel. Therefore, the problem is modeled as a knapsack 0-1 problem because each channel has an individual weight and a value that will be used by GAA users to decide whether to accept or reject the channel. There are some well-known algorithms available to solve the knapsack problems i.e., greedy algorithms [12], dynamic programming [13], branch and bound [14], e-t-c. A dynamic programming algorithm is used in the scenario where a problem can be segmented into sub-problems. To solve a problem, a dynamic programming algorithm solves individual sub-problems to get a solution for a particular sub-problem. In the end, it joins all the solutions to get an optimal solution. Whereas genetic algorithms find the optimal solution from a list of available solutions. Genetic algorithms are suitable when there is already a solution set is available. Branch and bound algorithms are suitable for combinatorial and discrete optimization problems. The distributed relaxation method is a well-known assignment algorithm that works like an auction. The distributed relaxation method allows the persons to bid simultaneously for multiple items with an option to raise the bids. When all the bids are in, the item is awarded to the highest bidding person. In our scenario, the distributed relaxation method is used, as it allows the GAA users to bid for the multiple PAL reserved channels in a spectrum pool. Once the SAS receives the bids from all participating GAA users then the PAL channel will be allocated to the GAA user with the highest bid. The distributed relaxation method is competitive and suitable for large problems as compared to the existing method. Hence, we select the dynamic programming and distributed relaxation method for comparison with our proposed algorithm because these two algorithms give the best solution in comparison with the existing methods and algorithms and our numerical results show that the SAS-QLA algorithm outperforms these two algorithms if SAS gives preference to GAA users.

Above all, the paper provides the following contributions.

1) An algorithm, SAS-QLA, based on Q-learning that uses reinforcement learning is proposed to enhance the

bidding strategy of GAA users to access PAL reserved radio channels.
2) We modeled the proposed scenario using the Knapsack problem and solved the problem using the dynamic programming method and distributed relaxation method.
3) A detailed comparison of the three algorithms is given that shows that the proposed SAS-QLA algorithm achieves guaranteed QoS for GAA users while satisfying rules proposed by FCC and also maximize the PAL operator's profit.

The remainder of the paper is organized as follows. Section II summarizes the related work. Section III presents the network model, proposed GAA bidding scenario, and detailed problem formulation. Section IV gives the detailed problem solution with the proposed SAS-QLA algorithm, formulation of the Knapsack 0-1 problem, and its solution i.e., dynamic programming method and distributed relaxation method. Furthermore, Results and performance evaluation are provided to analyze the performance of bidding and allocation algorithms ins section V. Finally, we conclude our work in Section VI.

II. RELATED WORK

In the last decade, the problems related to dynamic spectrum sharing have been discussed and investigated in many research efforts. The key focus of these research efforts was to design adaptive mechanisms and algorithms to allocate radio resources as well as admission control, interference management, spectrum pricing, and end users' requirements of quality of service [15]–[17]. In 3GPP release 15 [18], the concept of physical resource sharing between mobile network operators was proposed.

In this section, we summarize related work on spectrum trading and its impact on the economical model of dynamic spectrum access in 5G and beyond. Moreover, recent trends and techniques for allocation of radio resources in spectrum access system (SAS) based architecture proposed for citizens broadband radio service (CBRS) 3.5 GHz band are also discussed.

Generally, Spectrum trading methods are adopted by licensed mobile network operators (MNOs) to enhance their financial gains and spectrum utility by leasing their unused spectrum to unlicensed users with proper marketing strategies [19]–[21]. There are a lot of surveys published in the domain of spectrum trading, marketing strategies, spectrum pricing mechanisms, controlling dynamic network traffic, power allocation for radio resource management and spectrum hand-off, etc. [22]–[25] Authors in [26]–[29] study the diversity of existence of licensed and unlicensed users in a spectrum band using different strategies including game-theory based spectrum allocation and spectrum leasing to secondary users based on different pricing strategies. Authors in [30] considered heterogeneity using a game-theory-based approach for the coexistence of wireless infrastructure providers, end-user devices, and virtual network operators. Moreover, in [31] authors proposed an algorithm to investigate the uniqueness of equilibrium points using an iterative three-layer game model. As a game-theory-based approach is efficient in case of static physical scenarios so to meet the quality of service (QoS) requirements of licensed and unlicensed users in a real-time environment is still a major challenge.

To address the issues of generating additional profit for cellular and satellite network operators, researchers are paying much attention to share the idle spectrum held by network operators with cognitive users [32]. In comparison with classical wireless networks, the radio spectrum allocation problem in satellite-terrestrial networks is different in the context of the presence of high priority satellite users with their individual bandwidth requirements, the presence of heterogeneous cognitive users, and their unpredictable demands [33]–[35]. It is important to mention that the work proposed in [33]–[35] is based on the assumption that idle radio spectrum band is available continuously. However, the licensed users dynamically keep joining and leaving the spectrum. Hence, the assumption of continuous availability of radio spectrum is not practical as the approaches are unable to solve the discrete bandwidth strategies.

Authors in [36] improved the heterogeneous spectrum sensing using big data-based intelligent spectrum sensing technique. The authors discussed the machine learning techniques i.e., k-means clustering, distributed learning, extreme learning, reinforcement learning, kernel-based learning, deep learning, and transfer learning. The authors proposed a machine learning-based big spectrum data clustering mechanism to detect vast accessible spectrum resources for heterogeneous spectrum communications. In [37] authors proposed a solution for finding idle channels with guaranteed spectrum sensing performance with higher detection probability, high spectrum access probability, and lower error probability in cognitive radio networks. The authors proposed the multi-slot double threshold spectrum sensing with Bayesian fusion based on reinforcement learning to sense big industrial spectrum data. The authors modeled the channel prediction and selection using reinforcement learning based on Thompson sampling which results in efficiently finding the required idle channels by sensing the big spectrum data accurately. The problem of the age of information (AoI)- aware radio resource management for manhattan grid vehicle-to-vehicle network is investigated in [38]. The authors proposed a proactive algorithm that includes decentralized online testing at the vehicle user equipment (VUE) pairs and a centralized offline training at roadside units (RSUs) and modeled the stochastic decision-making procedure as a discrete-time single-agent Markov decision process (MDP). The proposed algorithm shows significant performance gains over the existing baseline algorithms.

Authors in [39] proposed the spectrum allocation to unlicensed users using reinforcement learning and game theory to achieve Nash equilibrium and to minimize the interference of licensed users caused by unlicensed users. However, the pricing and auction strategies to allocate available spectrum

to unlicensed users are still left open. In [40], the authors used a carrier sensing mechanism to formulate the super-radio formation algorithm to investigate the co-existence of users in shared spectrum access for the 3.5 GHz CBRS band. A channel allocation method to share the radio resources among different categories of users is proposed by authors in [41]. The proposed algorithm achieves the minimum throughput requirement for each category by assigning resource blocks among all stakeholders. The implementation and evaluation of CBRS-based SAS architecture, hardware experiments, and field trials are discussed in [42] and [43]. In recent research [44], the authors proposed the privacy techniques to protect high priority incumbent users from low priority licensed and unlicensed users. The authors used the concept of beamforming with constraints of limiting transmit power to alleviate interference and increase the detection probability of incumbent users. The resource allocation techniques and auction or pricing strategies are not considered. The authors in [45] considered the privacy of users in the SAS framework. The authors use the concept of blockchain technology to get the details of GAA users including their physical location, identity, and spectrum usage. The cryptographic methods were applied to protect the information. Authors in [46] proposed to use the concept of decentralization for the SAS-CBRS-based framework. However, FCC proposed the centralized SAS in the CBRS band. Moreover, in the case of decentralized SAS, the privacy of incumbent users can be compromised. So, this concept is practically not feasible.

Radio spectrum allocation to heterogeneous wireless technologies i.e., wireless fidelity (Wi-fi), Worldwide Interoperability for Microwave Access (WiMAX) networks, and cellular networks are considered in [47]. The authors applied a genetic algorithm and a Hungarian algorithm to solve the problem that was modeled as a multi-objective optimization problem. The concept of using the genetic algorithm to find the optimal solution for the stated problem is a good idea, but it takes a lot of time to find the optimal solution. Hence, the efficiency of the system is not taken into account. In [48], we proposed an improved Hungarian method to improve the computational efficiency for allocating idle PAL reserved channels to GAA users. We proposed the concept to find the optimal value of achieving higher data rates at lower costs. However, the revenue of PAL operators was not considered. Moreover, PAL operators are the main stakeholders of the CBRS-SAS architecture so it is important to consider the profit of PAL operators while considering the quality-of-service requirements of incumbent access, primary access licensee, and general authorized access users.

The recent research papers discussed in Section II show the tremendous efforts of the researchers towards radio network selection and dynamic allocation of radio spectrum in multi-tier environment; However, a need of proper mechanisms is still required to consider practical scenarios for allocation of radio channels to GAA users in presence of high priority users i.e., incumbent access users and primary access licensees while strictly following the rules proposed by FCC for allocation of radio channels in a multi-tier environment. Moreover, a pricing strategy is also required to enhance the spectrum utility and increase the revenue of commercial PAL operators.

III. SYSTEM MODEL

A. NETWORK SCENARIO

The spectrum access system (SAS) is proposed to be a central entity to manage the 3.5 GHz (3550 MH - 3700 MHZ) citizens broadband radio service (CBRS) band usage. The SAS-based CBRS framework is shown in Figure 1. The SAS is the central entity, which authorizes the tiered users in the CBRS band i.e., incumbent access (IA) users, primary access licensee (PAL) users, and general authorized access (GAA) users. SAS is also responsible to maintain the required quality of service (QoS) of IA and PAL users but does not protect the GAA users from the interference of guaranteed access to the spectrum. As the CBRS band was initially reserved for the US navy radar system and is now available for commercial usage. So, IA users can access the whole 150MHz (3550 MHz - 3700 MHz) spectrum. 70 MHz (3550 MHz - 3620 MHz) spectrum is reserved for PAL users. This 70 MHz radio spectrum band is divided into seven bands of 10 MHz each and assigned to a maximum of four PAL operators through competitive bidding.

The CBRS-SAS system consists of the following network elements.

- Spectrum Access System (SAS)
- Environmental Sensing Capability (ESC) sensor
- Citizen Broadband Radio Service Device (CBSD)
- Domain Proxy (DP)
- Network Management System (NMS);
- FCC external databases
- End devices

SAS is the main authorizing entity to allocate channels to IA, PAL, and GAA users and involves in radio spectrum management and distribution. It is also the responsibility of SAS to protect IA users from lower-tier users and also protect PAL users from interference caused by GAA users to ensure QoS is provided to PAL users. SAS also manages the transmission power to citizen broadband radio service devices CBSD. The CBSD is an eNodeB that supports transmissions in the 3.5 GHz radio spectrum band and supports protocols defined by the federal communications commission FCC in its technical specification [49]. SAS uses external databases to store the information of all users, CBSDS, transmit power, and the information of radio channels that are occupied or still available for allocation. The detection of navy radars and IA users is carried out using environmental capability ESC sensors. SAS uses the information provided by ESC sensors to vacate the radio channels for use of IA users. SAS uses domain proxy (DP) to manage CBSDs aggregation and proxy functions to manage large-scale networks. DP can also be integrated with an element management system (EMS) or network management system (NMS). It is an optional element to scale large networks. In Figure 1 end-user devices (EUD)

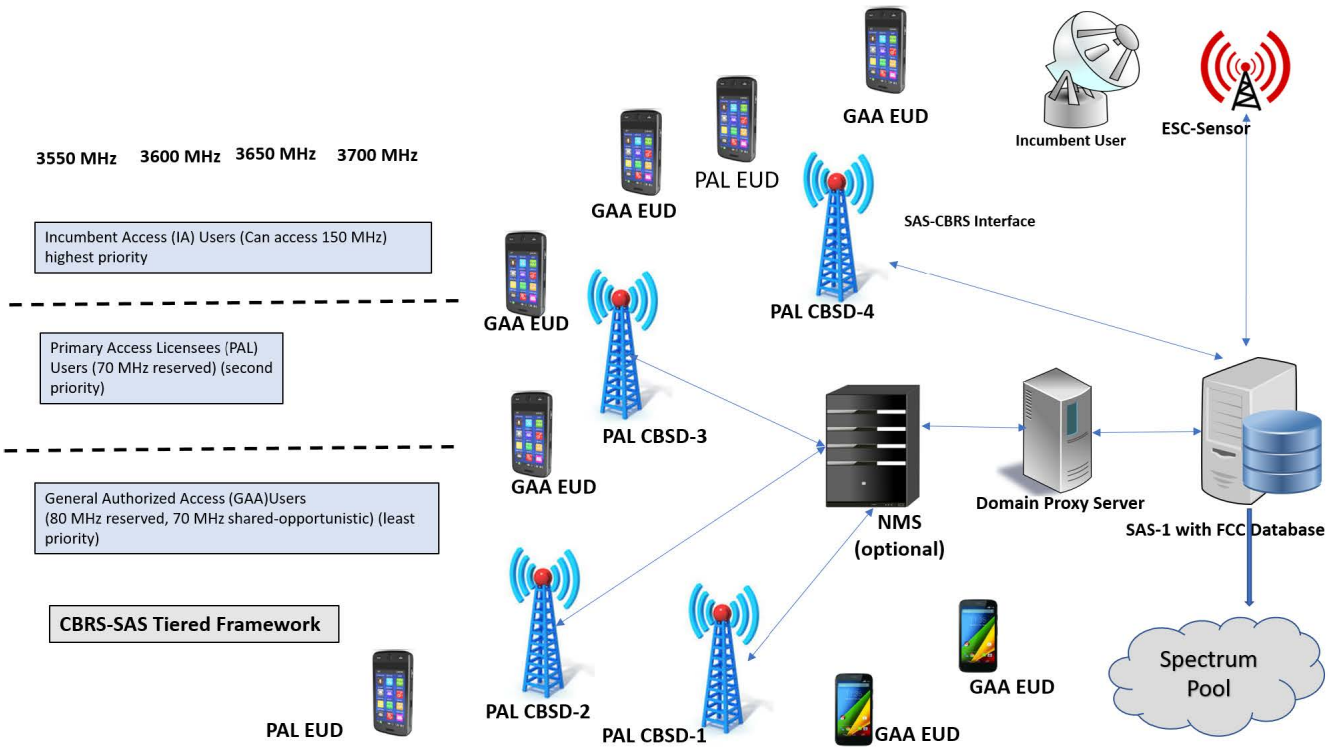

FIGURE 1. CBRS-SAS architecture.

are the electronic intelligent mobile devices that support frequency transmission in the 3.5 GHz radio spectrum band.

In this paper, a scenario is considered where multiple GAA users are seeking opportunities to access the radio spectrum and bid for the available idle PAL reserved channel to get guaranteed QoS. The bidding strategy proposed for GAA users is based on reinforcement learning that also increases the revenue of PAL operators. The GAA users bid according to the future reward expected, current state, and QoS requirements. The radio channel assignment will also be implemented using dynamic programming i.e., a solution for the knapsack problem and distributed relaxation method.

Consider a CBRS-based SAS framework scenario, where there are multiple idle PAL reserved channels available and several GAA users are seeking the available opportunities to access idle channels. The availability of PAL reserved idle channels can be modeled as a two-state Markov chain as the PAL users are not active all the time and the joining and leaving network is discontinuous. The available channels are assumed to be perfectly orthogonal eliminating the chances of experiencing interference if some constraints for out-of-band emissions are applied.

Let the number of incumbent users be i. The vector of incumbent user's set is represented as:

$$i = \{i_1, i_2, \ldots\ldots, i_x\}$$

The set of PAL users is given by:

$$p = \{p_1, p_2, \ldots\ldots .p_y\}$$

and GAA user's set is written as

$$g = \{g_1, g_2, \ldots\ldots .g_z\}$$

Let d represents the set of frequency spectrum available in a census tract, j denotes the set of PAL reserved channels in a census tract such that $j \subseteq d$ and h represents the set of GAA reserved channels. and PAL reserved channels set is represented as:

$$j = \{j_1, j_2, \ldots\ldots, j_l\}$$

The set of GAA reserved channels is shown as:

$$h = \{j_{l+1}, j_{l+2}, \ldots\ldots, j_n\}$$

and the set of all channels in a census tract is given as

$$d = \{j_1, j_2, \ldots, j_l, j_{l+1}, j_{l+2}, \ldots\ldots .j_n\}$$

The vector j represents the vector containing PAL reserved channels and the value of each element of vector j can be 0 or 1 i.e., $j \in \{0,1\}$. To see, whether the PAL users are active in any particular channel from the vector j can be implemented as a Poisson process of switch 2 state S i.e., idle or busy. The state S of PAL users is represented as $S = j * a_g^t$, where a_g^t shows whether the transmission opportunity to GAA user g is available or not at time t. The value of a_g^t is 0 or 1, $a_g^t \in \{0,1\}$. State 0 shows that the channel is idle and state 1 represents the channel is occupied. We assume GAA users can access the available Idle channel from vector j and transmit at constant power. Moreover, we assume that GAA

TABLE 1. Notations Used.

Set	Description
i	Vector of incumbent users
p	Vector of PAL users in PAL reserved channels
g	Set of GAA users
j	Set of PAL reserved frequency channels
h	Set of GAA reserved frequency channels
d	Set of all available frequency channels
a^t	Transmission opportunity, available or busy
S	Set of possible state space

users move slowly such that their channel conditions variate slowly. The notations used in problem formulation are listed in Table 1.

B. RADIO SPECTRUM BIDDING SCENARIO

Considering the fact that PAL users are not always active so PAL operators can take advantage to trade the radio channels with GAA users by using different marketing approaches. This scenario can be modeled as an auction or trading market, where PAL operators can sell the idle radio channels to GAA users to increase their revenue. The GAA users look for available opportunities and bid accordingly, while the PAL operators offer their available idle channels for bidding. The scenario creates a win-win business opportunity for both parties.

We made the following assumptions in this article.

1) GAA users can bid for one or more than one channel simultaneously and get the required QoS as committed by the PAL operator.
2) SAS can receive offers from all GAA users who applied for radio spectrum. SAS will select the GAA users on the basis of the maximum bid for each available channel.
3) The bids of GAA users are not shared with each other. Hence assuming a symmetric independent private value (SIPV) [50].
4) SAS keeps the information of bids offered by all GAA users as SAS has to decide on the basis of offered bids to allocate the radio channels to GAA users.
5) SAS allocates a common channel with GAA users to carry the information used in the SAS-QLA algorithm.

Moreover, SAS creates a pool of available PAL reserved channels that can be used for trading with GAA users. The spectrum pool of available PAL reserved channels of all four PAL operators is shown in Figure 2 The characteristics of PAL reserved channels of all operators may vary from each other and depends on channel fading and interference level. Random distribution of GAA users and complex spectrum environments leads to quality diversity. Due to this diversity, GAA users can make a rational selection from the spectrum pool as the prices of ideal channels are high. Furthermore, to maximize the profit of PAL operators, a proper pricing mechanism is required for SAS, which is based on GAA users bidding behaviors and QoS provided to GAA users.

C. Q-LEARNING MODEL

Q-learning is a model-free off-policy reinforcement learning used to find the next course of action based on the current state of the agent. The objective of the Q-learning model is to maximize the total reward and selects the action at random as it learns from the actions that lie outside the current policy therefore a policy is not required. In the scenario of using Q-learning, an agent that in our scenario is a GAA user; performs a task at a particular state to see the consequences of the actions in terms of the immediate reward or the penalty for performing the particular action in the current state. The value of the state is estimated for each action in the particular state. It will similarly try all actions in all states to learn the best action for a particular state based on the achieved long-term discounted reward. The optimal policy will be achieved when the GAA users gained the maximum expected discounted reward. Q-learning is an iterative process, the GAA users can exploit the reward based on the discount factor after performing an action for its particular state to initialize it from start. In Q-learning, a sequence of multiple stages is defined as incremental dynamic programming for the GAA user's experience. The GAA users can perform the following tasks in their nth stage:

- Observe the current state.
- Choose an action to perform for the current state.
- After the action it can observe the next state.
- It receives an immediate reward.
- Adjust the initial values using a learning factor.

The detailed formulation of the Q-learning process in the proposed SAS-QLA algorithm is discussed in the following section.

D. SAS-QLA FORMULATION

GAA users' main intention is to get guaranteed QoS at a reasonable price. GAA users can also opt for high-quality available channels with increased transmission capacity at a higher cost. Moreover, even with the same budget, GAA users can also select the channels based on differentiated services (DiffServ) with different service classes. Furthermore, GAA users can improve their bidding strategy in the SAS channel auction by using Q-learning based on reinforcement learning. GAA users also maintain a bidding vector that contains the best cost factor for each available idle PAL reserved channel. We define a function $F_{QSAS} = \zeta_g^{t+1}(S_g^t, \delta_g^t)$ based on Q-learning for GAA users that return the auction policy by getting observations as input.

To implement a Q-learning-based reinforcement learning algorithm a policy with a value function and reward function is required to make a decision strategy for channel allocation to GAA users. For each moment t, each GAA user g emits an action δ_g^t. In the next moment $t+1$ each GAA user receives a reward in response to its action given by ω_g^t and move to next state S_g^{t+1}. We consider a scenario, where each GAA user in CBRS-SAS architecture can make spectrum bidding policy individually. The SAS-QLA algorithm is modeled according

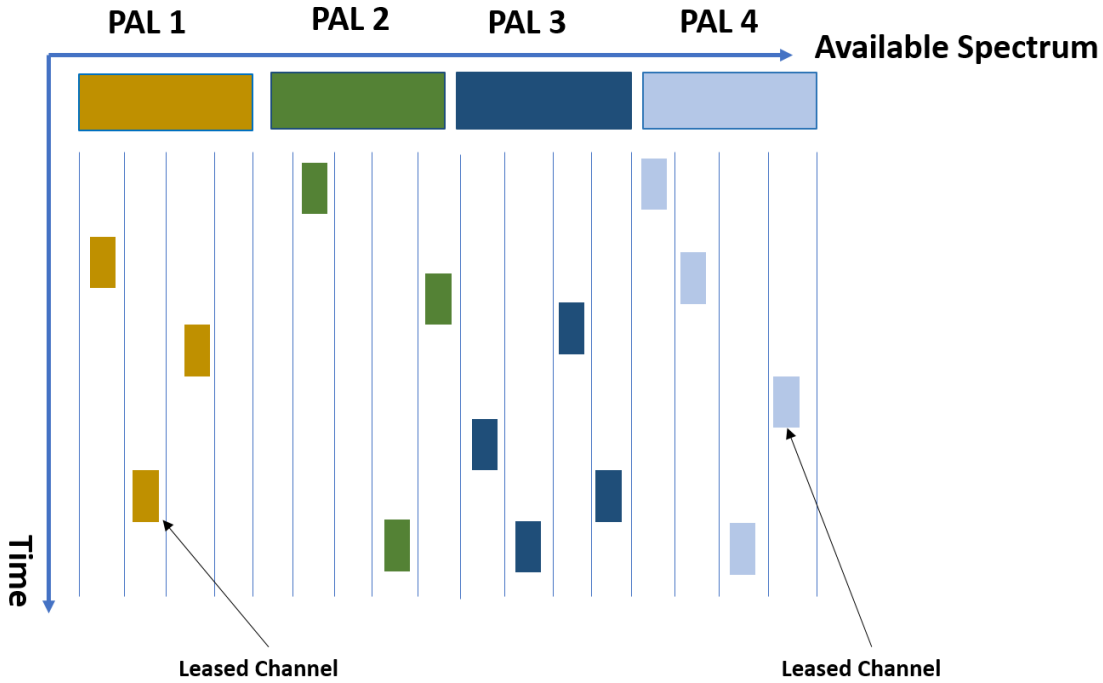

FIGURE 2. Spectrum Pool.

to the utility perceived for the current selection of action and the history of states visited to allocate the channels to GAA users. Hence, it is important to define states, actions, rewards, and learning policies for GAA users before applying the SAS-QLA algorithm.

1) STATES

In our defined scenario in section III-B, GAA user current occupied channel is defined as a state $S_g^t = \{\gamma_g^t\} \in S$. S represents the finite set of possible state space. Let's denote $S = \{S_k\}$, where k shows the number of states. $S_k = \{g_n\}$, where $g_n = \{0, 1\}$ and $n = 1, 2, \ldots, z$. State transition of GAA users from S_g^t to S_g^{t+1} can be determined by two stochastic events. First one is whether the channel is occupied and another is spectrum opportunity is not available. The GAA users will move to next state S_g^{t+1}, when one of these two events occur.

2) ACTIONS

SAS will perform an action when it has to assign the available PAL reserved channels to bidding GAA users. We define action of GAA users as $\delta_g^t = \{\alpha_g^t, P_g^t\}$, where α_g^t shows the preferred channel and P_g^t represents the bidding price. GAA users evaluate each action δ_g^t that is based on Q-learning function for every state $S_g^t \in S$. In result to this evaluation, an immediate reward ω_g^t is received at GAA users' end and the state of GAA users change from S_g^t to S_g^{t+1}.

3) REWARD FUNCTION

GAA users select the channel from available idle PAL reserved channel that gives the maximum reward that is received from the reward function $\omega_g^t\{S_g^t, \delta_g^t\}$ for each action δ_g^t at particular state S_g^t. So, maximizing the reward value for GAA users at each state with particular actions can be defined as

$$\omega_g^t = \sum_{j=1}^{l} \alpha_{g,j}^t . P_{g,j}^t \tag{1}$$

subject to the constraints:

$$\alpha_{g,j} \in \{0, 1\}, \sum_{g} \alpha_{g,j}^t \leq j$$

where j is the PAL reserved channel set available for lease to GAA users $g \in (1, 2, 3, \ldots, z)$ at time t. The reward ω_g^t for GAA users depends on the transmission capacity of each channel j by paying price $P_{g,j}^t$.

4) LEARNING POLICY

Learning policy deals with the mapping of history of visited states, utility received and probability of selected action δ_g^t into currently selected action.

For this purpose a Q-learning function $\zeta_g^{t+1}(S_g^t, \delta_g^t)$ for GAA users is

$$\begin{aligned}(S_g^t, \delta_g^t) = (1 - \eta_g)\zeta_g^t(S_g^t, \delta_g^t) \\ +\eta_g[(\omega_g + \sigma_g . max \zeta_g^t(S_g^{t+1}, \delta_g^{t+1})]\end{aligned} \tag{2}$$

The optimal policy $(\delta_g^{t+1})^*$ is defined as an expected sum of ω_g, that is discounted by σ_g^t. The Q-learning modifies the action value Q-learning function to obtain the true value using the optimal policy $(\delta_g^{t+1})^*$. i.e.,

$$(\delta_g^{t+1})^* = argmax \ \ \zeta_g^{t+1}(S_g^t, \delta_g^t) \tag{3}$$

Q-learning is an off-policy reinforcement learning algorithm that learns from the actions that are random and finds the best action to take for the given current state. The parameters used in the Q-learning process are the learning rate, discount factor, and maximum reward. The learning factor defined in our proposed SAS-QLA algorithm is η_g which varies from 0 to 1. This factor shows the learning ability of the Q-learning function. If it is set to 0, it means nothing is learned and Q-values are not updated as it is using exclusively the prior knowledge to decide. The high value of the learning factor shows that the learning process will be quick and the algorithm will converge speedily because it will ignore the old prior information and just considers the most recent information. In case of problems with deterministic solutions, the learning rate is set to 1. In stochastic problems just like in our scenario, the Q-learning function in SAS-QLA algorithm converges under technical conditions to use the prior knowledge that is why it is set to 0.1. The discount factor σ_g in the SAS-QLA algorithm also varies between 0 to 1. The discount factor is the deciding factor that helps the Q-learning function choose between future rewards and immediate rewards and determines the importance of future rewards. If a high value is set for the discount factor i.e., 1, then the future reward will be high but, in our scenario, it makes the base price set by the PAL users too high than the bidding price offered by GAA users, so it will be good to use the less value for the discount factor to converge the algorithm quickly and make the auction successful. The third important parameter for the Q-learning function is the maximum reward i.e., $\eta_g.\sigma_g.max\zeta_g^t(S_g^{t+1}, \delta_g^{t+1})$ defined in equation 2. It shows the maximum reward that can be achieved from state S_g^{t+1} weighted by the learning factor and the discount factor. The Q-learning is an iterative algorithm so initial conditions are assumed before the next update. The first reward can be used to reset the initial conditions.

Based on the defined optimal policy, each GAA user will make a positive preference decision. η_g is the learning rate that remains in the range of $0 < \eta_g < 1$. This parameter determines the Q-function updating speed. The reward for GAA users change too frequently if the value of η_g is close to 1. σ_g^t is the discount factor that is used to determine the current value i.e., $0 < \sigma_g^t < 1$ of future reward. If the value of discount factor σ_g^t approaches 1, then it shows that future interaction plays an important role to define total utility values. The Q-function $\zeta_g^{t+1}(S_g^t, \delta_g^t)$ takes two values to update its evaluation.

1) The projected value (ζ-value) of new state $\zeta_g^t(S_g^{t+1}, \delta_g^{t+1})$.
2) The instantaneous reinforcement value ω_g^t

IV. PROPOSED SOLUTIONS

A. SAS-QLA ALGORITHM

In an example of an auction market, GAA users must offer higher bids than SAS-QLA iterative price for GAA users to have more chances of getting access to the radio spectrum. In this scenario, SAS acts as an auctioneer on behalf of PAL operators to give access to GAA users with maximal bidding prices. The price requested by PAL operators may vary according to the market fluctuations and GAA users offer to payoff based on their own preference for the channel and QoS requirements. GAA users also use evaluation of future behaviors and current transaction state to set their preferences.

To implement this auction scenario using the SAS-QLA algorithm, it is assumed that GAA users strictly follow the truth-telling policy, and there is no motivation to misrepresent their information.

1) GAA BIDDING PRICE

GAA user g maintains a preference list $PL_{g,j}^t$ over time t to access channel j. The optimal bid a GAA user can is higher than or equal to $P_{g,j}^t$ i.e., $PL_{g,j}^t \geq P_{g,j}^t$. Hence, a preference list is defined for GAA user g to access channel j over time t as:

$$PL_{g,j}^t = \psi.B_j^t + \zeta_g^{t+1}(S_g^t, \delta_g^t) \tag{4}$$

where B_j^t is a buffer that holds the total accumulative packets, $\zeta_g^{t+1}(S_g^t, \delta_g^t)$ is the future reward and ψ is a factor to regulates the trade-off between future market and current packet expectations.

We define a random variable λ_j^t independent of time here, that stores the number of packets arrived in the buffer in time slot t. The arrival rate is supposed to follow Poisson distribution with λ packets per second. The buffer capacity is set to be C_g. The buffer state S_g of GAA user g is calculated as:

$$B_{S_g}^t = min\{(B_{S_g}^{t-1} - R_{S_g}^{t-1})^+ + \lambda_g^t, C_g\} \tag{5}$$

where, $R_{S_g}^t$ shows the immediate gain received after transmitting packets. The factor $(B_{S_g}^{t-1} - R_{S_g}^{t-1})^+ = max(0, (B_{S_g}^{t-1} - R_{S_g}^{t-1}))$

2) SAS-QLA IMPLEMENTATION

In the SAS-QLA algorithm, we made an assumption that SAS will auction the available PAL reserved channels for a time period of T and the cost payable to PAL operators and reward payoff of GAA users remain the same for this period. SAS releases the radio channels to GAA users when the transaction is completed and the received bidding price satisfies the received price constraint $C_{p,j}$ for the channel j. After receiving payoff $P_{g,j}^t$ from the GAA user, SAS completes the transaction and calculates the reward defined by:

$$P_{g,j}^t = \psi.B_j^t \tag{6}$$

So, SAS aims to maximize the profit for PAL operators and ensures that the radio channel will be leased to GAA users not less than the reserved price. Accordingly, the optimization problem for maximizing PAL operator's profit can be written as:

$$\tau(S_g, P_g^t) = argmax(P_{P_i,j}^t) \tag{7}$$

subject to:

$$P_g^t > C_{p,j}^t \tag{8}$$

and

$$P_{P_i,j}^t = P_{g,j}^t \tag{9}$$

3) SAS-QLA ALGORITHM CONVERGENCE

The convergence of SAS-QLA algorithm depends on the time varying learning factor η_g. The factor η_g uses the results derived from Robbins-Monro theory [51] and set the following conditions to be met for convergence of equation (2) to optimal value uniformly over $\zeta_g^{t+1}(S_g^t, \delta_g^t)^*$ S_g^t and δ_g^t with probability of 1. The conditions are as:

1) The state space and action space must be finite.
2) Variable $\omega_g^t\{S_g^t, \delta_g^t\}$ must be finite.
3) $\sum_{t=0}^{+\infty} \eta_g = \infty$ and $\sum_{t=0}^{+\infty} (\eta_g)^2 < \infty$.
4) If the variable ζ_g^t approaches to 1 then it represents that all strategies converge to cost-free terminal state with probability 1.

In Section III-D, the variables S_g^t and δ_g^t are defined. The variable shows the available channels. The radio channels that can be occupied or released are from the finite set. Hence the first condition is satisfied for condition 1. The reward function defined in equations (1) and (2) shows that $\omega_g^{min} < \omega_g^t < \omega_g^{max}$. As $(\omega_g^t)^2$ is finite. So, variable $\omega_g^t = E(\omega_g^t)^2 - (E(\omega_g^t))^2$ is also finite. It proves that condition 2 holds. In SAS-QLA algorithm η_g is defined as:

$$\eta_g = \begin{cases} \frac{1}{t} & t>0 \\ 0 & t=0 \end{cases} \tag{10}$$

So, this proves that condition 3 holds as well. If the factor $\zeta_g^t = 1$, then the model implemented shows the optimal strategy for achieving maximum gains, whereas, the objective is to maximize the reward function. In this scenario, all policies and strategies approach a terminal state with a probability factor of 1 which is true for the finite horizon model. This shows that condition 4 is also satisfied.

B. KNAPSACK PROBLEM AND DYNAMIC PROGRAMMING ALGORITHM

Knapsack-based dynamic resource allocation model proposed in [52] allows the SAS to select the most suitable GAA users based on their bids to allocate idle PAL reserved channels. The 0-1 Knapsack problem is modeled as a combinatorial optimization problem. It is a scenario in which a constrained knapsack with a fixed size is filled with the most important and profitable elements. In this article, SAS considers a pool of available idle PAL reserved channels as a limited sack of capacity C. Each GAA user that can be placed in the sack has a certain weight w_g and a factor of profit p_g. So, the problem defined here is to give space to GAA users in the sack while SAS ensures that maximum profit is guaranteed to PAL operators, and required QoS is provided to GAA users are modeled as constraints. So, the knapsack problem for allocation of radio channels to GAA users can be modeled as a combinatorial optimization problem, formulated as:

$$max \sum_{j=1}^{l} \sum_{g=1}^{z} p_{j,g} x_{j,g} \tag{11}$$

subject to

$$\sum_{g=1}^{z} w_g x_g \leq C \tag{12}$$

where, x_g represents the binary variable that shows whether a GAA user is selected or not.

$$x_g = \begin{cases} 1 & \text{if the GAA user } g \text{ is selected} \\ 0 & \text{if the GAA user is not selected} \end{cases}$$

Suppose there are g GAA users seeking opportunities to access idle PAL channels available in the sack managed by SAS. The GAA user pays a price (profit of PAL operators for leasing the PAL idle reserved channel) p_j, weight w_g and capacity C. where p_j, w_g and C are the positive integers. The optimal solution of the problem modeled in equation (11) is given by:

$$\upsilon_g = max\{\upsilon_{g-1}(C), \upsilon_{g-1}(C)\} \tag{13}$$

where, υ_g represents the optimal solution. The optimization problem defined in equation (11) is actually a multi-dimensional Knapsack problem and modeled as NP-hard problem [53]. There are different solutions available to solve resource allocation problem that is modeled as knapsack problem. The well-known solution to this problem is the use of dynamic programming (DP) approaches.

Dynamic programming solution modeled in [54] is used to solve knapsack optimization problems. Dynamic programming aims to achieve optimal solution. An optimal solution to a particular problem is combined with similar, overlapping and smaller sub-problems. The key steps of dynamic programming solution are:

1) Decomposition of problem into sub-problems
   Suppose, the maximum cost w is obtained and stored in an array M(g,w) by selecting GAA users $0 < g < z$ while satisfying the maximum load of PAL operators C. If all entries of this array are computed, then the array entry $M(g, w_g)$ contains the maximum profit for PAL operators in terms of cost is the solution to our problem.
2) The recursive equation
   After selecting an optimal solution next step is to recursively define the optimal solutions of the sub-problems. In our scenario, we have two sub-problems; whether a GAA user is given access to PAL reserved channel or rejected of low bid. The recursive equations are

mathematically represented as:

$$M(g, W) = \begin{cases} 0 & \text{if } g = 0 \\ M(g-1, w) & \text{if } w_g > \text{w} \\ \upsilon_g & \text{else} \end{cases} \quad (14)$$

After the step of recursive equations, a dynamic programming solution is achieved for the problem modeled in equation (11).

C. DISTRIBUTED RELAXATION METHOD

Distributed relaxation method was first proposed in [55] to solve the classical assignment problems. The algorithm acts like an auction, where unassigned GAA users bid for the idle PAL reserved channels. SAS will receive bids from all users, once SAS completes the bidding process, then the particular radio channel for which the highest bid is received is assigned to that GAA user. In the worst scenario, the time complexity to assign GAA users a channel becomes $O(gAlog(gC)$, where g is the number of GAA users. A is representing the number of pairs i.e., GAA users and PAL reserved idle channels and C is the maximum cost that a GAA user offer for the required channel.

The distributed relaxation method solves the problem given in equation (11) with the condition of having $g \times j$, where G is the number of GAA users and j represents the number of available PAL reserved channels. To solve this optimization problem, distributed relaxation method uses the concept of cost equilibrium bidding strategy. In the practical scenario, for all available idle PAL reserved channels j, every GAA user has different interests to meet their requirements. Let's suppose SAS issues a pool of n radio channels. A GAA user g has to pay a certain price P to get the access. Based on the interest of each GAA user for the particular radio channel, an interest factor ι_g for g^{th} GAA user to determine the value of interest for each GAA user. Preference for each GAA user can be determined as:

$$Pref_g = \iota_g - P \quad (15)$$

To successfully allocate the idle PAL reserved channel j from the pool of channels j vector to GAA user g, the factor $Pref_g$ must be at maximum. For all GAA users with maximum $Pref_g$ for a particular channel meet the cost equilibrium and assigned with their desired channel. The main steps of distributed relaxation method are given below.

1) The first step deals with the initialization that deals with the random allocation of GAA users to idle PAL reserved channels.
2) SAS assigns the minimum bid for each idle channel from the pool j to $\frac{1}{g-1}$, where g is the number of GAA users participating in the auction process.
3) In third step a GAA user who is not satisfied from the allocation by SAS in step 1 increase the bidding price i.e., $Newprice = oldprice + pref_g + increment$.
4) Profits for PAL operators are updated by subtracting the updated bidding cost for allocating channels to each GAA user and updating the profit array.
5) Repeat step 3 to select a GAA user with maximum unsatisfactory factor and complete the auction process using step 4 and step 5. Complete the auction process till every GAA user is satisfied with the channel assigned and also meets the cost equilibrium.

V. NUMERICAL RESULTS

A. PARAMETERS SETTING

In order to evaluate the performance of the proposed scheme in the SAS-based CBRS spectrum sharing framework, we consider a scenario where there are four PAL operators offering their available radio frequency channels to be auctioned and added to a spectrum pool managed by SAS. To see the impact of GAA user's load in the SAS system, the number of GAA users who took part in a spectrum auction held by SAS is varied from 10 to 500. The scenario is depicted in Figure 1. PAL operators add the information of their unused frequency channels through a common channel to SAS. The SAS creates a pool of available PAL unused frequency channels and holds an auction. The GAA users who need guaranteed QoS for delay-sensitive applications take part in the auction process. We proposed the SAS-QLA algorithm to bid intelligently in the auction based on the available current state, future reward, and transmission requirements. The data rate available to GAA users is modeled using the Shannon capacity theorem [56] expressed as:

$$Datarate = B_g^j log_2\left(1 + \frac{S_g^j}{N_g^j}\right) \quad (16)$$

where B_g^j is the bandwidth available to GAA user g for PAL reserved channel j and $\frac{S_g^j}{N_g^j}$ is the signal to noise ratio received by GAA user g for using PAL reserved channel j.

The following simulation parameters for GAA users are defined for the implementation of the scenario discussed above. GAA users' transmit power is restricted to 1W to limit interference caused by GAA users and remain same for all GAA users. Bandwidth B of sub-carriers is defined as 10 kHz per carrier to find the data rate in the equation 16 as defined in [11]. The default value of discount factor is selected as 0.5 because the effective price where the bidding price is higher than the reserve price is in range of 0.3 to 0.8. Learning rate is varied from 0 to 1. The macro-cell based urban propagation model is selected for the simulations with the Rayleigh multi-path fading model and cell radius of 1 Km. The simulation parameters used in our simulations are shown in Table 2.

All the simulation experiments are executed 50 times to obtain the mean value to minimize the randomness and to get stable results.

B. PERFORMANCE EVALUATION

In this article, we evaluate the performance of our proposed algorithm SAS-QLA to allocate the idle PAL reserved radio channels to GAA users using spectrum trading based on

TABLE 2. Network Parameters.

Parameter	Value
GAA power constraint	1W
Bandwidth of Subcarriers	10 kHz
Discount Factor	0.5 (Default)
Timing period	0.1 Sec
Learning Rate	0-1
Propagation model	$128.1 + 37.6Log_{10}(R)$
Cell radius	1 Km
Fading Model	Rayleigh multi-path

Q-learning that uses reinforcement learning. The proposed radio spectrum allocation algorithm is compared with the dynamic programming solution for the knapsack problem and distributed relaxation method based on the auction method to assign channels to GAA users.

The cumulative distribution function (CDF) of radio channels allocation to 300 GAA users is represented for the scenario in Figure 3, where multiple PAL operators took part in the auction process. Separate readings are calculated for the auction process where a single PAL operator is present and also considered for the auction process with multiple PAL operators. In case when only one PAL operator took part in the auction process, the number of available channels for competing GAA users in an auction was too less, that is the reason only 40% of the GAA users get access to PAL reserved channels using the SAS-QLA algorithm. The percentage is up to 20% less in case of using dynamic programming and distributed relaxation method. When more PAL operators took part in the auction, more radio channels in the pool become available to GAA users so the percentage of GAA users getting access to PAL reserved channels increased from 40% to approximately 100% when 300 GAA users applied. In the case of more than 300 GAA users, the percentage may vary because of the limited availability of radio channels. This result shows that the proposed SAS-QLA algorithm is accommodating up to 20% more GAA users in available PAL reserved channels in comparison with distributed programming and distributed relaxation method.

In this article, we analyzed the impact of discount factor σ_g^t and learning factor η_g defined in equation (2) and equation (10), respectively, for the convergence of our proposed SAS-QLA algorithm. The value of the discount factor remains between 0 and 1. So, we compare the effect of the σ_g^t factor for values between 0 and 1. The impact of the discount factor on the convergence of the SAS-QLA algorithm is represented in Figure 4. It is clear from the Figure 4 that the bidding price and the reserve price grow much faster when the value of σ_g^t is greater than 0.8. We can see that when the discount factor σ_g^ts is greater than 0.8 the reserved prices increased up to 5 times when it approaches the value 1. The price remains under $ 50 when the discount factor is less than 0.8. It means that the future reward defined in equation (1), (2) and (3) has the high impact on SAS-QLA algorithm. When

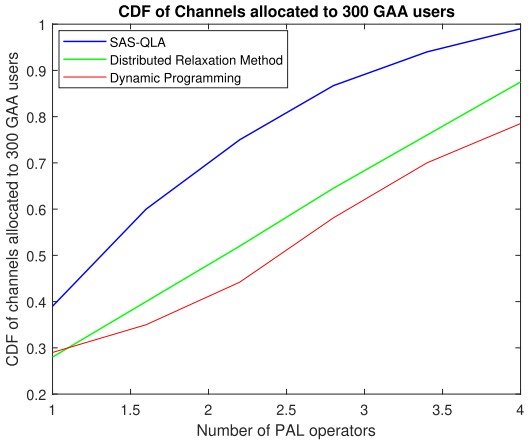

FIGURE 3. CDF of channels allocated to 300 GAA users.

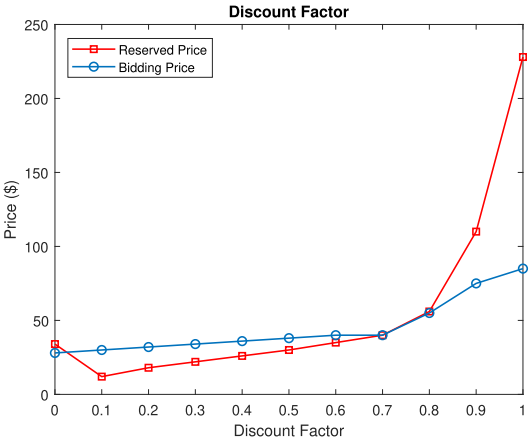

FIGURE 4. Effect of Discount Factor.

the discount factor σ_g^t is too less or too high, the bidding price of GAA users becomes lower than the reserved price of PAL operators set by SAS. So, for the successful trading of radio channels, the bidding price of GAA users must be greater than the reserved price set by PAL operators. Hence, discount factor σ_g^t must be in the range of 0.05 to 0.7 for successful allocation of available radio channels to GAA users. If the discount factor σ_g^t is greater than 0.8 or less than 0.05 then the bidding price of GAA users becomes lesser than the PAL operators reserved price. In this case, SAS will not consider the bidding price of GAA users. In our simulations and parameters settings, we consider the default value of discount factor σ_g^t as 0.5.

The speed of updating the SAS-QLA algorithm depends on the learning rate η_g of the algorithm defined in equation (2) and (10). The effect of learning factor η_g is depicted in Figure 5. Learning factor η_g is not much useful in the convergence of the SAS-QLA algorithm but its main effect is on the speed of updating the SAS-QLA algorithm to converge. We can see in Figure 5 that the bidding price of GAA users is quite high in comparison with the reserved price set by PAL operators. In the case of all values ranging from 0 to 1, the

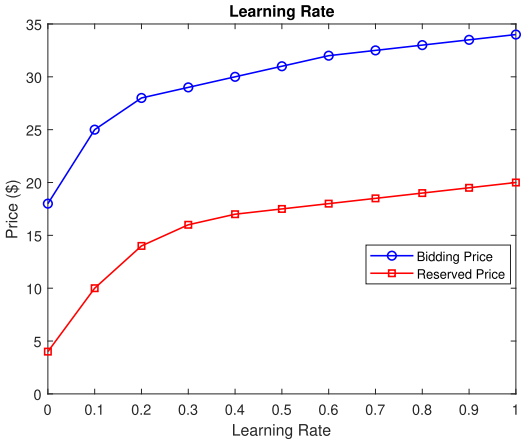

FIGURE 5. Learning Rate.

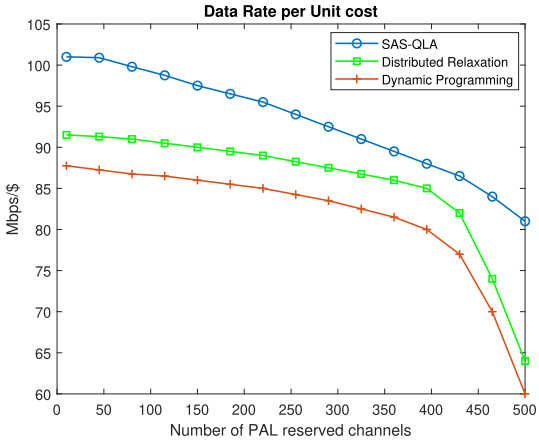

FIGURE 6. Average data rate per unit cost.

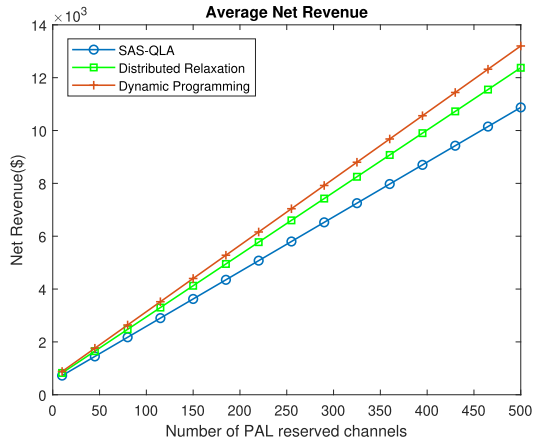

FIGURE 7. Average Net Revenue.

bidding price remains high from the reserved price, we can use any values for successful trade but need to satisfy the constraints defined in section IV-A3.

The average data rate per unit cost is shown in Figure 6. The experiment is conducted for GAA users ranging from 10 to 500. When there is less number of GAA users participating in the auction managed by SAS, then the overall demand for the available PAL idle channels in the pool is less. So, GAA users get the best QoS at less price. As the number of GAA users increases, the participants in the auction of the PAL idle reserved channels become more competitive as a result of the overall net data rate per unit cost being reduced which is also evident from the graph.

One of the purposes of the PAL reserved channels auction was to increase the revenue of PAL operators as PAL users are not active all the time so PAL operators can take advantage by leasing the channels to GAA users to increase their revenue. The net revenue gained by PAL operators is shown in Figure 7. The GAA users use the SAS-QLA algorithm to bid intelligently based on the current state and future reward. In the case of using dynamic programming and distributed relaxation method, both methods select the maximum cost offered by GAA users without offering intelligent bidding. If the preference of SAS is to maximize the net revenue for PAL operators, then selecting the dynamic programming method will return the maximum revenue. In the case of using the SAS-QLA algorithm, the SAS maintains a preference list with respect to giving the benefit to the GAA users by allocating the maximum data rate at minimum cost. Hence, the net revenue for PAL operators in the case of using the SAS-QLA algorithm will be less as compared to dynamic programming and distributed relaxation method. In the scenario, where giving the benefit to the GAA users is the priority, then the SAS-QLA algorithm outperforms the other two methods, which is evident from the Figure 6 because of the GAA user's ability to bid intelligently. Hence, it is clear from Figure 7 that if SAS opts to prefer the PAL operators over GAA users, PAL operators will get 23% more revenue in comparison with the SAS-QLA algorithm and 9% more revenue in comparison to the distributed relaxation method while satisfying FCC rules and also GAA users getting their desired QoS at the best price using our proposed SAS-QLA algorithm. If the priority is to prefer the best data rate per unit cost, then it is evident from the results as depicted in Figure 6, that SAS will opt for the SAS-QLA algorithm as the proposed algorithm gives a 43% improved data rate per unit cost and a 32 % more data rate per unit cost in comparison with the dynamic programming and the distributed relaxation method respectively.

Figure 8 shows the overall fairness of the algorithms based on Jain's fairness index (JFI) on a scale of 0 to 1. The fairness index of the algorithms depends on the factor that how many GAA users are assigned channels out of the total GAA users who became part of the spectrum auction. We can see from the graph that the overall fairness of the algorithm reduces as GAA users in the system are increased. It is evident from the spectrum trading that as GAA users in the spectrum auction increase, more GAA users will remain unassigned because of the limited availability of radio channels. Hence, the overall Jain's fairness index reduces as the number of GAA users increases. On contrary, when the number of

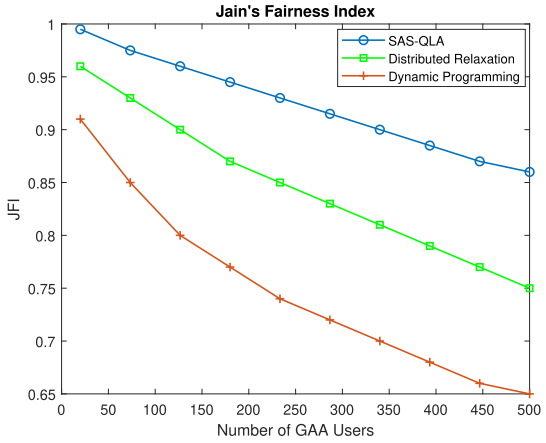

FIGURE 8. GAA users satisfaction level.

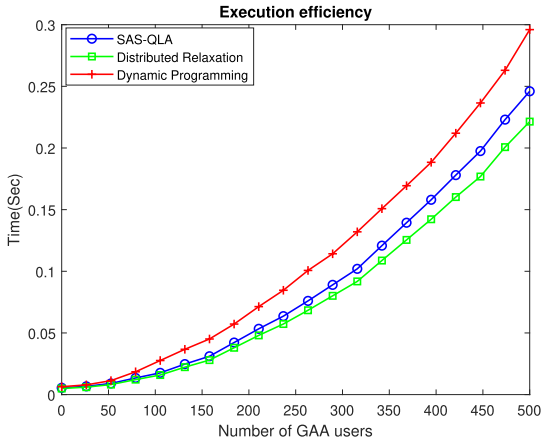

FIGURE 9. Execution efficiency.

GAA users in the auction is limited the JFI value approaches 1 approximately. In comparison to dynamic programming and distributed relaxation method, the proposed algorithm SAS-QLA is accommodating more GAA users with the best QoS and the overall JFI remains better than other algorithms.

Figure 9 depicts the execution efficiency of the proposed SAS-QLA algorithm, dynamic programming method, and distributed relaxation method to allocate idle PAL reserved channels to GAA users. The distributed relaxation method outperforms other algorithms because it allocates the available preferred channels to GAA users without using any learning strategy. The proposed SAS-QLA algorithm allocates 500 GAA users in almost 250ms. Dynamic programming takes 300 ms to allocate available channels to GAA users because this algorithm searches for the best match in the worst scenario; the time complexity of dynamic programming is $O(n^3)$. The proposed SAS-QLA algorithm is quite efficient as compared to dynamic programming but in comparison with distributed relaxation method, the SAS-QLA algorithm is achieving the best data rate per unit cost and fairness index.

The simulation results presented show that the proposed SAS-QLA algorithm efficiently allocates the GAA users according to their requirements. The SAS-QLA algorithm allows the GAA users to bid according to the future reward, their current states, and transmission requirement. The bidding price and reserved price of GAA users are executed locally. When preference of the SAS is to give benefits to GAA users instead of PAL operators than the SAS-QLA algorithm outperforms the dynamic programming method and distributed relaxation method to achieve better bidding efficiency, the data rate per unit cost, and better GAA load management. Maximum GAA users are accommodated while meeting their desired QoS.

VI. CONCLUSION

In this paper, we proposed a Q-learning-based auction algorithm that in turn is based on reinforcement learning in order to meet the requirements of GAA users in CBRS-SAS architecture. GAA users are the least priority users in CBRS-SAS architecture. The SAS as a central controlling entity does not guarantee to provide the required QoS to GAA users. For delay-sensitive real-time applications, GAA users need guaranteed QoS. GAA reserved channels are not a good option for delay-sensitive applications. As a matter of fact, PAL users are licensed users and have guaranteed QoS provisioning from SAS but PAL users are not active all the time to fully utilize the PAL reserved channels. PAL operators paid a license fee to get access to these PAL reserved channels. In order to increase the revenue, the PAL operators can use this opportunity to auction these channels. The PAL operators share a list of available idle channels with SAS, where SAS manages a pool of available idle channels shared by all PAL operators who want to take part in spectrum trading with GAA users.

Our proposed algorithm allows the GAA users to bid according to the future reward, their current state, and experienced environment. The PAL operators also share a vector of reserved price i.e., SAS only accepts the bids of GAA users for available idle channels when the bid price is higher than the reserved price. Finally, the practicality of the SAS-QLA algorithm is validated using SAS-QLA convergence analysis.

The simulation results of the SAS-QLA algorithm confirm that the proposed algorithm is much more efficient in allocating maximum numbers of GAA users while satisfying their QoS requirements. SAS-QLA algorithm also maximizes the data rate per unit cost while the execution efficiency is not disturbed. Jain's fairness index (JFI) shows that when the number of GAA users is less, who take part in the auction process the JFI approaches 1 i.e., the maximum limit. When the GAA users in the spectrum auction process increase to 500, the SAS-QLA algorithm outperforms the other algorithms and accommodates 10% more users in comparison with distributed relaxation and 20% more users in comparison to dynamic programming. The proposed SAS-QLA algorithms provide an approximate optimal solution that converges fast.

REFERENCES

[1] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road towards 6G," 2021, *arXiv:2107.03059*.

[2] J Holdren and E. Lander, "Realizing the full potential of government-held spectrum to spur economic growth," Executive Office President, President's Council Advisors Sci. Technol. (PCAST), Federal Commun. Commission, Washington, DC, USA, Tech. Rep., Jul. 2012. [Online]. Available: http://www.whitehouse.gov/administration/eop/ostp

[3] Fcc, "Shared commercial operations in the 3550–3650 MHz band," Report and Order and Second Further Notice of Proposed Rulemaking, Federal Commun. Commission, Washington, DC, USA, Tech. Rep., Jun. 2015.

[4] CBRS Alliance, "CBRS network service technical specifications," OnGo Alliance, Beaverton, OR, USA, Tech. Rep., CBRSA-TS-1002, Feb. 2018.

[5] FCC, "Electronic code of federal regulations, title-47: Telecommunication, part 96-CBRS," Federal Commun. Commission, Washington, DC, USA, Tech. Rep. CFR-2016-title47-vol5-part96, Jul. 2015.

[6] FCC, "Promoting investment in the 3550–3700 MHz band," Federal Commun. Commission, Washington, DC, USA, Tech. Rep., Oct. 2018.

[7] S. Bhattarai, J.-M. J. Park, B. Gao, K. Bian, and W. Lehr, "An overview of dynamic spectrum sharing: Ongoing initiatives, challenges, and a roadmap for future research," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 2, pp. 110–128, Jun. 2016.

[8] M. M. Sohul, M. Yao, T. Yang, and J. H. Reed, "Spectrum access system for the citizen broadband radio service," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 18–25, Jul. 2015.

[9] G. Saha and A. A. Abouzeid, "Optimal spectrum partitioning and licensing in tiered access under stochastic market models," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 1948–1961, Oct. 2021.

[10] K. B. S. Manosha, S. Joshi, T. Hanninen, M. Jokinen, P. Pirinen, H. Posti, K. Horneman, S. Yrjola, and M. Latva-aho, "A channel allocation algorithm for citizens broadband radio service/spectrum access system," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–6.

[11] CBRS Alliance, "CBRS network services use cases and requirements," OnGo Alliance, Beaverton, OR, USA, Tech. Rep., CBRSA-TS-1001 v4.0.0, Mar. 2021.

[12] Y. Akçay, H. Li, and S. H. Xu, "Greedy algorithm for the general multidimensional knapsack problem," *Ann. Oper. Res.*, vol. 150, no. 1, pp. 17–29, Feb. 2007.

[13] K. Chebil and M. Khemakhem, "A dynamic programming algorithm for the knapsack problem with setup," *Comput. Oper. Res.*, vol. 64, pp. 40–50, Dec. 2015.

[14] S. Coniglio, F. Furini, and P. S. Segundo, "A new combinatorial branch-and-bound algorithm for the knapsack problem with conflicts," *Eur. J. Oper. Res.*, vol. 289, no. 2, pp. 435–455, 2021.

[15] N. Zhao, F. R. Yu, H. Sun, and M. Li, "Adaptive power allocation schemes for spectrum sharing in interference-alignment-based cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3700–3714, May 2016.

[16] T. X. Quach, H. Tran, E. Uhlemann, and M. T. Truc, "Secrecy performance of cooperative cognitive radio networks under joint secrecy outage and primary user interference constraints," *IEEE Access*, vol. 8, pp. 18442–18455, 2020.

[17] C. Gan, R. Zhou, J. Yang, and C. Shen, "Cost-aware learning and optimization for opportunistic spectrum access," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 1, pp. 15–27, Mar. 2019.

[18] T.-K. Le, U. Salim, and F. Kaltenberger, "An overview of physical layer design for ultra-reliable low-latency communications in 3GPP releases 15, 16, and 17," *IEEE Access*, vol. 9, pp. 433–444, 2021.

[19] A. Abdelhadi, H. Shajaiah, and C. Clancy, "A multitier wireless spectrum sharing system leveraging secure spectrum auctions," *IEEE Trans. Cognit. Commun. Netw.*, vol. 1, no. 2, pp. 217–229, Jun. 2015.

[20] L. Sendrei, J. Pastircak, S. Marchevsky, and J. Gazda, "Cooperative spectrum sensing schemes for cognitive radios using dynamic spectrum auctions," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2015, pp. 159–162.

[21] Q. Wang, J. Huang, Y. Chen, X. Tian, and Q. Zhang, "Privacy-preserving and truthful double auction for heterogeneous spectrum," *IEEE/ACM Trans. Netw.*, vol. 27, no. 2, pp. 848–861, Apr. 2019.

[22] F. Benedetto, L. Mastroeni, and G. Quaresima, "Auction-based theory for dynamic spectrum access: A review," in *Proc. 44th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2021, pp. 146–151.

[23] A. Upadhye, P. Saravanan, S. S. Chandra, and S. Gurugopinath, "A survey on machine learning algorithms for applications in cognitive radio networks," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2021, pp. 01–06.

[24] T. T. Thanh Le and S. Moh, "Comprehensive survey of radio resource allocation schemes for 5G V2X communications," *IEEE Access*, vol. 9, pp. 123117–123133, 2021.

[25] F. S. Samidi, N. A. M. Radzi, W. Ahmad, F. Abdullah, M. Z. Jamaludin, and A. Ismail, "5G new radio: Dynamic time division duplex radio resource management approaches," *IEEE Access*, vol. 9, pp. 113850–113865, 2021.

[26] W. Gulzar, A. Waqas, H. Dilpazir, A. Khan, A. Alam, and H. Mahmood, "Power control for cognitive radio networks: A game theoretic approach," *Wireless Pers. Commun.*, vol. 123, no. 1, pp. 745–749, 2021.

[27] T. LeAnh, N. H. Tran, S. Lee, E.-N. Huh, Z. Han, and C. S. Hong, "Distributed power and channel allocation for cognitive femtocell network using a coalitional game in partition-form approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3475–3490, Apr. 2017.

[28] X. Liao, J. Si, J. Shi, Z. Li, and H. Ding, "Generative adversarial network assisted power allocation for cooperative cognitive covert communication system," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1463–1467, Jul. 2020.

[29] F. Li, K.-Y. Lam, X. Li, X. Liu, L. Wang, and V. C. M. Leung, "Dynamic spectrum access networks with heterogeneous users: How to price the spectrum?" *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5203–5216, Jun. 2018.

[30] D. B. Rawat, A. Alshaikhi, A. Alshammari, C. Bajracharya, and M. Song, "Payoff optimization through wireless network virtualization for IoT applications: A three layer game approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2797–2805, Apr. 2019.

[31] N. N. Sapavath and D. B. Rawat, "Wireless virtualization architecture: Wireless networking for Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5946–5953, Jul. 2020.

[32] B. Li, Z. Fei, Z. Chu, F. Zhou, K.-K. Wong, and P. Xiao, "Robust chance-constrained secure transmission for cognitive satellite–terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4208–4219, May 2018.

[33] L. Wang, F. Li, X. Liu, K.-Y. Lam, Z. Na, and H. Peng, "Spectrum optimization for cognitive satellite communications with Cournot game model," *IEEE Access*, vol. 6, pp. 1624–1634, 2018.

[34] F. Li, K.-Y. Lam, N. Zhao, X. Liu, K. Zhao, and L. Wang, "Spectrum trading for satellite communication systems with dynamic bargaining," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4680–4693, Oct. 2018.

[35] F. Li, K.-Y. Lam, J. Hua, K. Zhao, N. Zhao, and L. Wang, "Improving spectrum management for satellite communication systems with hunger marketing," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 797–800, Jun. 2019.

[36] X. Liu, Q. Sun, W. Lu, C. Wu, and H. Ding, "Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5G," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 67–73, Oct. 2020.

[37] X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng, and P. Li, "Reinforcement learning-based multislot double-threshold spectrum sensing with Bayesian fusion for industrial big spectrum data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3391–3400, May 2021.

[38] X. Chen, C. Wu, T. Chen, H. Zhang, Z. Liu, Y. Zhang, and M. Bennis, "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2268–2281, Apr. 2020.

[39] Z. Youssef, E. Majeed, M. D. Mueck, I. Karls, C. Drewes, G. Bruck, and P. Jung, "Concept design of medium access control for spectrum access systems in 3.5 GHz," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2018, pp. 1–8.

[40] X. Ying, M. M. Buddhikot, and S. Roy, "Coexistence-aware dynamic channel allocation for 3.5 GHz shared spectrum systems," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Mar. 2017, pp. 1–2.

[41] I.-P. Belikaidis, A. Georgakopoulos, E. Kosmatos, V. Frascolla, and P. Demestichas, "Management of 3.5-GHz spectrum in 5G dense networks: A hierarchical radio resource management scheme," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 57–64, Jun. 2018.

[42] N. N. Krishnan, N. Mandayam, I. Seskar, and S. Kompella, "Experiment: Investigating feasibility of coexistence of LTE-U with a rotating radar in CBRS bands," in *Proc. IEEE 5G World Forum (5GWF)*, Jul. 2018, pp. 65–70.

[43] A. Kliks, P. Kryszkiewicz, L. Kułacz, K. Kowalik, M. Kołodziejski, H. Kokkinen, J. Ojaniemi, and A. Kivinen, "Application of the CBRS model for wireless systems coexistence in 3.6–3.8 GHz band," in *Proc. Int. Conf. Cogn. Radio Oriented Wireless Netw.* Cham, Switzerland: Springer, 2017, pp. 100–111.

[44] S. Biswas, A. Bishnu, F. A. Khan, and T. Ratnarajah, "In-band full-duplex dynamic spectrum sharing in beyond 5G networks," *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 54–60, Jul. 2021.
[45] M. Grissa, A. A. Yavuz, B. Hamdaoui, and C. Tirupathi, "Anonymous dynamic spectrum access and sharing mechanisms for the CBRS band," *IEEE Access*, vol. 9, pp. 33860–33879, 2021.
[46] Y. Xiao, S. Shi, W. Lou, C. Wang, X. Li, N. Zhang, Y. T. Hou, and J. H. Reed, "Decentralized spectrum access system: Vision, challenges, and a blockchain solution," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 220–228, Feb. 2022.
[47] X. Dong, L. Cheng, G. Zheng, and T. Wang, "Network access and spectrum allocation in next-generation multi-heterogeneous networks," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 8, 2019, Art. no. 1550147719866140.
[48] W. Abbass, R. Hussain, J. Frnda, N. Abbas, M. A. Javed, and S. A. Malik, "Resource allocation in spectrum access system using multi-objective optimization methods," *Sensors*, vol. 22, no. 4, p. 1318, Feb. 2022.
[49] *SAS to CBSD Protocol Technical Report-B*, Spectrum Sharing Committee Work Group-3, Wireless Innovation Forum, Reston, VA 20191, USA, document WINNF-15-P-0062, Mar. 2016.
[50] Y. Li, "Optimal reserve prices in sealed-bid auctions with reference effects," *Int. J. Ind. Org.*, vol. 71, Jul. 2020, Art. no. 102624.
[51] D. B. Rokhlin, "Robbins–Monro conditions for persistent exploration learning strategies," in *Modern Methods in Operator Theory and Harmonic Analysis*. Cham, Switzerland: Springer, 2018, pp. 237–247.
[52] G. R. Bitran and A. C. Hax, "Disaggregation and resource allocation using convex knapsack problems with bounded variables," *Manage. Sci.*, vol. 27, no. 4, pp. 431–441, Apr. 1981.
[53] U. Pferschy, J. Schauer, and C. Thielen, "Approximating the product knapsack problem," *Optim. Lett.*, vol. 15, no. 8, pp. 2529–2540, Nov. 2021.
[54] S. T. T. Sin, "The parallel processing approach to the dynamic programming algorithm of knapsack problem," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (ElConRus)*, Jan. 2021, pp. 2252–2256.
[55] D. P. Bertsekas, "The auction algorithm: A distributed relaxation method for the assignment problem," *Ann. Oper. Res.*, vol. 14, no. 1, pp. 105–123, Dec. 1988.
[56] C.-F. Yang, S.-C. Chang, and C.-Y. Hsu, "Hierarchical game theoretic design of frequency assignment and channel selection for general authorized accesses," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Apr. 2019, pp. 448–452.

WASEEM ABBASS received the B.Sc. degree in computer engineering from COMSATS University Islamabad (CUI), in 2011, and the M.S. degree in computer engineering from the University of Engineering and Technology (UET), Taxila, in 2014. He is currently pursuing the Ph.D. degree with CUI, with a focus on architecture and protocol design of wireless networks especially for cognitive radio network. He is also a Lecturer with the Department of Computer Science, CUI. His research interests include D2D communication, cognitive radio networks, wireless sensor networks, and the Internet of Things.

RIAZ HUSSAIN received the B.S. degree (Hons.) in electrical engineering from the University of Engineering and Technology, Peshawar, Pakistan, the master's degree in networks from North Carolina State University, Raleigh, NS, USA, and the Ph.D. degree from the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2013. His dissertation was entitled "Modeling, Analysis and Optimization of Vertical Handover Schemes in Heterogeneous Wireless Networks." He is currently an Associate Professor with the Department of Electrical Engineering, COMSATS University Islamabad. His current research interests include cognitive radio networks, device to device communication, and the Internet of Things.

JAROSLAV FRNDA (Senior Member, IEEE) was born in Slovakia, in 1989. He received the M.Sc. and Ph.D. degrees from the Department of Telecommunications, VSB–Technical University of Ostrava, Czechia, in 2013 and 2018, respectively. He is currently working as an Assistant Professor at the University of Žilina, Slovakia. He has authored and coauthored 24 SCI-E and nine ESCI papers in WoS. His research interests include quality of multimedia services in IP networks, data analysis, and machine learning algorithms.

IRFAN LATIF KHAN received the B.Sc. degree in electrical engineering from The University of Azad Jammu and Kashmir, Pakistan, in 1999, the M.S. degree in telecommunication engineering from the University of Management and Technology Lahore, Pakistan, in 2004, and the Ph.D. degree from the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Pakistan, in 2018. He has seven years of industrial experience as an Assistant Divisional Engineer with Pakistan Telecommunication Company Ltd. Since 2008, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, COMSATS University Islamabad. His research interests include spectrum sensing, medium access control, and resource management in cognitive radio networks.

MUHAMMAD AWAIS JAVED (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Engineering and Technology Lahore, Pakistan, in August 2008, and the Ph.D. degree in electrical engineering from The University of Newcastle, Australia, in February 2015. From July 2015 to June 2016, he worked as a Postdoctoral Research Scientist with the Qatar Mobility Innovations Center (QMIC) on SafeITS Project. He is currently working as an Associate Professor with COMSATS University Islamabad, Pakistan. His research interests include intelligent transport systems, vehicular networks, protocol design for emerging wireless technologies, and the Internet of Things.

SHAHZAD A. MALIK received the B.S. degree in electrical engineering from the University of Engineering and Technology Lahore, Lahore, Pakistan, in 1991, and the M.S. degree in communication systems and networks and the M.Phil. degree in digital telecommunication systems from the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, France, in 1997 and 1998, respectively. He was a Postdoctoral Research Fellow/Student Project Advisor with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada, from 2003 to 2004. He was an Assistant Professor with the College of Electrical Engineering and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan, from 2004 to 2007. Since 2007, he has been with the Department of Electrical Engineering, COMSATS University Islamabad, Pakistan, where he is currently a Full Professor and the Dean of electrical engineering. His current research interests include wireless multimedia information systems, mobile computing, QoS provisioning and radio resource management in heterogeneous wireless networks (mobile cellular-2.5/3G/4G, HSPA, LTE, WLANs, WiMAX, MANETs, and WSN), modeling, simulation, and performance analysis, network protocols, architecture and security, wireless application development, embedded system design, and the Internet of Things.