

RESEARCH

Open Access



Resource allocation in shared spectrum access communications for operators with diverse service requirements

Mirza Golam Kibria^{*}, Gabriel Porto Villardi, Kentaro Ishizu, Fumihide Kojima and Hiroyuki Yano

Abstract

In this paper, we study inter-operator spectrum sharing and intra-operator resource allocation in shared spectrum access communication systems and propose efficient dynamic solutions to address both inter-operator and intra-operator resource allocation optimization problems. For inter-operator spectrum sharing, we present two competent approaches, namely the subcarrier gain-based sharing and fragmentation-based sharing, which carry out fair and flexible allocation of the available shareable spectrum among the operators subject to certain well-defined sharing rules, traffic demands, and channel propagation characteristics. The subcarrier gain-based spectrum sharing scheme has been found to be more efficient in terms of achieved throughput. However, the fragmentation-based sharing is more attractive in terms of computational complexity. For intra-operator resource allocation, we consider resource allocation problem with users' dissimilar service requirements, where the operator supports users with delay constraint and non-delay constraint service requirements, simultaneously. This optimization problem is a mixed-integer non-linear programming problem and non-convex, which is computationally very expensive, and the complexity grows exponentially with the number of integer variables. We propose less-complex and efficient suboptimal solution based on formulating exact linearization, linear approximation, and convexification techniques for the non-linear and/or non-convex objective functions and constraints. Extensive simulation performance analysis has been carried out that validates the efficiency of the proposed solution.

Keywords: Shared spectrum access, Resource allocation, Delay constraint service, NLP relaxation, Linear approximation

1 Introduction

Frequency spectrum is an extremely valuable and important natural resource. The exponential increase in demand for the technologies like Wi-Fi or smart electricity grids means we must utilize this finite radio resource very efficiently. But matching this exponentially growing demand for wireless connectivity is harder in the absence of unused or vacant spectrum. In traditional exclusive licensing systems, many frequency bands are spatially and temporally underutilized. Due to the deficiency of the spectrum resources and to support the predicted enormous wireless traffic explosion in future, it is important

to make full use of the existing radio resources. Spectrum sharing presents a supplementary approach to conventional license-exempt and exclusive licensing schemes and can be realized to cope with the existing network infrastructure with the support of new technologies. Even though many applications still depend on exclusive access to spectrum, spectrum sharing [1–3] is increasingly recognized as the breeding framework for wireless innovation that triggers the development and deployment of more resilient and flexible wireless technologies.

Spectrum sharing among operators can appear in many different scenarios. One example is co-primary sharing, where the spectrum regulator licenses a frequency band to multiple operators without specifying the boundaries between the bands of spectrum sharing operators and all the operators have equal right to access the shareable

^{*}Correspondence: mirza.kibria@nict.go.jp
Wireless Network Research Institute, National Institute of Information and Communications Technology, 3-4 Hikarino-oka, Yokosuka, 239-0847 Kanagawa, Japan

spectrum. Another example is licensed shared access scenario, where an incumbent user licenses its frequency band to multiple operators for shared usage in a certain geographical location and for a certain time period. Spectrum sharing is coordinated in accordance with sharing rules under a well-defined set of conditions and mutual agreement. Shared spectrum access [4–9] facilitates efficient utilization of the available spectrum in the 5th-generation (5G) and beyond networks and will become unquestionably mandatory in order to accommodate the predicted enormous wireless traffic explosion. It acts as an intermediary solution between conventional unlicensed and licensed strategies in which the spectrum sharing operators share the licensed spectrum under a decided set of coverage restrictions and time period. Furthermore, spectrum sharing represents a supplementary approach to conventional license-exempt and exclusive licensing schemes, and can be realized to cope with the existing network infrastructure with reasonable and feasible modifications [10].

1.1 Related works and issues on inter-operator spectrum sharing

Dynamic and flexible inter-operator spectrum sharing among the participating operators is very important in shared spectrum access scenarios. A large number of issues have to be considered, such as the spectrum sharing policy, operators' individual traffic demands, structure of the shared spectrum, i.e., contiguous/non-contiguous, the operating environment, channel propagation characteristics, and inter-operator interference. Unlike resource allocation in other systems, inter-operator resource allocation in shared spectrum communication depends on various factors such as license agreement policy and traffic demands along with other conventional constraints [11, 12].

The work in [12] considered a shared spectrum access model, where each operator is allocated with a fragment of the shareable spectrum. However, [12] does not consider dynamic fragment sharing among the operators, which can have a significant impact on the system performance. This is because the achievable throughput for a *spectrum* sharing operator over different fragments of the shared spectrum can vary quite significantly depending on the types of applications and channel characteristics. In [13], a centralized approach for spectrum sharing across multiple operators has been proposed based on a coordinated scheduling algorithm. The authors of [14, 15] considered orthogonal frequency division multiple access (OFDMA)-based scheme and proposed spectrum sharing approaches from a game theoretic perspective under cognitive radio context, where the spectrum sharing operators are classified as primary and secondary. Unfortunately, each of the abovementioned shared spectrum allocation works

considered either subcarrier gain-based or fragmentation-based schemes, not both.

Furthermore, during the inter-operator spectrum sharing process, if an operator needs to ensure that all the users have approximately same data rates or each user should be able to transmit at a minimum rate, some notion of fairness has to be incorporated in the optimization process that gives the users the way of being treated in accordance with the fairness notion. In this study, we aim at treating all the users equally in terms of allocating resources to them. Along with achieving overall higher system throughput, maximizing fairly shared spectrum efficiency is very important, especially from the spectrum sharing operators' perspectives.

1.2 Related works and issues on intra-operator resource allocation

From the allocated spectrum in the inter-operator spectrum sharing stage, each operator then allocates the radio resources to its own users depending on system objective, users' applications types, and other constraints. In general, the operators can perform such resource allocation independently from each other. For the intra-operator resource allocation, we consider non-overlapping subcarrier allocation and the operators support users with heterogeneous service requirements. There are many works in literature dealing with the problem of resource allocation in OFDMA system under various system constraints [16–20]. For instance, the authors of [16, 17] have shown that the overall system capacity of an OFDMA system is optimized when each subcarrier is assigned to the user with the best channel gain. The *max-min* optimization problem is addressed in [16], where all the users are assured to achieve a similar data rate through the maximization of the worst users' capacity.

The algorithm proposed in [17] is aimed at the maximization of data rate under total transmitting power and target bit error rate requirements. In [18, 19], the authors claim that non-convexity is not an issue for the resource allocation problem in an OFDMA system if the number of subcarriers is very large. In [20], the authors proposed an iterative resource allocation algorithm to minimize the total transmitting power under fixed user data rates and bit error rate constraints. In [21], the authors proposed a best-effort fairness scheme that ensures minimum number of subchannels for all the users. In [24, 25], adaptive resource allocation in OFDMA system is considered under partial channel state information. In [26–29], the authors formulate an optimization problem, which balances the trade-off between capacity and fairness among the users. Proportional fairness is assured, i.e., ensures that the rates of different users are proportional, by imposing a set of non-linear constraints.

We consider intra-operator resource allocation for a system with users with delay constraint service requirements. The users are categorized under two difference sets, i.e., delay constraint (DC) users and non-delay constraint (NDC) users. Unlike [17, 20, 23], which consider homogenous traffic for DC users and NDC users, respectively, in our considered system model, both DC and NDC traffics can be supported simultaneously. Due to the presence of non-linear structure in the objective function and constraints, and non-sharing nature of the subcarrier allocation among the users, the optimization problem becomes a mixed-integer non-linear programming (MINLP) problem, which is computationally very expensive. In [22], the authors studied the resource allocation problem in a system with users requiring delay differentiated services and consider fairness in terms of delay sensitive users. They proposed a suboptimal solution by introducing time-sharing variables, and therefore, the system model employed differs from the original OFDMA system.

1.3 Contributions

For inter-operator spectrum sharing in shared spectrum access communications, we propose two solutions that are computationally inexpensive. For both of the solutions, optimizing total system throughput has been the objective metric while allocating spectrum resources to the operators. The first solution (a.k.a. subcarrier gain-based spectrum sharing) is iterative in nature. In this proposed scheme, unlike [13–15], we emphasize on fairness issues not only for the spectrum sharing operators but also for the users served by the operators, by taking care of the sharing policy measures based on well-defined sharing conditions, traffic demands, and propagation environments. The second proposed solution (a.k.a. fragmentation-based spectrum sharing) is based on allocating spectrum fragments to the operators instead of subcarriers as in the case of the first solution, where each fragment is a set of larger number of contiguous subcarriers. Note that each operator can dynamically obtain multiple non-contiguous fragments from the shared spectrum if the available spectrum for sharing is non-contiguous. Contrary to the shared spectrum scheme proposed in [13], the proposed solution in this work allows dynamic sharing of frequency spectrum among the operators depending on the types of applications, e.g., short-range communications, the technologies it operates on, and channel propagation characteristics.

For the intra-operator resource (spectrum and power) allocation problem, we propose a computationally efficient (if not, at least solvable) solution based on some linearization techniques (exact linearization or linear approximations) considering the structures of the optimization problem and the constraints. The performance

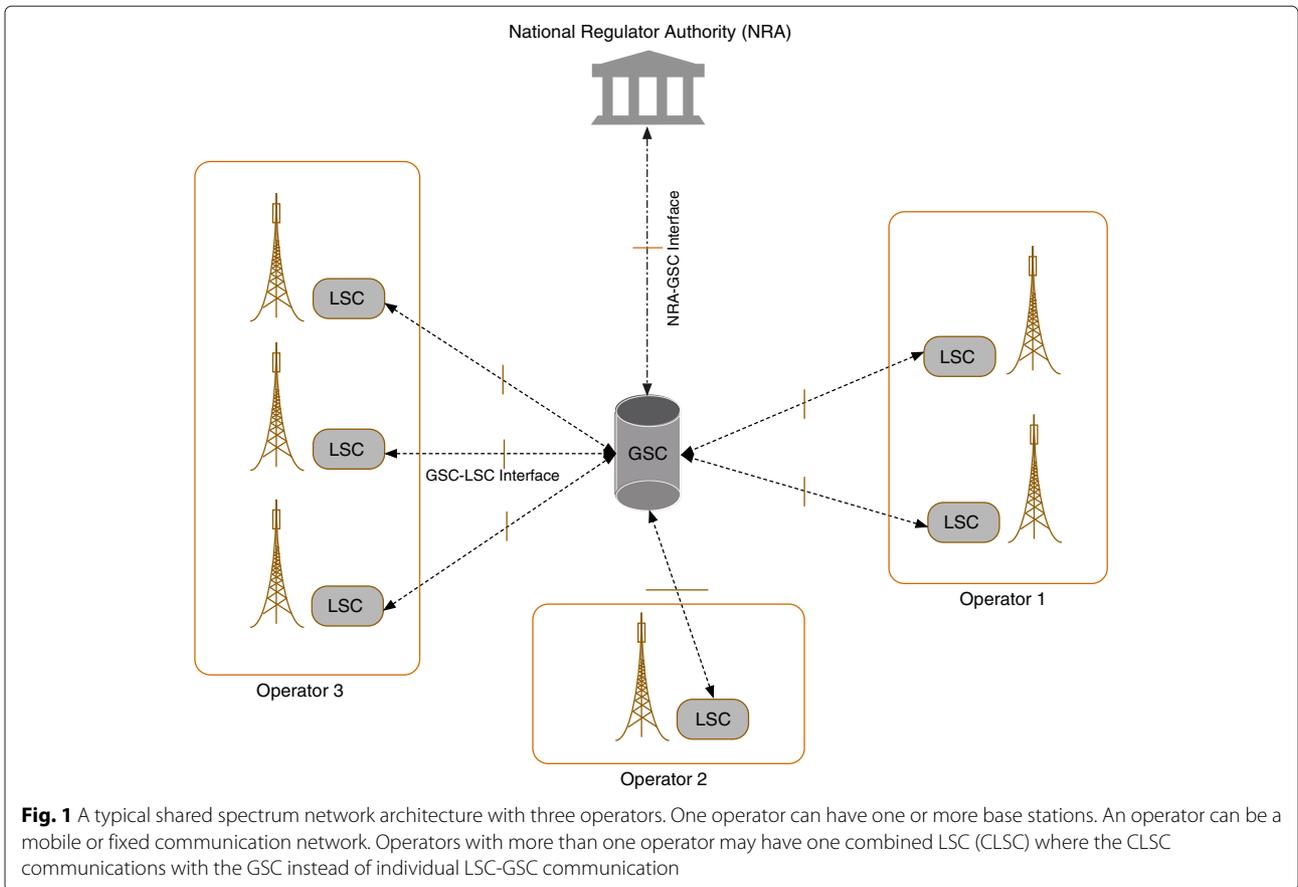
of the proposed solution is impressive when compared to the original MINLP and other existing solutions. In particular, we transform the computationally expensive non-convex MINLP into a convex problem by introducing a series of efficient linearization, linear approximation, and convexification techniques, therefore, significantly reducing the computational time.

The remainder of this paper is structured as follows. The system model and the problem statement are discussed in Section 2. In Section 3, we discuss two competent solutions for inter-operator spectrum sharing. In Section 4, the proposed scheme for intra-operator resource allocation for users with dissimilar services is discussed. In Section 5, we describe the simulation parameters and evaluate the performances of the proposed solutions. Finally, we conclude the paper in Section 6.

2 System model

We consider a co-primary or horizontal spectrum sharing communication model with N_{op} operators, where all the operators have equal right to access the spectrum. The National Spectrum Regulatory Authority (NRA) licenses a shareable spectrum band to N_{op} operators participating in shared spectrum access, without fixed boundaries between spectrum bands of different operators. The operators coordinate their spectrum usage according to certain sharing rules and mutual agreement. All the N_{op} participating operators employ orthogonal frequency division multiple access or multi-carrier waveforms [12], e.g., OFDMA, filter bank multi-carrier waveforms, and spectrum sharing is achieved in a coordinated way. Different operators have the flexibility to use different air interfaces that support scalable bandwidth and flexible size of discrete Fourier transform (DFT). An entity called the global spectrum controller (GSC) carries out the coordinated spectrum sharing related tasks. The GSC can be either a virtual entity implemented in a distributed way in the base stations or an separate/independent entity in the network infrastructure. A typical shared spectrum network architecture with three operators is depicted in Fig. 1.

An operator or a shared spectrum licensee is an entity operating a mobile/fixed communication networks (MFCN), which holds individual rights of use to the shared spectrum resource. An operator can serve its users by one or more base stations. The GSC supports the entry and storage of shared spectrum resources availability informations and is able to convey the related availability informations to authorized licensed shared spectrum controllers (LSC) and is also able to receive and store acknowledgement informations sent from the LSCs. The GSC also provides means for the NRA to monitor the operation of the shared spectrum system and to provide the shared spectrum system with information on the *Sharing Framework* (set of sharing rules or sharing



conditions for the band, information on spectrum that can be made available for shared use, and the corresponding technical and operational conditions for its use) and the shared spectrum licensees. The GSC ensures that the shared spectrum system operates in conformance with the Sharing Framework and the licensing regime.

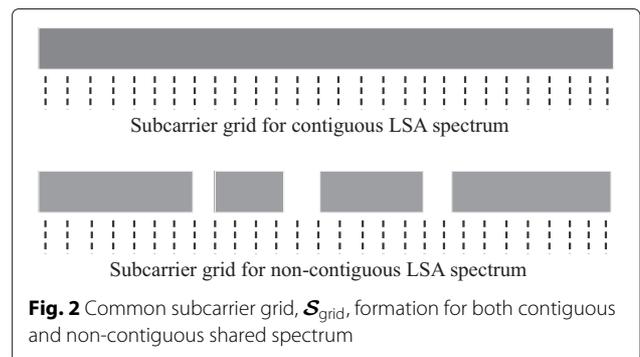
LSC is located within the shared spectrum licensee's domain and enables the shared spectrum licensee to obtain shared spectrum resource availability informations from the GSC and to provide acknowledgment information to the GSC. The LSC interacts with the licensee's MFCN in order to support the mapping of available informations into appropriate radio transmitter configurations and receive the respective confirmations from the MFCN. Each base station is associated with one LSC. Multiple LSCs of the same or different shared spectrum licensee(s) are connected to one GSC.

In our considered system model, a common subcarrier grid, \mathcal{S}_{grid} , is firstly formed. The number of subcarriers N_{sub} and subcarrier spacing Δ_{sub} depend on various parameters such as coherence bandwidth B_{coh} , coherence time T_{coh} , carrier frequency offset (CFO) tolerance CFO_{tol} , Doppler frequency f_D , the size of the smaller band of non-contiguous spectrum B_{low} , if the shareable

spectrum is non-contiguous, etc. The parameter Δ_{sub} is carefully designed such that all the subcarriers in \mathcal{S}_{grid} experience flat fading. In general, B_{coh} and T_{coh} decide the maximum/minimum value of Δ_{sub} . Therefore,

$$\{N_{sub}, \Delta_{sub}\} = f(B_{low}, B_{coh}, T_{coh}, CFO_{tol}, f_D). \quad (1)$$

A uniform common subcarrier grid is assumed, i.e., even if the shared spectrum is non-contiguous, the subcarrier grid covers the whole shared spectrum under operation. The formation of common subcarrier grid for both contiguous and non-contiguous shared spectrums is depicted in Fig. 2.



Each operator first calculates the minimum required spectrum size in terms of Hz or the number of subcarriers, calculated based on its users' data rate requirements and transmitting power budget. Thereupon, it sends this information along with the subcarriers' signal-to-noise ratios (SNRs) of its users to the GSC. The GSC associates the operators with the parameters $\{\rho_1, \rho_2, \dots, \rho_{N_{op}}\}$ that reflect the sharing agreement policy and quantify the operators' priority when the spectrum is shared. We consider that operators utilize this priority scheme in which higher priority incurs in upfront payment to the other operators. Any operator that desires higher priority, for instance, to guarantee more spectrum to its customers, shall compensate the other operators by a predefined fee also proportional to the degree of prioritization. The GSC, thereafter, determines the amount of spectrum for each operator depending on the sharing agreement policy, traffic demands, and subcarrier SNRs. The common rule is that the operator who pays more should have access to a larger amount of shareable spectrum. Let $\{\Delta_{min}^{req,1}, \dots, \Delta_{min}^{req,N_{op}}\}$ denote the minimum amount of spectrums requested by the operators to satisfy their users' service requirements, calculated based on average subcarrier SNR and transmitting power constraint, where $\{\mathbf{q}_1, \dots, \mathbf{q}_{N_{op}}\}$ represent the sets of channel quality informations, i.e., subcarrier SNRs, calculated by different operators. The shared spectrum allocation to different operators is obtained as

$$\{\mathcal{S}_1, \dots, \mathcal{S}_{N_{op}}\} = f\left(\{\rho_1, \dots, \rho_{N_{op}}\}, \{\mathbf{q}_1, \dots, \mathbf{q}_{N_{op}}\}, \{\Delta_{min}^{req,1}, \dots, \Delta_{min}^{req,N_{op}}\}\right), \tag{2}$$

where $\{\mathcal{S}_1, \dots, \mathcal{S}_{N_{op}}\}$ are the sets of subcarriers assigned to spectrum sharing operators and $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \dots \cup \mathcal{S}_{N_{op}} = \mathcal{S}_{grid}$. Note that the subcarriers in \mathcal{S}_n can be from a small fragment (a contiguous band) of shared spectrum or can be scattered over the whole shared spectrum.

Let us consider that spectrum sharing operator n supports K_n non-cooperative users with a single receiving antenna each. The data transmissions of different users are assumed to be subject to slowly varying, independent frequency-selective Rayleigh fading. Perfect channel state information is assumed to be available and a non-sharing subcarrier allocation scheme is considered, i.e., a subcarrier can be allocated to a single user only. The data transmission is subject to regulated maximum transmitting power constraint, $P_{Max}^{(n)}$. Let us consider that $||\mathcal{S}_n|| = L_n$ and $B_n = \Delta_{sub}L_n$ is the total frequency bandwidth allocated to operator n . Then, the capacity achieved by user k of operator n when transmitting data over subcarrier l is given by

$$r_{k,l}^{(n)} = \log_2 \left(1 + p_{k,l}^{(n)} h_{k,l}^{(n)} \right), \tag{3}$$

with $h_{k,l}^{(n)} = |z_{k,l}^{(n)}|^2 / \sigma_n^2$, where $z_{k,l}^{(n)}$ defines the frequency gain on subcarrier l of user k and $||\mathbf{x}||$ denotes the cardinality of \mathbf{x} . $\sigma_n^2 = N_0 B_n / L_n$ is the variance of additive white Gaussian noise (AWGN) over subcarrier l , where N_0 is the noise power spectral density. The quantity $h_{k,l}^{(n)} = |z_{k,l}^{(n)}|^2 / \left(N_0 \frac{B_n}{L_n} \right)$ is defined as the effective SNR on subcarrier l allocated to user k of operator n . $p_{k,l}^{(n)}$ is the amount of power allocated to user k corresponding to subcarrier l . The total rate achieved by user k of operator n is given by $R_k^{(n)} = \sum_{l=1}^{L_n} c_{k,l}^{(n)} r_{k,l}^{(n)}$, where $c_{k,l}^{(n)}$ is the subcarrier assignment function. $c_{k,l}^{(n)} = 1$ refers to the subcarrier allocation in which user k of operator n is assigned with subcarrier l . If subcarrier l is not assigned to user k , $c_{k,l}^{(n)}$ is equal to 0.

3 Solutions for inter-operator spectrum sharing

In our proposed inter-operator spectrum sharing approaches, spectrum sharing among the operators is transformed into an optimization program which maintains a fair priority requirements based on the sharing policy and traffic demands of the operators while maximizing the total system throughput. The spectrum allocation among the operators is performed in such a way that coarsely fulfills the relationship as follows

$$\frac{||\mathcal{S}_i||}{||\mathcal{S}_j||} \approx \frac{\rho_i^{act}}{\rho_j^{act}} \quad \forall i, j \in \mathcal{N}, i \neq j, \tag{4}$$

where $\mathcal{N} \triangleq \{1, 2, \dots, N_{op}\}$ and $\sum_{n=1}^{N_{op}} \rho_n^{act} = 1$. The relationship defined in (4) states that the amount of spectrum resources allocated to the operators are proportional to each other, and ρ_n^{act} defines the active priority measure of operator n . Note that $\rho_i, \forall i, i \in \mathcal{N}$ are the original priority measures of the operators depending solely on sharing rules and mutual agreement. While $\rho_i^{act}, \forall i, i \in \mathcal{N}$ are the active priority measures calculated considering additionally the current traffic demands, $\rho_i^{act}, \forall i, i \in \mathcal{N}$ decide the final spectrum allocation. The values of $\rho_i^{act}, \forall i, i \in \mathcal{N}$ may or may not be equal to the values of $\rho_i, \forall i, i \in \mathcal{N}$.

3.1 Subcarrier gain-based spectrum sharing

The proposed subcarrier-based inter-operator spectrum sharing approach is iterative in nature, and opts to fairly and flexibly allocate the available shareable spectrum among the operators. The subcarriers allocated to any operator scatter over the whole subcarrier grid. We can express the spectrum allocation problem as

$$\begin{aligned} & \max_{\{S_n\}} \sum_{n=1}^{N_{op}} \sum_{k=1}^{K_n} R_k^{(n)} \\ & \text{subject to } \frac{\|S_i\|}{\|S_j\|} \approx \frac{\rho_i^{act}}{\rho_j^{act}} \quad \forall i, j \in \mathcal{N}, i \neq j. \\ & S_1 \cup S_2 \cup \dots \cup S_{N_{op}} = S_{grid}. \end{aligned} \quad (5)$$

The GSC calculates an active set of $\{\rho_1, \rho_2, \dots, \rho_{N_{op}}\}$, defined as $\{\rho_1^{act}, \rho_2^{act}, \dots, \rho_{N_{op}}^{act}\}$ before entering into the actual iterative process. The determination of this active set and the iterative spectrum allocation process is provided in Fig. 3. The GSC calculates the active set of sharing parameters by following the process described in *phase 1* and finally the spectrum allocation among the operators is carried out by following the process illustrated in *phase 2*.

The spectrum allocation problem with the operators' desired amount of spectrum and preferred sets of subcarriers is transformed into an optimization program which maintains a proportional spectrum fairness requirements among the operators. The amount of spectrum and subcarriers are assigned to the operators in such a way that coarsely/closely fulfills the relationship given in (5). The advantage of introducing these normalized priority spectrum measures $\{\rho_1^{act}, \rho_2^{act}, \dots, \rho_{N_{op}}^{act}\}$ is that we can explicitly control the spectrum allocation ratio among the operators during an iterative allocation process. At any particular iteration, the operator i has the opportunity to get assigned with a subcarrier that has the highest SNR over all of its users if it complies with the condition:

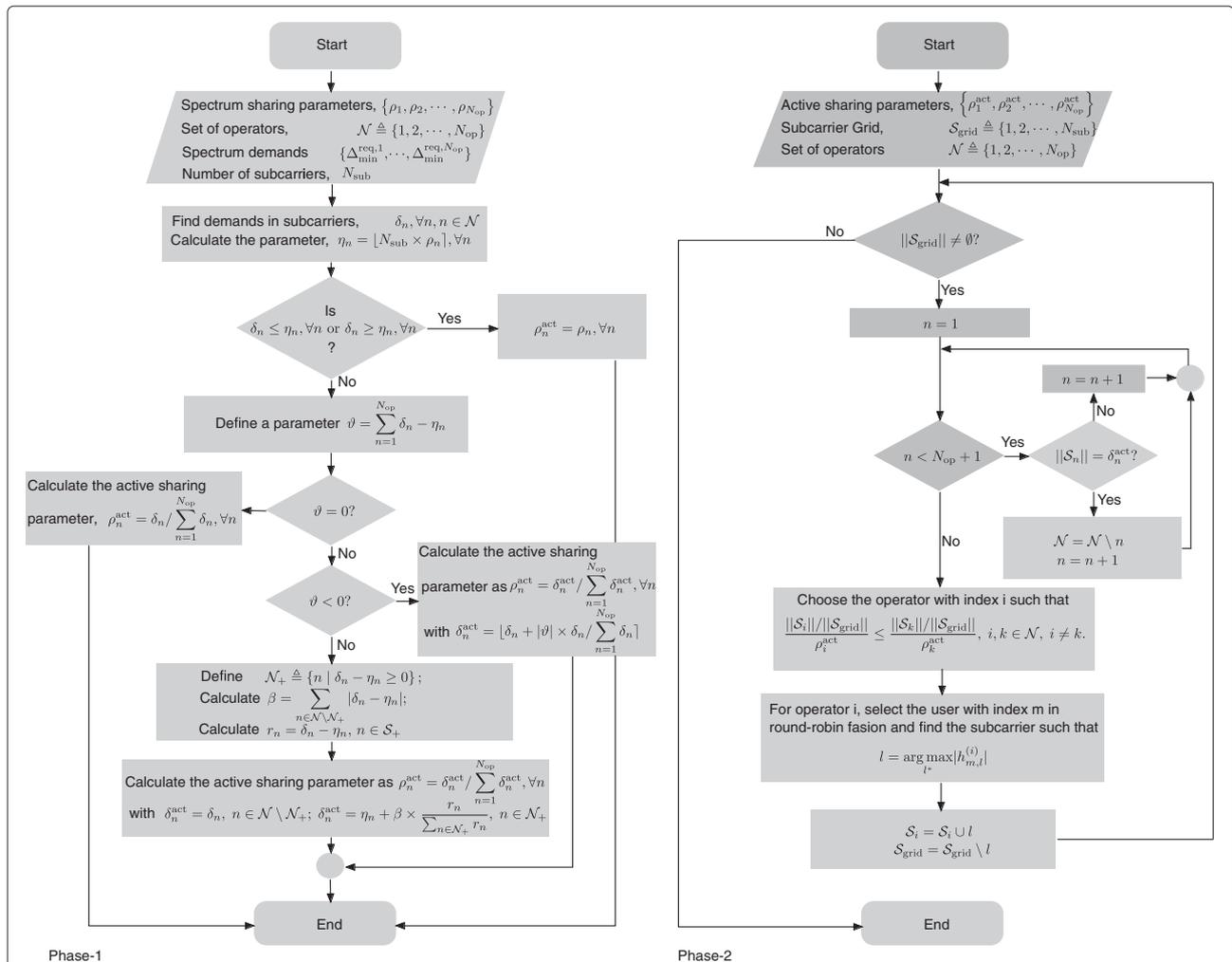


Fig. 3 Proposed subcarrier gain-based resource sharing in shared spectrum access communication systems. Phase 1 corresponds to the process for calculating the active set of sharing parameters $\{\rho_1^{act}, \rho_2^{act}, \dots, \rho_{N_{op}}^{act}\}$, and phase 2 corresponds to the resource allocation process among the spectrum sharing operators based on calculated $\{\rho_1^{act}, \rho_2^{act}, \dots, \rho_{N_{op}}^{act}\}$ in phase 1. Note that for fragmentation-based spectrum sharing, phase 1, remains the same

$$\frac{\|\mathcal{S}_i\|/\|\mathcal{S}_{\text{grid}}\|}{\rho_i^{\text{act}}} \leq \frac{\|\mathcal{S}_k\|/\|\mathcal{S}_{\text{grid}}\|}{\rho_k^{\text{act}}}, \quad i, k \in \mathcal{N}, \quad i \neq k. \quad (6)$$

Hence, this proposed spectrum allocation scheme assigns subcarriers to the operators depending on their desired amount of spectrum and the sharing policy.

A set of predefined rules is followed by the GSC for dynamic and fair allocation of shared spectrum. The amount of shared spectrum allocated to any operator n , δ_n (calculated based on $P_{\text{Max}}^{(n)}$ and average subcarrier SNR $\bar{h}^{(n)}$) scales with its traffic demand under some fairness measures. If the desired amounts of shared spectrum in terms of number of subcarriers, $\delta_n, \forall n, n \in \mathcal{N}$ are \geq or \leq their actual priority amounts $\eta_n, \forall n, n \in \mathcal{N}$ corresponding to the sharing agreement ($\{\rho_1, \rho_2, \dots, \rho_{N_{\text{op}}}\}$), the GSC just allocates the subcarriers proportionally according to the actual sharing agreement policy, i.e., $\rho_n^{\text{act}} = \rho_n, \forall n$.

When for some operators, $\delta_i < \eta_i, i \in \mathcal{N}$, and for the rest of the operators $\delta_j \geq \eta_j, j \in \mathcal{N}$ with $i \neq j$, then if $\sum_{n=1}^{N_{\text{op}}} \delta_n - \eta_n \leq 0$, the GSC allocates the spectrum according to the desired amount of spectrum irrespective of the sharing agreement as it will not violate the agreement. However, if $\sum_{n=1}^{N_{\text{op}}} \delta_n - \eta_n > 0$, the GSC allocates the additional spectrum due to the operators who have $\delta_n < \eta_n$ among the operators with $\delta_n > \eta_n$ proportionally depending on the values of $\delta_n - \eta_n$. Furthermore, for each operator, the subcarrier assignment is performed in such a way that maintains fairness among its users, i.e., each user gets equal opportunity in a round-robin manner to get assigned with its best subcarriers. Finally, after the GSC finds the subcarrier sets $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{N_{\text{op}}}\}$, it notifies all the spectrum sharing operators. Since all the network operators have access to the whole spectrum band and the users' device has the capability to tune to the whole band, each operator can give access to any user subscribed to one of the sharing network operators. However, in order to enable such spectrum sharing, infrastructure sharing and high coordination is required among the operators. Software defined networking and network function virtualization may be the viable solutions for such sharing in order to enable the system obtaining the benefit from multiuser diversity across the operators' domain.

Note that this subcarrier-based spectrum allocation process is iterative and thus requires significantly more time to find the allocation. If the operating environment is such that the channel is highly frequency-selective, then N_{sub} tends to become larger since the subcarrier spacing will become shorter, which in turn will increase the computation time. Note that in the current study, we have considered that the finest resource granularity for transmission is one subcarrier. However, the finest resource granularity can also be one resource block as in the case of Long Term Evolution-Advanced (LTE-A), which contains a group of successive subcarriers.

3.2 Fragmentation-based spectrum sharing

In order to reduce the computation time, the GSC can perform fragmentation-based shared spectrum allocation. For fragmentation-based spectrum sharing, we follow the same procedures for obtaining the active priority measures, i.e., phase 1 remains the same.

In fragmentation-based spectrum sharing, the operators have the option to inform the GSC about the favorable fragments they like to operate on by sending an extra variable α . The range of values in α depends on the number of operators participating in the spectrum sharing process. For example, if there are only two operators, α can be binary. If any operator wants to transmit on the lower end of the spectrum, it sends 0, or 1, otherwise. The same goes for the second operator. If there are three operators, α is of 2-b size while 00 points to the lower end and 11 points to the upper end of the spectrum. Any operator favors one fragment over other fragments in the shared frequency band depending on the types of applications, e.g., short-range communications, the technologies it operates on and channel propagation characteristics, and its achievable capacity on the fragment. The GSC also takes these features under consideration when it allocates the fragments to different operators.

When two or more operators request for the same fragment, the GSC prioritizes one over the other by judicious evaluation of the abovementioned features. If all the operators offer the same type of applications and have equal priority, then the contention is solved by random selection of one of the operators. When similar contention occurs again in future, the GSC performs the same random selection by ignoring the operator that was selected in the previous contention period. When fragmentation-based spectrum sharing is employed, each operator can have independent network deployment. Each operator can independently adjust its own transmission frame structure in accordance with the use case, traffic type, etc. The operators have also the flexibility to change the number of subcarriers thus the subcarrier spacing, Δ_{sub} within their fragments.

As an example, let us consider that there are only two operators participating in the shared spectrum access communications, and the available spectrum for sharing is contiguous. If both the operators have same applications types, e.g., cellular communications and $\rho_1 = \rho_2 = 0.5$ with $\alpha_1 = 0, \alpha_2 = 1$, respectively. The GSC calculates the active priority measures ρ_1^{act} and ρ_2^{act} in accordance with the traffic demands from the operators and sharing rules. The GSC then partitions the shareable spectrum into two fragments with respect to ρ_1^{act} and ρ_2^{act} and allocates the lower end fragment of size $\delta_1^{\text{act}} \Delta_{\text{sub}}$ Hz to operator 1 and the upper end fragment of size $\delta_2^{\text{act}} \Delta_{\text{sub}}$ Hz to operator 2. However, if α_1 and α_2 are equal, then the GSC randomly assigns the ends of the spectrum to the

operators. If the communications scenario is such that the type of applications and technologies one of the operators operates on, e.g., shorter-range communications, prefers higher end fragment to upper end fragment, the GSC will probably allocate the higher end fragment to that particular operator, even if the other operator seeks the same end. Therefore, fragmentation-based spectrum sharing is a situation-specific mechanism.

As we have already mentioned in the previous sections, the GSC can allocate the shared spectrum among the operators based on fragmentation depending on the operators' desired minimum fragment size and channel quality information. A fragment is defined as a contiguous band, and its bandwidth must be an integer multiple of the subcarrier bandwidth Δ_{sub} and complies with B_{low} and other system parameters. It is worth mentioning that if the fragmentation-based spectrum sharing is employed, then each operator has the flexibility to employ independent radio interface, flexible DFT size, etc. The work in [30] considers a spectrum sharing scenario where different operators employ different radio interfaces, i.e., waveforms. It is also possible to allocate different fragments from non-contiguous bands if the radio interface supports it [31]. If the operators are not synchronized in a way that operators of adjacent spectrum do not transmit at the same time, guard bands are created around each fragment to protect other operators from its out-of-band emission. The minimum size of the fragments depends on the desired guard band overhead, which in turn, depends on the pulse shape being used for data modulation.

Each operator can claim a minimum fragment size to limit the overall guardband overhead given as $\frac{B_{\text{guard}}}{B_{\text{frag}}}$, where B_{guard} is the total amount of spectrum belonging to all the guardbands and B_{frag} is the total amount of frequency spectrum belonging to all usable fragments. If the fragment size is too small, the guardband overhead would be too large. This is calculated based on its waveform especially the out-of-band emission level. From that, the operator calculates the required guardband, and considering the guardband size, the operator then claims a minimum fragment size. The GSC performs the fragmentation according to the traffic loads of each operator and may also depend on the preferred fragments of each operator.

Each operator transmits a signal in the allocated spectrum fragments by activating and deactivating subcarriers of the signal, i.e., only the subcarriers within the allocated spectrum fragments are activated, while the others are not. Since different operators do not necessarily have accurate mutual synchronization, the out-of-fragment radiation power of the signals of each operator has to be taken into account, which causes interference to the other operators. We also consider accurate inter-operator synchronization that can be realized in downlink if the

spectrum sharing operators share the radio access network or through GPS modules used in the base stations. In this proposed system model, each operator can transmit data to its own serving users independently without creating interference (assuming perfect synchronization among the operators or adequate guardbands between the fragments if fragmentation based spectrum sharing is employed) to the users served by other operators.

Both of the proposed inter-operator spectrum sharing solutions are suboptimal. In subcarrier or resource block based resource sharing optimization, the solution aims at optimizing the total system throughput and resource allocation is obtained through an iterative process instead of optimal exhaustive search in order to reduce the computational burden and ensures fairness among the operators as well as the users. In fragmentation-based spectrum sharing optimization, GSC judiciously evaluates the sharing rules and the informations received from the operators and allocates the favorable spectrum fragments accordingly. Note that GSC does not follow any strict mathematical process; instead, it takes dynamic and situation-specific measures to decide the fragment allocation.

4 Solution for intra-operator resource allocation

Soon after receiving the information about spectrum allocations from the GSC, each operator performs intra-operator resource allocation for the users with dissimilar service requirements. In practice, the regulatory scenario enforces a total transmitting or radiated power constraint. Therefore, the base stations of each operator work under maximum transmitting power constraint while satisfying its own users' service requirements. Note that there can be several base stations under one operator. In this study, we consider that each operator has only one base station to make the analysis simple and straightforward. Note that it is enough to consider resource allocation optimization for any particular operator since base station can transmit data to its own serving users independently without creating interference (with perfect synchronization or adequate guardbands between the fragments) to the users served by other operators.

The main objective of resource allocation optimization in this section is to perform efficient resource allocation for users with DC and NDC service requirements under total transmitting power constraint. Let us consider the resource allocation for operator n , for which $|\mathcal{S}_n| = L_n$, and K_1 users out of K_n users belong to NDC service requirements while $K_n - K_1$ users require DC services under the maximum transmitting power constraint $P_{\text{Max}}^{(n)}$. For notational brevity, we ignore the operator index n in the following presentation of the paper. Let us define \mathcal{K} and \mathcal{L} as the sets of $\{1, 2, \dots, K\}$ and $\{1, 2, \dots, L\}$, respectively. Now, the resource allocation optimization problem

can be cast as 0–1 optimization problem, which is given by

$$\begin{aligned}
 & \max_{\{c_{k,l}, p_{k,l}\}} \sum_{k=1}^{K_1} \sum_{l=1}^L c_{k,l} r_{k,l} \\
 & \text{subject to} \\
 & \text{C1: } c_{k,l} \in \{0, 1\}, \quad \forall k, l, k \in \mathcal{K}, l \in \mathcal{L} \\
 & \text{C2: } \sum_{k=1}^K c_{k,l} = 1, \quad \forall l, l \in \mathcal{L} \\
 & \text{C3: } p_{k,l} \geq 0, \quad \forall k, l, k \in \mathcal{K}, l \in \mathcal{L} \\
 & \text{C4: } \sum_{k=1}^K \sum_{l=1}^L c_{k,l} p_{k,l} \leq P_{\text{Max}} \\
 & \text{C5: } \sum_{l=1}^L c_{k,l} r_{k,l} \geq R_k^{\text{target}}, \quad k = K_1 + 1, \dots, K
 \end{aligned} \tag{7}$$

where R_k^{target} is the desired throughput of DC user k . A solution to this problem looks for a partition of groundset $\{1, 2, \dots, L\}$ into K subsets such that the measure associated with the subsets, P_{Max} fulfills some bounds. Note that the problem in (7) is non-linear since we have multiplication of variables in constraints and in the objective function. Also, we need to deal with the integer variables because of non-divisibility of resources (subcarriers); thus, it is a non-convex MINLP. Non-linear constraints are more difficult to handle. Therefore, it is very advantageous to incorporate linear constraints. In our proposed solution, we restrict our optimization model to contain only linear constraints. In the following, we formulate several linearization approaches to transform (7) into a convex program, which is easier to solve than solving a MINLP since combining both non-linearity and integrality can lead the MINLP to be undecidable [32].

For notational brevity and simplified analysis, we define a set \mathcal{M} that contains all the tuples of the indices, i.e., $\{k, l\}$ as $\{\{1, 1\}, \dots, \{1, L\}, \dots, \{k, 1\}, \dots, \{k, L\}, \dots, \{K, 1\}, \dots, \{K, L\}\}$, where \mathcal{M}_t denotes the t th tuple in \mathcal{M} . The corresponding indices in the t th subset are defined as $\{k, l\} = \{\lceil t/L \rceil, t - ((\lceil t/L \rceil - 1)L)\}$, where $\lceil \cdot \rceil$ is the ceiling operator. Similarly, we vectorize the elements $h_{k,l}$, $c_{k,l}$ and $p_{k,l}$, $\forall k, l$, in \mathbf{h} , \mathbf{c} and \mathbf{p} , respectively, in accordance with the sequences of the indices in \mathcal{M} . Now, we can rewrite the objective function as

$$\begin{aligned}
 & \max_{\{c_{\mathcal{M}_t}, p_{\mathcal{M}_t}\}} \sum_{t=1}^{K_1 L} c_{\mathcal{M}_t} \log_2(1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t}) = \log_2 \\
 & \left(\max_{\{c_{\mathcal{M}_t}, p_{\mathcal{M}_t}\}} \prod_{t=1}^{K_1 L} (1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}} \right)
 \end{aligned} \tag{8}$$

Now, due to the monotonicity of logarithmic function, i.e., if $x > y$, $\log_2(x) > \log_2(y)$, the log function in the subsequent optimization problems can be

eliminated. We can rewrite the objective function in (8) as $\max_{\{c_{\mathcal{M}_t}, p_{\mathcal{M}_t}\}} \prod_{t=1}^{K_1 L} (1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}}$, and consequently, the optimization problem in (7) can be expressed as

$$\begin{aligned}
 & \max_{\{c_{\mathcal{M}_t}, p_{\mathcal{M}_t}\}} \prod_{t=1}^{K_1 L} (1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}} \\
 & \text{subject to} \\
 & \text{C1: } c_{\mathcal{M}_t} \in \{0, 1\} \forall t, t = 1, \dots, KL \\
 & \text{C2: } \sum_{t \in \{l, l+kL\}} c_{\mathcal{M}_t} = 1, \forall l, l = 1, \dots, L \\
 & \quad \quad \quad k = 1, \dots, K \\
 & \text{C3: } p_{\mathcal{M}_t} \geq 0, \quad \forall t, t = 1, \dots, KL \\
 & \text{C4: } \sum_{t=1}^{KL} c_{\mathcal{M}_t} p_{\mathcal{M}_t} \leq P_{\text{Max}} \\
 & \text{C5: } \sum_{t=K_1 L+1+(k-1)L}^{K_1 L+L+(k-1)L} c_{\mathcal{M}_t} \log_2(1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t}) \geq R_k^{\text{target}}, \\
 & \quad \quad \quad k = K_1 + 1, \dots, K
 \end{aligned} \tag{9}$$

The objective function is still non-linear in its current form. However, note that since $c_{\mathcal{M}_t} \in \{0, 1\}$, we have the flexibility to transform each multiplicative term in the objective function, $(1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}}$ as

$$(1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}} = 1 + c_{\mathcal{M}_t} p_{\mathcal{M}_t} h_{\mathcal{M}_t}. \tag{10}$$

Obviously, this is not true for all $c_{\mathcal{M}_t}$, it is true only for binary $c_{\mathcal{M}_t}$, which is the case here. We take advantage of this transformational relationship in the linearization process. For each t , we create a new variable $\xi_{\mathcal{M}_t}$, and then add the constraints as given below

$$\xi_{\mathcal{M}_t} \leq 1 + \min\{p_{\mathcal{M}_t}, P_{\text{Max}} c_{\mathcal{M}_t}\} h_{\mathcal{M}_t}. \tag{11}$$

Consider the first case $c_{\mathcal{M}_t} = 0$, which means the value of $(1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t})^{c_{\mathcal{M}_t}}$ should be 1. The linear constraint $\xi_{\mathcal{M}_t} \leq 1 + \min\{p_{\mathcal{M}_t}, P_{\text{Max}} c_{\mathcal{M}_t}\} h_{\mathcal{M}_t}$ in (11) forces $\xi_{\mathcal{M}_t}$ to be ≤ 1 . Now, let us consider the case $c_{\mathcal{M}_t} = 1$, now the quantity should be equal to $1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t}$. Similarly, the constraint in (11) enforces the quantity $\xi_{\mathcal{M}_t}$ to be $\leq 1 + p_{\mathcal{M}_t} h_{\mathcal{M}_t}$. The reason for putting a “ \leq ” sign in (11) instead of the “ $=$ ” sign is to obtain a convex feasible region since a constraint in the form of “*linear/affine \geq concave*” generates a non-convex feasible region, which is a major problem. However, at the optimum, the inequality turns out to be an equality. The geometry of the linearization process of constraint C3 is depicted in Fig. 4a. The blue circles represent the exact or expected values of the variable $\xi_{\mathcal{M}_t}$ for two different cases ($c_{\mathcal{M}_t} = 0$ and $c_{\mathcal{M}_t} = 1$). The red lines show the range of values the variable can take when linearization is employed.

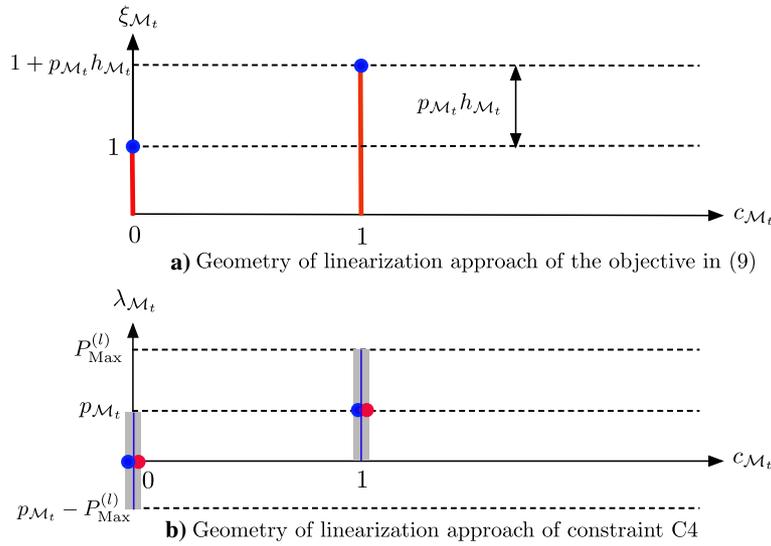


Fig. 4 a–b Geometry for linearization approach of the objective function and constraint C4 in (9)

Furthermore, maximization of the new objective function, $\prod_{t=1}^{K_1L} \xi_{\mathcal{M}_t}$ and maximization of $\left(\prod_{t=1}^{K_1L} \xi_{\mathcal{M}_t}\right)^{1/K_1L}$ will give the same values for the optimization variables. So, we can recast the optimization problem in (9) as the following

$$\max_{\{c_{\mathcal{M}_t}, \xi_{\mathcal{M}_t}, p_t\}} \left(\prod_{t=1}^{K_1L} \xi_{\mathcal{M}_t}\right)^{1/K_1L} \quad (12)$$

subject to

$$\text{C6: } \xi_{\mathcal{M}_t} \leq 1 + \min\{p_{\mathcal{M}_t}, P_{\text{Max}}c_{\mathcal{M}_t}\}h_{\mathcal{M}_t}$$

along with the constraints C1–C5 in (9). Now, the objective of the optimization problem in (12) becomes a concave function.

Now, let us turn our focus to the linearization of the constraint C4 in (9). Here, $c_{\mathcal{M}_t}$ (binary integer) and $p_{\mathcal{M}_t}$ (continuous) are variables in this mathematical program, and we have to deal with their product $c_{\mathcal{M}_t}p_{\mathcal{M}_t}$. If both the variables were continuous, we would have ended up having a quadratic problem, which may create issues with convexity if the quadratic terms appear in constraints. However, constraint C4 is special in the sense that at least one of $c_{\mathcal{M}_t}$ and $p_{\mathcal{M}_t}$ is binary, and the other variable is bounded. Under these assumptions, the product $c_{\mathcal{M}_t}p_{\mathcal{M}_t}$ can be linearized by introducing new slack variables $\lambda_{\mathcal{M}_t}$ and incorporating the following additional linear constraints

$$\begin{aligned} \min\{0, p_{\mathcal{M}_t}^{\text{lb}}\} &\leq \lambda_{\mathcal{M}_t} \leq \max\{0, p_{\mathcal{M}_t}^{\text{ub}}\} \\ p_{\mathcal{M}_t}^{\text{lb}}c_{\mathcal{M}_t} &\leq \lambda_{\mathcal{M}_t} \leq p_{\mathcal{M}_t}^{\text{ub}}c_{\mathcal{M}_t} \\ p_{\mathcal{M}_t} - p_{\mathcal{M}_t}^{\text{ub}}(1 - c_{\mathcal{M}_t}) &\leq \lambda_{\mathcal{M}_t} \leq p_{\mathcal{M}_t} - p_{\mathcal{M}_t}^{\text{lb}}(1 - c_{\mathcal{M}_t}) \end{aligned} \quad (13)$$

Therefore, we linearize each term in the sum separately. This way we end up with KN of the λ variables, for example $\lambda_{\mathcal{M}_t}$. The steps followed in the linearization of constraint C4 are:

- Introducing a variable $\lambda_{\mathcal{M}_t} = c_{\mathcal{M}_t}p_{\mathcal{M}_t}$ for each product.
- Finding upper and lower bounds for each $p_{\mathcal{M}_t}$.
- Introducing the constraints for each $\lambda_{\mathcal{M}_t}$ as in (13).
- Substituting the $\lambda_{\mathcal{M}_t}$ variables into the constraint $\sum_{t=1}^{KL} c_{\mathcal{M}_t}p_{\mathcal{M}_t} = P_{\text{Max}}$.

Therefore, due to the linearization of C4, we have these following linear constraints to be incorporated in the optimization problem.

$$\text{C4: } \begin{cases} \text{N1: } \lambda_{\mathcal{M}_t} \leq P_{\text{Max}}c_{\mathcal{M}_t}, \forall t, t = 1, \dots, KL \\ \text{N2: } \lambda_{\mathcal{M}_t} \geq 0, \quad \forall t, t = 1, \dots, KL \\ \text{N3: } \lambda_{\mathcal{M}_t} \leq p_{\mathcal{M}_t}, \quad \forall t, t = 1, \dots, KL \\ \text{N4: } \lambda_{\mathcal{M}_t} \geq p_{\mathcal{M}_t} - P_{\text{Max}}(1 - c_{\mathcal{M}_t}), \forall t, t = 1, \dots, KL \\ \text{N5: } \sum_{t=1}^{KL} \lambda_{\mathcal{M}_t} = P_{\text{Max}}, \end{cases} \quad (14)$$

Consider the first case $c_{\mathcal{M}_t} = 0$, which means the product $\lambda_{\mathcal{M}_t} = c_{\mathcal{M}_t} p_{\mathcal{M}_t}$ should be 0. The first pair of inequalities (N1, N2) says $0 \leq \lambda_{\mathcal{M}_t} \leq 0$, forcing $\lambda_{\mathcal{M}_t} = 0$. The second pair of inequalities (N3, N4) says $p_{\mathcal{M}_t} - P_{\text{Max}} \leq \lambda_{\mathcal{M}_t} \leq p_{\mathcal{M}_t}$, and $\lambda_{\mathcal{M}_t} = 0$ satisfies those inequalities. Now, consider the case $c_{\mathcal{M}_t} = 1$, so that the product should be $\lambda_{\mathcal{M}_t} = p_{\mathcal{M}_t}$. The first pair of inequalities becomes $0 \leq \lambda_{\mathcal{M}_t} \leq P_{\text{Max}}$, which is satisfied by $\lambda_{\mathcal{M}_t} = p_{\mathcal{M}_t}$. The second pair says $p_{\mathcal{M}_t} \leq \lambda_{\mathcal{M}_t} \leq p_{\mathcal{M}_t}$, forcing $\lambda_{\mathcal{M}_t} = p_{\mathcal{M}_t}$ as desired. This linearization approach, in particular, equates to splitting the feasible regions into two subregions, one where $c_{\mathcal{M}_t} = 0$ and $f(c_{\mathcal{M}_t}, p_{\mathcal{M}_t}) = c_{\mathcal{M}_t} p_{\mathcal{M}_t} = 0$ (trivially linear) and the other where $c_{\mathcal{M}_t} = 1$ and $f(c_{\mathcal{M}_t}, p_{\mathcal{M}_t}) = p_{\mathcal{M}_t}$ (also linear).

The geometry of the linearization process of constraint C4 is depicted in Fig. 4b. The blue circles represent the expected values of the optimization variable $\lambda_{\mathcal{M}_t}$. The gray shaded areas represent the range of values that the variable can take due to linearization process. The red circles stands for the obtained values of the optimization variable. Note that the expected and obtained values of $\lambda_{\mathcal{M}_t}$ are perfectly matched, i.e., exact linearization is obtained.

Now, we are left with the constraint C5 of (9), which is also non-linear. Following the relationship in (10), constraint C5 can be stated as

$$\prod_{t=K_1L+1+(k-1)L}^{K_1L+L+(k-1)L} (1+c_{\mathcal{M}_t} p_{\mathcal{M}_t} h_{\mathcal{M}_t}) \geq 2^{R_k^{\text{target}}}, k=K_1+1, \dots, K \quad (15)$$

which further can be equivalently stated as the following

$$\text{C5: } \left(\prod_{t=K_1L+1+(k-1)L}^{K_1L+L+(k-1)L} \xi_{\mathcal{M}_t} \right)^{1/L} \geq 2^{R_k^{\text{target}}/L}, k=K_1+1, \dots, K \quad (16)$$

where the left side of (16) is the geometric-mean of the optimization variables $\xi_{\mathcal{M}_t}$ and it is concave. After performing linearization and/or convexification of the non-linear constraints (outer approximating the feasible region) and the objective function, the MINLP optimization problem in (7) now becomes a convex, which is comparatively much easier to solve.

5 Performance analysis

In all simulation results presented in this section, the wireless channel is modeled as a frequency-selective channel consisting of six independent Rayleigh multipaths. The multipath components are modeled by Jakes' flat fading model [33]. The power delay profile is exponentially decaying with $e^{-\alpha l}$ with $\alpha = 1$ (decay factor), where l

defines the multipath index. A maximum delay spread of 5 μs and maximum doppler of 30 Hz are assumed. The relative power of the six multipath components are [0, -4.35, -8.69, -13.08, -17.43, -21.78] dB. The power spectral density of AWGN is -170 dBm/Hz. We also assume that the width of the shared spectrum is 10 MHz and it is contiguous. A common subcarrier grid is generated with 512 subcarriers and the operators employ OFDMA for carrying users' data. In the performance analysis for fragment-based spectrum allocation, all the operators employ same subcarrier spacing and radio interfaces with OFDM waveforms, however, the DFT sizes vary depending on the sizes of frequency fragments (i.e., depending on the number of subcarriers as fragments are integer multiples of subcarrier bandwidth) the operators are assigned with. For example, let us say, during one scheduling, the cardinalities of the sets of subcarriers (contiguous) assigned to three spectrum sharing operators are given by $|\mathcal{S}_1| = 112$, $|\mathcal{S}_2| = 213$, and $|\mathcal{S}_3| = 187$. Then, the DFT sizes employed by operator 1, operator 2, and operator 3 are 128, 256, and 256, respectively. Note that the DFT size is chosen as a power of 2 greater than or equal to the number of subcarriers one operator is assigned with. On the other hand, in subcarrier gain-based allocation, all the operators employ a DFT size of 512 irrespective of the number of subcarriers each operator is assigned with since all the operators operate over the whole shared spectrum. Each BS may perform subchannelization of the subcarriers it is assigned with for intra-operator resource allocation if data transmission occurs in terms of resource block. Note that we compare the performance of subcarrier- and fragmentation-based spectrum sharing in terms of achieved throughput. Standard water-filling algorithm is followed to distribute the power optimally among the subcarriers per user per operator. The non-linear optimization solver KNITRO [34] along with mathematical programming modeling language AMPL [35] have been employed to solve the intra-operator resource allocation optimization problem.

In Fig. 5, we have investigated the impact of varying the number of spectrum sharing operators on achieved system throughput and compared the performances of subcarrier gain-based and fragmentation-based spectrum sharing. It can be clearly seen that subcarrier-based spectrum distribution achieves higher throughput, and the performance gap increases as the number of operators is increased under a given set of subcarriers. This is due to there fact that when a large number of operators participate in the shared spectrum access communication, the multiuser diversity improves. Compared to the fragmentation-based spectrum sharing, subcarrier-based spectrum sharing has higher flexibility in allocating the subcarriers to the operators, i.e., to the users who have the highest gains for the subcarriers. The only

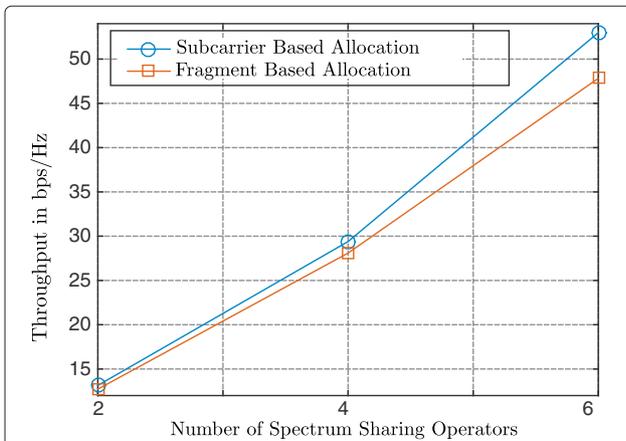


Fig. 5 Impact of varying the number of spectrum sharing operators on achieved system throughput. For simplicity, we have assumed $\rho_1 = \rho_2 = \dots = \rho_{N_{op}}$ with $N_{sub} = 512$. Each operator transmits its signal with 4 W, irrespective of the number of subcarriers it is assigned with. When there is increase in the number of operators, there is linear increase in the total transmit power $\sum_{n=1}^{N_{op}} p_{Max}^{(n)}$. For example, (i) $N_{op} = 2$, total power is 8 W, and (ii) $N_{op} = 6$, total transmission power in the shared spectrum access system is 24 W, regardless of the fact that the number of subcarriers $||S_{grid}||$ remains the same

disadvantageous fact about the subcarrier-based distribution is that the GSC needs to send each operator the indices of the subcarriers, which, in turn, increases the backhaul load. It is worthy to mention that for the performance evaluation in Fig. 5, we have considered a spectrum sharing scenario where all the operators are overloaded and $\rho_1 = \rho_2 = \dots = \rho_{N_{op}}$. Therefore, when there are two operators sharing the spectrum, all the available subcarriers are assigned to two operators equally. When there are more (> 2) operators, the available spectrum is assigned to the operators in accordance with the sharing rules and active priority measures. We have already mentioned that the operators have the flexibility in employing independent radio interface, e.g., multi-carrier waveform, DFT size. Under this circumstance, the operators employ varying DFT sizes depending on the numbers of subcarriers they are assigned with.

In Fig. 6, we study the impact of multiuser diversity on shared spectrum access system in terms of difference in achieved throughputs between the subcarrier gain-based and fragmentation-based spectrum sharing solutions by varying the number of operators and also the number of users per operator. We can clearly distinguish three distinct regions on this figure: (i) *region 1* is determined by the steepest ascent in the performance gap curves and it corresponds to scenarios in which there are few users per operator; (ii) *region 3* is determined by the steepest descent in the performance gap curves and corresponds to scenarios in which the number of users per operator is (almost) maximum; (iii) *region 2* lies within region 1

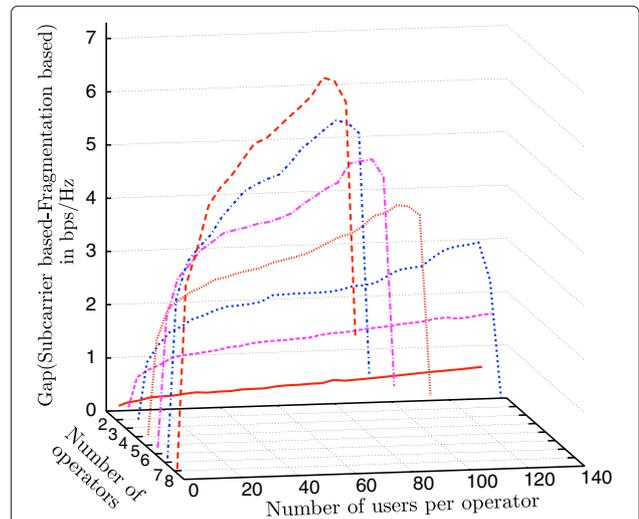


Fig. 6 Multiuser diversity impact on the performances of the proposed spectrum sharing approaches. For simplicity, we have assumed $\rho_1 = \rho_2 = \dots = \rho_{N_{op}}$ with $p_{Max}^{(n)} = 4 W, \forall n$

and region 3 and corresponds to the scenarios most likely found in real systems. Region 1 and region 3 can be treated as boundary regions where both of the proposed spectrum sharing approaches perform almost equally or where the gap is small. That is to say, region 1 represents the, unlikely yet possible, scenario in which a single user gets assigned with all the subcarriers belonging to its operator, for both the subcarrier-based and fragment-based approaches. On the other hand, region 3 should be thought as the scenario in which there are as many users as the number of available subcarriers and each user is assigned with a single subcarrier.

Note the the higher gap between the subcarrier-based and fragment-based allocation methods is achieved when the system has more spectrum sharing operators. This is due to the fact that the set size of available subcarriers inherent to each allocation method decreases with different rates as the number of spectrum sharing operators grows. For instance, although there is throughput improvement due to multiuser diversity under the fragment-based solution, such improvement is significantly affected as each operator gets assigned with a smaller fragment due to the existence of more operators. Therefore, from the whole spectrum sharing communication system perspective, the maximum gap depends on the size of the subcarrier grid, sharing parameters $\{\rho_1, \rho_2, \dots, \rho_{N_{op}}\}$, and the number of spectrum sharing operators. For any particular operator, the gap will solely depend on the set size of the subcarriers it is assigned with and the number of users the operator serves. We can still observe that the analysis shown in Fig. 6 could be useful for an entity managing the shared spectrum in order

to make a judicious selection of allocation method. For instance, suppose that the aforementioned GSC decides that a performance gap within 1 bps/Hz is acceptable. In a region where three operators having up to 20 users each share the licensed spectrum, the fragmentation based allocation method should be utilized since the performance requirement is satisfied with reduced computational time.

We evaluate the achieved spectrum efficiency by employing the proposed linearization-based intra-operator resource allocation approach for users with DC and NDC service requirements in Fig. 7 and compare it with the original MINLP-based solution and other existing methods. Here, two out of the total four users have DC service requirements of 1.4 and 1.6 bps/Hz, respectively. Note that the spectrum efficiencies achieved by solving the original MINLP problem and proposed approach are very close. The upper bound curve is obtained by solving the proposed method in [22]. The solution gives an upper bound on the achievable maximum sum rate of all NDC users under the individual rate requirement for each DC user and the total transmit power constraint. But, the proposed convex relaxation technique in [22] permits time-sharing of each subchannel; therefore, the system model differs from the original OFDMA system since it does not perform mutually exclusive subchannel assignment. If we strictly follow non-overlapping subcarrier allocation, the correct bound is MINLP solution. That is to say that the optimal solution of the MINLP is also the optimal solution of the linearized problem, i.e., the upper bound is given by the MINLP solution. In other words, for integer programming problems (i.e., $c_{M_t} \in \{0, 1\}$), MINLP solution evolves as the upper bound.

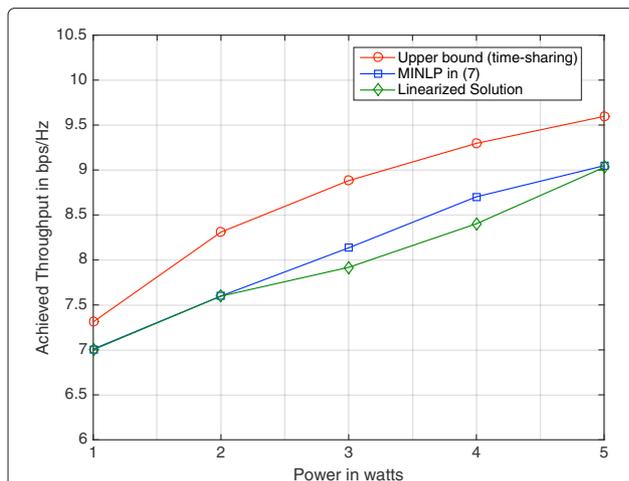


Fig. 7 Comparison of spectral efficiencies achieved by solving the original MINLP and solving our proposed linearized formulation for an operator with four (two DC users and two NDC users) users and eight subchannels

Exact solution of the linearized and convex optimization problem depends on factors like tolerance of the feasibility error and relative optimality tolerance of the solver. Smaller values of absolute feasibility tolerance and relative optimality tolerance result in a higher degree of accuracy in the solution with respect to the feasibility and optimality, respectively, but the solution will be more expensive. We have kept the default values for tolerance of the feasibility error and optimality error, which is $1.0e-6$ for both the cases. As this value is quite small, even if we cannot guarantee the optimality, we can at least claim that the solution provided by the solver is very close to the optimal solution.

It can be noticed that there exists a difference in performance between the original MINLP and linearized program solutions. One possible reason for this performance gap is that although in Fig. 4b we certainly obtain perfect matching of the expected and obtained values of the slack variables, we may not obtain such perfect matching in Fig. 4a. Furthermore, it should be noted that the obtained values of the slack variables in Fig. 4a are always lower than the expected/maximum values. Therefore, when the MINLP is solved, it is very likely that the expected values will be obtained. However, when the linearized problem is solved, the obtained values may not always match the expected values. Since, the obtained values are always \leq the expected values, there may exist a difference in performance of the linearized/convexified and MINLP solutions. In order to find the probable cause responsible for this gap, we followed the following concept: after optimization, if the slack variable ξ is strictly smaller than the right side of (11), it would be easy to increase ξ until the inequality in (11) becomes an equality. This would eventually benefit the objective function, and it would not violate any constraint. The constraint C5 in (16) will even benefit from an increase of ξ . However, we found that the constraint in Eq. (11): $\xi_{M_t} \leq 1 + \min\{p_{M_t}, P_{Max}c_{M_t}\}h_{M_t}$ is not that main cause of the performance loss. Interestingly, we noticed that the values of $\xi_{M_t} - \min\{p_{M_t}, P_{Max}c_{M_t}\}h_{M_t}$ are very close to 1 and, even after increasing ξ until the inequality in (11) becomes an equality, does not really help much in improving the gap. Thus, this gap may be the consequence of numerical inaccuracies of the solver we employed.

From the point of view of computational complexity of the intra-operator resource allocation problem, as far as the worst-case complexity goes, MINLP problems are provably unsolvable [32], but linearized problems are solvable (sometimes, perhaps not very quickly). Furthermore, non-linear programming is more difficult than linear programming, especially when the feasible region is non-convex since it can be hard to determine which points are actually optimal. Therefore, if we have a choice, a linearized formulation is probably a better choice because

the constraints are easier to deal with, and the LP relaxation gives an LP optimum. However, it is not straightforward to prefer one formulation over the other just based on computational complexity. One should however make such decisions based on a judicious analysis of existing tools and algorithms and perform experiments to find out the most efficient approach. In general, in the current state-of-the-art, the linearized formulation tends to win.

6 Conclusions

In this paper, we considered a shared spectrum access communication system and proposed suboptimal solutions both for inter-operator and intra-operator resource allocation problems. For inter-operator spectrum allocation, two efficient algorithms (subcarrier and fragmentation based) are proposed that take care of the mutual sharing policy and fairness issues. Subcarrier-based spectrum allocation scheme has been found to be more efficient in terms of achieved throughput. However, fragmentation-based allocation scheme is more suitable in terms of computational complexity. For the intra-operator resource (spectrum and power) allocation problem, we considered resource allocation for a system with users with delay constraint service requirements and formulated computationally efficient (if not, at least solvable) solutions based on some linearization techniques (exact linearization or linear approximations) considering the structures of the optimization problems and the constraints. The performances of the proposed solutions are impressive when compared to the original MINLP and other existing solutions.

In the current status of the proposed spectrum sharing algorithms, we have not considered the loss of spectral efficiency due to additional guardband being used by the operators that employ inefficient waveforms. Using spectrally inefficient waveforms by some of the operators would decrease the overall shared spectrum access spectral efficiency and at the same time, the GSC could be unfair to the operators who employ spectrally efficient waveforms. Therefore, some mechanisms need to be developed to identify the operators and quantify the loss and thus, penalize the operators employing inefficient waveforms accordingly, which will be investigated in our future works.

In order to simplify the user terminal complexity, the availability of large and contiguous spectrum is preferable. User terminal might require to perform aggregation of multiple frequency bands with different characteristics if large contiguous spectrum for sharing is unavailable. Furthermore, the radio frequency front end of the user terminal needs to be tuneable and configurable to operate at a particular frequency band depending on shareable spectrum availability for its operator to support different spectrum of operation. Therefore, designs of frequency

agile front end for user terminal and flexible air interface for supporting dynamic usage of spectrum need to be investigated.

Acknowledgements

This research was conducted under a contract of R&D for radio resource enhancement, organized by the Ministry of Internal Affairs and Communications, Japan.

Competing interests

The authors declare that they have no competing interests.

Received: 14 September 2015 Accepted: 15 July 2016

Published online: 29 July 2016

References

1. T Irnich, J Kronander, Y Selen, G Li, in *IEEE International Symposium Personal Indoor Mobile Radio Communications*. Spectrum sharing scenarios and resulting technical requirements for 5G systems, (2013), pp. 127–132
2. M Matinmikko, H Okkonen, M Palola, S Yrjola, P Ahokangas, M Mustonen, Spectrum sharing using licensed shared access: the concept and its workflow for LTE-advanced networks. *IEEE Trans. Wireless Commun.* **21**(2), 72–79 (2014)
3. S Yrjola, E Heikkinen, in *IEEE International Conference on Cognitive Radio Oriented Wireless Networking and Communications*. Active antenna system enhancement for supporting Licensed Shared Access (LSA) concept, (2014), pp. 291–298
4. M Mustonen, T Chen, H Saarnisaari, Matinmikko M, S Yrjola, M Palola, Cellular architecture enhancement for supporting the european licensed shared access concept. *IEEE Trans. Wireless Commun.* **21**(3), 37–43 (2014)
5. J Khun-Jush, P Bender, B Deschamps, M Gundlach, in *ETSI Workshop on Reconfigurable Radio Systems*. Licensed shared access as complementary approach to meet spectrum demands: benefits for next generation cellular systems, (2012), pp. 1–5
6. M Gundlach, J Hofmann, C Markwart, E Mohyeldin, in *IEEE International Workshop on Cognitive Cellular Systems*. Recent advances on LSA in standardization, regulation, research and architecture design, (2014), pp. 1–5
7. M Abitbol, PJ Muller, in *IEEE International Symposium on Dynamic Spectrum Access Networks*. Licensed Shared Access: an innovation in European radio spectrum policy, (2014), pp. 259–262
8. M. D Mueck, V Frascolla, B Badic, in *IEEE International Workshop on Cognitive Cellular Systems*. Licensed Shared Access—state-of-the-art and current challenges, (2014), pp. 1–5
9. M Palolo, T Rautio, M Matinmikko, J Prokkola, M Mustonen, M Heikkila, T Kippola, S Yrjola, V Hartikainen, L Tudose, A Kivinen, J Paavola, J Okkonen, M Makelainen, T Hanninen, H Kokkinen, in *IEEE International Conference on Cognitive Radio Oriented Wireless Networking and Communications*. Licensed Shared Access (LSA) trial demonstration using real LTE network, (2014), pp. 498–502
10. P Ahokangas, M Matinmikko, S Yrjola, M Mustonen, H Posti, E Luttinen, A Kivimaki, in *IEEE International Symposium on Dynamic Spectrum Access Networks*. Business models for mobile network operators in Licensed Shared Access (LSA), (2014), pp. 263–270
11. E Perez, K J Friederichs, I Viering, D J Naranjo, in *IEEE International Conference on Cognitive Radio Oriented Wireless Networking and Communications*. Optimization of authorised/licensed shared access resources, (2014), pp. 241–246
12. J Luo, J Eichinger, Z Zhao, E Schulz, in *IEEE International Symposium on Dynamic Spectrum Access Networks*. Multi-carrier waveform based flexible inter-operator spectrum sharing for 5G systems, (2014), pp. 449–457
13. L Anchora, L Badia, E Karipidis, M Zorzi, in *International Symposium on Wireless Communications Systems*. Capacity gains due to orthogonal spectrum sharing in multi-operator lte cellular networks, (2012), pp. 286–290
14. H Kamal, M Coupechoux, P Godlewski, in *IEEE International Symposium on Personal Indoor Mobile Radio Communications*. Inter-operator spectrum sharing for cellular networks using game theory, (2009), pp. 425–429

15. M Bennis, S Lasaulce, M Debbah, Inter-operator spectrum sharing from a game theoretical perspective. *EURASIP J. Adv. Signal Process.* **2009**, 295739 (2009)
16. W Rhee, JM Cioffi, in *IEEE Vehicular Technology Conference*. Increase in capacity of multiuser OFDM system using dynamic subchannel allocation, (2000), pp. 1085–1089
17. J Jang, KB Lee, Transmit power adaptation for multiuser OFDM systems. *IEEE J. Selected Areas Commun.* **21**(2), 171–178 (2003)
18. W Yu, R Lui, Dual methods for non-convex spectrum optimization of multicarrier systems. *IEEE Trans. Wireless Commun.* **54**(7), 1310–1322 (2006)
19. DWK Ng, ES Lo, R Schober, Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity. *IEEE Trans. Wireless Commun.* **11**(10), 3618–3631 (2012)
20. CY Wong, RS Cheng, KB Lataief, RD Murch, Multiuser OFDM system with adaptive subcarrier, bit, and power allocation. *IEEE J. Selected Areas Commun.* **17**, 1747–1758 (1999)
21. C Bae, D-H Cho, Fairness-aware adaptive resource allocation scheme in multihop OFDMA systems. *IEEE Commun. Lett.* **11**, 134–136 (2007)
22. M Tao, Y-C Liang, F Zhang, Resource allocation for delay differentiated traffic in multiuser OFDM systems. *IEEE Trans. Wireless Commun.* **7**(6), 2190–2201 (2008)
23. YM Tsang, RS Cheng, in *IEEE Wireless Communications and Networking Conference*. Optimal resource allocation in SDMA/MIMO/OFDM systems under QoS and power constraints, (2004)
24. Z Hu, G Zhu, Y Xia, G Liu, in *IEEE Wireless Communications and Networking Conference*. Multiuser subcarrier and bit allocation for MIMO-OFDM systems with perfect and partial channel information, (2004), pp. 1188–1193
25. P Xia, S Zhou, G Giannakis, Adaptive MIMO-OFMA based on partial channel state information. *IEEE Trans. Signal Process.* **52**, 202–213 (2004)
26. Z Shen, JG Andrews, BL Evans, Adaptive resource allocation in multiuser OFDM systems with proportional data rate fairness. *IEEE Trans. Wireless Commun.* **4**(6), 2726–2737 (2005)
27. IC Wong, Z Shen, J. G Andrews, B. L Evans, in *IEEE International Workshop on Signal Processing Systems*. A low complexity algorithm for proportional resource allocation in OFDMA systems, (2004)
28. S Sadr, A Anpalagan, K Raahemifar, in *IEEE Vehicular Technology Conference*. A novel subcarrier allocation algorithm for multiuser OFDM system with fairness: User's perspective, (2007), pp. 1772–1776
29. C Mohanram, S Bhashyam, A sub-optimal joint subcarrier and power allocation algorithm for multiuser OFDM. *IEEE Commun.Letts.* **9**(8), 685–687 (2005)
30. MG Kibria, GP Villardi, K Ishizu, F Kojima, in *IEEE Wireless Personal Multimedia Communications*. Coexistence of systems with different multicarrier waveforms in LSA communications, (2015), pp. 1–5
31. J McMenamy, I Macaluso, N Marchetti, L Doyle, in *IEEE International Symposium on Dynamic Spectrum Access Networks*. A methodology to help operators share the spectrum through an enhanced form of carrier aggregation, (2014), pp. 334–345
32. RC Jeroslow, There cannot be any algorithm for integer programming with quadratic constraints. *Oper Res.* **21**(1), 221–224 (1973)
33. TS Rappaport, *Wireless Communications: Principles and Practice*. (NJ: Prentice-Hall, Upper Saddle River, 2002)
34. RH Byrd, J Nocedal, RA Waltz, *KNITRO: an integrated package for nonlinear optimization, large-scale nonlinear optimization*, vol. 83. (Springer-Verlag, 2006), p. 35–59
35. R Fourer, DM Gay, BW Kernighan, AMPL: a modeling language for mathematical programming. *Manag. Sci.* **36**, 519–554 (1990)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
