IEEE Open Journal of the Communications Society

Received 30 July 2020; revised 31 August 2020; accepted 7 September 2020. Date of publication 21 September 2020; date of current version 13 October 2020.

Digital Object Identifier 10.1109/OJCOMS.2020.3024778

Unified User Association and Contract-Theoretic Resource Orchestration in NOMA Heterogeneous Wireless Networks

MARIA DIAMANTI[1], GEORGIOS FRAGKOS[2] (Graduate Student Member, IEEE), EIRINI ELENI TSIROPOULOU[2] (Member, IEEE), AND SYMEON PAPAVASSILIOU[1] (Senior Member, IEEE)

[1]School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Zografou, Greece

[2]Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA

CORRESPONDING AUTHOR: S. PAPAVASSILIOU (e-mail: papavass@mail.ntua.gr)

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty Members and Researchers and the Procurement of High-Cost Research Equipment Grant" under Project HFRI-FM17-2436. The work of Georgios Fragkos and Eirini Eleni Tsiropoulou was supported by the NSF under Grant CRII-1849739.

ABSTRACT Lately, the deployment of heterogeneous wireless networks has emerged, as part of the 5G vision, to cope with the users' exaggerated service demands. In this context, the application of Non-Orthogonal Multiple Access (NOMA) technique constitutes a promising solution to facilitate a balance between spectral efficiency and system complexity. In this article, we consider the problem of joint user association and uplink power allocation, in heterogeneous 5G wireless networks, employing NOMA technology. The coupled problem is treated under an incomplete information scenario, where the Base Stations (BSs) have statistical only knowledge of the users' channel conditions. To deal with the incompleteness of Channel State Information (CSI), a Contract Theory (CT) based approach is introduced. A Reinforcement Learning (RL) based methodology, capitalizing on the provided feedback from the communication environment is initially adopted, in order to achieve the users to BS association in an iterative and distributed manner. The problem of uplink power allocation is subsequently formulated as a contract between each BS and its corresponding users. The optimal power is thus obtained as the solution of the optimization of each BS's utility function, while ensuring the optimality of the utility function of each associated user, given the unique communications characteristics and type of each user. Detailed numerical evaluation of the performance of the proposed unified user association and power allocation framework is provided, via modeling and simulation, illustrating its operation, features and benefits, under densely deployed heterogeneous environments.

INDEX TERMS Contract theory (CT), heterogeneous networks, incomplete information, non-orthogonal multiple access (NOMA), power allocation, reinforcement learning (RL), user association.

I. INTRODUCTION

The advent of 5G technology gives rise to an increased usage of bandwidth-intensive and delay-sensitive services, competing for the scarce radio resources [1]. To this end, the introduction of small cells to the existing macrocell-based infrastructure has provided an effective shift in wireless network architecture. Low power Base Stations (BSs), such as femto, pico or micro BSs with a coverage area of a few tens to hundreds of meters, increase network capacity, while providing the required Quality of Service (QoS) [2]. The heterogeneity of network infrastructure is further promoted by the advancement in Networked Flying Platform (NFP)-assisted communications. Due to their flexible deployment, strong Line-of-Sight (LOS) links and controlled mobility [3], NFPs can serve as a means of temporary data traffic management and provide coverage in hotspot areas or emergency situations [4]. Nevertheless, cell densification is yet insufficient by itself to address the spectrum scarcity problem, calling for intelligent radio access and resource management approaches. To further fill the requirements for spectrum efficiency, resource usage fairness and massive connectivity, Non-Orthogonal Multiple Access (NOMA) provides one of the most promising radio access techniques in next-generation wireless communications. In contrast to conventional Orthogonal Multiple Access (OMA) techniques, where users occupy dedicated resources [5], NOMA allows multiple users to multiplex in the power domain over the same time/frequency/code resources, preventing the under-utilization of system bandwidth [6], [7]. Advanced physical layer and multi-user detection techniques, such as Successive Interference Cancellation (SIC), are then, applied at the receiver to decode the received superimposed signal and deal with the interference problem. In the case of code-domain multiplexing, different users are allocated different codes and multiplexed over the same time/frequency resources, such as multiuser shared access, sparse code multiple access, and low-density spreading [8]. Furthermore, the interference alignment technique has been proposed in the literature towards improving the users' achievable data rate according to the interference levels and channel conditions [9], especially in the case of multiple-input multiple-output (MIMO) channels [10], [11].

Despite the plethora of research works that have adopted and exploited the application of NOMA in heterogeneous wireless networks, there are still important challenges to be addressed, pertinent to its realization. In view of the intrinsic interference caused by different users transmitting over the same resources, advanced interference management techniques are crucial to guarantee the performance gain of NOMA. The latter, in turn, heavily depend on the coupled problems of user scheduling and power control, stressing the need for joint resource optimization schemes. However, one of the key challenges faced by the existing resource allocation approaches, directly affecting their effectiveness and applicability, is the lack of perfect knowledge of the Channel State Information (CSI) at the base station. Owing to the inherent uncertainty introduced by the rapidly varying channels and the increased backhaul signaling overhead induced by the deluge of user devices, perfect CSI is practically difficult to be achieved [12]. Nevertheless, when considering the situation where only statistical CSI is available at the BS, the private information of the user device regarding its experienced channel gain could be utilized to steer the resource allocation procedure. In that sense, the problem of resource allocation under statistical CSI can be modeled and treated as an incomplete information problem [13].

To this end, a synergistic approach between the BSs and their serving user devices is necessary to accommodate the incompleteness of CSI on BSs' behalf. Especially, a holistic user-centric resource allocation mechanism could additionally ameliorate users' satisfaction and facilitate temporal network deployments (i.e., NFPs), against a continuously evolving heterogeneous environment. This poses the need to scrutinize novel formal methods, in the field of wireless networks, to mathematically formulate and treat the timely resource allocation problem of joint user association and power allocation from the users' perspective. Our work in this article aims exactly at addressing this problem, by capturing and modeling the relationships among the actors involved in such an interdependent NOMA-based wireless system, while properly driving their behavior to mutual beneficial operation points.

A. RELATED WORK

Driven by the fundamental concept of NOMA to multiplex users in the power-domain, several research works have been devoted, so far, to the essential problem of power control in NOMA-operated networks. Early work by the authors in [14] introduced a fixed power allocation strategy that was further extended in [15] to account for the distinct channel state of each multiplexed user. Towards overcoming the drawbacks of fixed power allocations, dynamically adjusted power allocation schemes were designed in [16], [17] for two-user and multiple-user NOMA systems, respectively. Based on this groundwork, more recent efforts targeted at addressing various joint resource allocation problems in NOMA-based networks, such as joint rate and power control [18], and joint spectrum and power control [19].

Nonetheless, when employing NOMA in a multi-cell network scenario, the interdependence between user to cell association and power allocation should be taken into account. Centralized solutions that iteratively optimize user association and power allocation have been presented in [20], [21], based on game theory (e.g., coalition formation and matching games) and advanced optimization techniques. Considering that such centralized approaches usually fail to apply in more complex systems, significant attention has also be drawn in the design of distributed solutions. A semi-distributed approach towards user association, transmission mode selection, and power allocation has been developed in [22]. Moreover, a cluster formation and power-bandwidth allocation algorithm executed exclusively by each cell is proposed in [23]. Despite their distributed structure and reduced computational complexity, both works in [22], [23] optimize the subsequent problems under investigation at different stages and independent of each other, instead of jointly addressing them. More importantly, all aforementioned works [20]–[23] presume perfect CSI at the BSs during the resource allocation procedure, which significantly limits their exploitability and applicability.

Meanwhile, the issue of imperfect/partial CSI has received considerable attention to secure the performance gain of NOMA in practical implementations. Based on the long-term statistics of channel realizations, stochastic methods that model either the CSI or the channel estimation error have been extensively used in the literature to evaluate the performance of NOMA and to design effective resource allocation schemes. Primary efforts focused on identifying the impact of partial CSI on different performance metrics, such as the outage probability and the average sum rate [24], or the user fairness [17]. Power allocation strategies under statistical CSI have also been derived by the works in [25], [26], while more combinatorial resource allocation frameworks for networks operating under Multi-Carrier NOMA (MC-NOMA) have been proposed in [27], [28]. All the aforementioned resource allocation works (i.e., [25]–[28]) formulate centralized optimization problems using probabilistic constraints on some outage event. Following a series of observations and appropriate simplifications, the constraints are subsequently reduced, and the corresponding problems are transformed into non-probabilistic ones.

An alternative formal method to mathematically formulate resource allocation problems under statistical CSI, that promotes users' involvement in the allocation procedure, has been introduced in the literature of wireless networks, based on the concept of Contract Theory (CT). Contract theory is a field of economics that provides the mathematical foundations to create mutually agreeable contracts or arrangements between economic players in presence of complete or incomplete information (often referred to as asymmetric information) [29]. Under this concept, a principal/employer creates contract bundles based on agents'/employees' private information, i.e., type, to motivate them provide back their effort and hence, reveal their actual type. In wireless communications, contract theory has been already used to provide mutual agreements between network/service providers and user devices. Exemplified works in [30], [31] have applied contract theory in Device-to-Device (D2D) and cognitive communications, respectively, to incentivize users' participation and contribution towards enhancing network's capacity.

Consequently, contract theory constitutes a particularly prominent way to design resource allocation schemes that consider imperfect/partial CSI, under a user-centric flavor. Preliminary contract-based solutions of relevant resource allocation schemes have been discussed in [13], [32], [33]. Both [32] and [33] model the problem of relay node selection in wireless networks, while each research work employs a different multiple access technique, namely NOMA and Orthogonal Frequency Division Multiple Access (OFDMA), respectively. In these settings, the CSI of the prospective relay nodes is regarded as their private information and only the probability distribution of their types, i.e., channel gain, is known at the BS. Embracing the idea of probabilistic CSI at the BS, the authors in [13] confront the joint user association, and spectrum and power allocation problem from a contract-theoretic perspective, while aiming at mitigating the resulting interference in OFDMA networks. Nevertheless, the OFDMA-based specific formulation assumed in [13] renders this approach inapplicable in NOMA-operated networks, while its centralized nature restricts its applicability in terms of network scalability.

B. CONTRIBUTIONS & OUTLINE

To the best of our knowledge, our work is the first one in the literature that aims at exactly filling the aforementioned research gaps, by removing the corresponding assumptions of perfect CSI knowledge and treating the emerging challenges associated with resource allocation in NOMA-operated heterogeneous networks. In particular, we propose a novel framework that jointly tackles the user-to-BS association and uplink power allocation in heterogeneous wireless networks. The problem is formulated and solved under an incomplete CSI scenario, by introducing formal methods based on Reinforcement Learning (RL) and Contract Theory (CT).

In our approach, the users learn autonomously and in a distributed manner their most beneficial association to the available BSs by observing the provided feedback from the communications environment. Along with the user-to-BS association, the users' uplink transmission power levels are, also, optimized via a contract-theoretic model. More precisely, the users' association in each iteration of the proposed RL mechanism is probabilistically reinforced with respect to their associated BS's network-related and social characteristics. With the term social characteristics, we refer to the long-term BSs' reputation formulated by the users' expressed subjective opinion regarding the service that they enjoy by the BS that they are associated with. The latter is realized in our work through the novel use of the Bayesian Truth Serum mechanism.

Subsequently, considering that perfect CSI is hard to attain, we leverage the principles of contract theory and introduce a power allocation mechanism that operates under the scenario of incomplete CSI. A labor economics-based relationship is developed among each BS and its associated users. The BS rewards the users with personalized rewards that depend on the user's type, which is extracted based on the user's channel gain conditions. Accordingly, appropriate utility functions are formulated for both the BSs and the users. A power control optimization problem of each BS's utility function is introduced, while ensuring the optimality of the utility function of each associated user, given the unique communications characteristics of each user and its type. The above optimization problem is solved and the optimal users' uplink transmission powers are determined.

Based on the above theoretical foundations, we study and demonstrate via modeling and simulation the inherent characteristics of the contract-theoretic uplink power allocation under the cases of complete and incomplete CSI. Moreover, the operation of the RL-based user-to-BS association mechanism is illustrated in terms of the convergence to a beneficial users' association. A detailed comparative numerical evaluation of the proposed approach against other user-to-BS association mechanisms is performed, showing the benefits of the overall proposed framework in terms of power saving, achieved data rate, and fairness among the users within the heterogeneous wireless network.

The remainder of the paper is organized as follows. The overall considered model and assumptions, in terms of

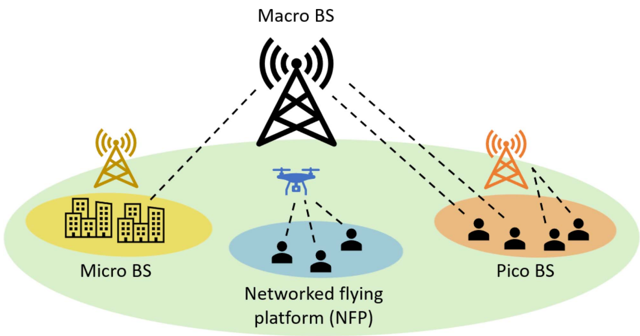


FIGURE 1. Heterogeneous wireless network topology.

TABLE 1. Simulation parameters.

Parameter	Value
MBS's bandwidth	1.8 MHz
PBSs' bandwidth	720 kHz
UAVs' bandwidth	360 kHz
MBS to user path loss	$128.1 + 37.6 \log_{10}(d[km])$ dB
PBS to user path loss	$140.7 + 36.7 \log_{10}(d[km])$ dB
UAV to user path loss	$92.45 + 20 \log_{10}(d[km])$ dB
Shadowing standard deviation	8 dB
AWGN power spectral density	-174 dBm/Hz
Users' max transmit power	23 dBm
Carrier frequency	2 GHz
UAVs' height	20 m

system, network topology and user characteristics, are introduced in Section II. Section III presents the contract-theoretic power allocation design under both complete and incomplete information scenarios. In Section IV, the autonomous user-to-BS association mechanism is introduced via exploiting a reinforcement learning technique and the concept of Bayesian Truth Serum. Also, the unified user association and power allocation algorithm is discussed in the same section. In Section V, a detailed numerical evaluation of the performance of the proposed unified user association and power allocation framework is provided, through modeling and simulation, illustrating its operation, features and benefits. Finally, Section VI concludes the paper.

II. SYSTEM MODEL

We consider the uplink communication of a heterogeneous wireless network consisting of $|C|$ BSs, and let us denote the corresponding set by $C = \{1, \ldots, |C|\}$. Each BS in the network, depending on its physical characteristics, features, and capabilities, can be of different type. Indicative examples of such different types of BSs are macro BSs, small BSs (i.e., micro, pico or femto) or NFPs, as illustrated in Fig. 1. Let $U = \{1, \ldots, |U|\}$ denote the set of users to be served by the network. Subsequently, a set $\mathbb{U}_c$ of cardinality $|\mathbb{U}_c|$ represents the users associated with a BS c at any given time. As a means of efficiently managing complexity and interference mitigation, we assume that the total system available bandwidth B is sub-divided into $|C|$ different orthogonal frequency chunks, where accordingly bandwidth B_c has been allocated to each BS c, such that $B = \sum_{c \in C} B_c$. Moreover, we presume that the allocated frequency chunks (i.e., channels) are exposed to frequency-flat block fading, meaning that they can be considered constant over their bandwidth, but can vary independently with each other. Following NOMA principles, multi-user spectrum sharing within the assigned bandwidth of a specific BS is achieved through power-domain multiplexing, capitalizing on the exploitation of channel gain difference among users. As a result, the users associated with a BS c transmit their signals on the entire BS's available bandwidth B_c, and distinguish themselves according to their transmission power level, while SIC technique is applied at the receiver side to have the signals properly decoded.

A. UPLINK TRANSMISSION DATA RATE

Let us define as p_u^c the uplink transmission power of user u communicating with BS c. The corresponding wireless communication's channel gain is indicated by G_u^c and is differentiated among the different BSs, e.g., Macro BSs (MBSs), Pico BSs (PBSs), Networked Flying Platforms (NFPs). More information regarding the channel gain formulation specifics is provided in Table 1. Also, it is noted that within the scope of our paper and analysis, all the BSs considered to be equipped with a single antenna. Based on the principle of SIC, without loss of generality, the channel gains observed by a BS c are sorted in ascending order as $G_{|\mathbb{U}_c|}^c \leq \cdots \leq G_u^c \leq \cdots \leq G_1^c$ and the signal of the highest channel gain user is decoded first. When a signal is decoded, it is subtracted from the superposition signal before further decoding takes place. Hence, the interference sensed by a user u associated with BS c can be expressed as:

$$I_u^c = \sum_{u' \geq u+1}^{|\mathbb{U}_c|} G_{u'}^c p_{u'}^c + I_0, \tag{1}$$

where I_0 is the power of zero-mean Additive White Gaussian Noise (AWGN). Adopting Shannon's capacity formula, the achievable data rate of user u pertinent to BS c is

$$R_u^c = B_c \log_2\left(1 + \gamma_u^c\right), \tag{2}$$

where γ_u^c corresponds to the Signal-to-Interference-plus-Noise Ratio (SINR) given by

$$\gamma_u^c = \frac{G_u^c p_u^c}{I_u^c}. \tag{3}$$

B. CONTRACT BUNDLE

Given a user-to-BS association, each BS incentivizes its serving users to transmit with an appropriately selected power that enables the decoding of their signal, by offering them a contract bundle of {effort, reward}. The contract bundle $\{p_u^c, r_u^c\}$ designed by a BS c for a user $u \in \mathbb{U}_c$ consists of the user's effort, i.e., transmission power p_u^c, and its offered reward by the BS c, denoted by r_u^c. The offered reward is considered inversely proportional to the user's sensed interference I_u^c, defined as $r_u^c = \rho / I_u^c$, where $\rho \in \mathbb{R}^+$ is a reward factor. It is noted that the user's sensed interference I_u^c is determined by applying the SIC technique at the receiver, i.e., BS, as described in Section II-A, and it is computed and determined following Eq. (1). The physical meaning and interpretation of the reward is that the BS provides a greater reward to users that experience less interference, i.e., the worse channel gain users, since it expects greater transmission power to decode their signals. This design encourages fairness among NOMA users, who usually experience unequal data rates due to SIC prerequisites. Thus, the BSs provide the rewards to the users in order the latter ones to be incentivized to adapt their transmission power levels towards enabling their transmitted signals to be decoded by the receiver without burdening the overall communication environment with excessive interference. This process is further detailed later in Section II-D.

C. USER TYPE

Each user associated with a BS is characterized by a type that captures its private information related to its established channel quality. We define the type of a user $u \in \mathbb{U}_c$ communicating with BS c as $t_u^c = (G_u^c \cdot \sum_{u=1}^{|\mathbb{U}_c|}\{1/G_u^c\})^{-1/2}$, $t_u^c \in (0, 1]$, where G_u^c corresponds to its channel gain. The proposed novel formulation of the user's type captures the user's relative quality of its channel gain within the examined wireless communication environment with respect to the rest of the users. It is noted that the introduced user's type is unique for each user within the system and acts as its personal identity and characteristic. In the following analysis, we consider a fully heterogeneous networking environment in terms of users' channel gain characteristics and their corresponding types. Specifically, each user is characterized by a unique type and thus, there are $|\mathbb{U}_c|$ types of users associated with a BS c. By sorting the user types in ascending order as $t_1^c < \cdots < t_u^c < \cdots < t_{|\mathbb{U}_c|}^c$, it is revealed that a user of worse channel conditions is of higher type. Naturally, a higher type implies that the user invests greater effort (i.e., transmission power) and thus, is rewarded more by the BS.

D. USER'S AND BS'S UTILITY

We define the user's utility function U_u^c that depicts its perceived satisfaction from the reward r_u^c provided by the BS c, as well as its cost to provide its effort to the BS, i.e., transmission power p_u^c, as follows:

$$U_u^c\left(p_u^c\right) = t_u^c \cdot e\left(r_u^c\right) - p_u^c. \tag{4}$$

The first term expresses the user's satisfaction depending on its type t_u^c and its evaluation function of reward $e(r_u^c)$. The evaluation function is strictly increasing and concave with respect to the user's received reward (i.e., $e(0) = 0, e'(r_u^c) > 0, e''(r_u^c) < 0$). For demonstration purposes and without loss of generality, in the following we consider $e(r_u^c) = \sqrt{r_u^c}$. The physical meaning of the user's utility function captures the user's "profit" by transmitting its data to the receiver within a fully heterogeneous NOMA-based communications environment.

Similarly, a BS c experiences a utility $U_c^u = p_u^c - \mathcal{C} \cdot r_u^c$ by each user's u provided effort, while accounting for its personal cost to provide the reward to the user, with $\mathcal{C} > 0$ representing its unit cost. For simplicity in the presentation and without loss of generality, we consider that all the BSs experience the same unit cost $\mathcal{C}$. However, the rest of the analysis would hold true even if each BS was characterized by a personalized unit cost $\mathcal{C}_c$, as the latter is a constant value. Also, it is noted that the BSs should sophisticatedly evaluate their unit cost $\mathcal{C}$ in order to guarantee that the users' available battery is sufficient to participate in the personalized contract bundles. In the typical and realistic scenario that the BS is unaware of the user's type, the BS c estimates the probability λ_u^c that the user u is of type t_u^c, with $\sum_{u=1}^{|\mathbb{U}_c|}\lambda_u^c = 1$. Therefore, for a number of $|\mathbb{U}_c|$ users communicating with BS c, its utility function is represented as follows:

$$U_c = \sum_{u=1}^{|\mathbb{U}_c|}\left[\lambda_u^c \cdot \left(p_u^c - \mathcal{C} \cdot r_u^c\right)\right]. \tag{5}$$

The physical meaning of the BS's utility function captures the probabilistic "profit" that the BS will experience by incentivizing the users to adapt their transmission power levels in order to avoid creating excessive interference in the overall system, while enabling the BS to decode their received signals.

III. CONTRACT-THEORETIC POWER ALLOCATION

In the following, the proposed contract-theoretic power allocation mechanism, under both complete and incomplete CSI scenarios is presented, based on the already defined contract bundles, user types, and user and BS utilities. The power allocation is formulated as a fully distributed optimization problem executed by each BS of the system independently. The objective of this distributed optimization problem is to maximize the respective BS's utility function subject to its corresponding users' constraints, which in turn ensure their acceptance of the contract. As stated before, the contract designed by a BS c for a user $u \in \mathbb{U}_c$ consists of the user's transmission power (i.e., user's effort), and its corresponding offered reward by the BS c. Accordingly, a set of contract bundles between the BS and its serving users is the outcome that we seek from this procedure. A thorough analysis on how to derive optimal contract bundles, meeting the optimization objectives defined under complete and incomplete information scenarios, is pursued in the subsequent sections. It is noted here that in the following subsections we assume that the users have already been associated with a specific BS. However, later on, in Section IV, we provide a fully distributed user-to-BS association procedure and framework.

A. OPTIMAL CONTRACT UNDER COMPLETE INFORMATION

In this section, the complete information scenario is considered and analyzed, in the sense that the BS knows a priori the type of its serving users. This ideal case serves as a baseline to verify the effectiveness of the resource allocation performed under the incomplete CSI scenario. Indeed, since the BS is assumed to be completely aware of the users' types and corresponding channel characteristics, it can fully exploit their effort and maximize its utility, while guaranteeing that the users accept its offered contract. In other words, the BS has to ensure that the users experience a non-negative utility, i.e., the optimal contract bundles satisfy the individual rationality condition, as defined formally below.

Definition 1 (Individual Rationality (IR)): A contract bundle $\{p_u^c, r_u^c\}$ satisfies the individual rationality constraint if each user receives a non-negative utility, i.e.,

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq 0,\ \forall c \in C,\ \forall u \in \mathbb{U}_c. \tag{6}$$

Therefore, the problem of determining the optimal contracts for a BS c under the complete information of the users' types, can be written as follows:

$$\max_{\{p_u^c, r_u^c\}_{\forall u \in \mathbb{U}_c}} \quad U_c^u = p_u^c - \mathcal{C} \cdot r_u^c,\ \forall c \in C \tag{7a}$$

$$\textbf{s.t.} \quad t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq 0,\ \forall u \in \mathbb{U}_c. \tag{7b}$$

In the contract design under complete CSI, the BS will target to maximize its own utility, by providing the minimum acceptable utility to its serving users. In this case, it will decrease r_u^c until $t_u^c \cdot e(r_u^c) - p_u^c = 0$, meaning that the constraint of Eq. (7b) can be considered as equality. Accordingly, solving the equality of Eq. (7b) with respect to r_u^c and substituting in Eq. (7a) we get $U_c^u = p_u^c - \mathcal{C} \cdot (p_u^c/t_u^c)^2$. Thus, a closed-form solution can be derived for the optimization problem of Eq. (7a)-(7b) by solving the following equation:

$$\frac{\partial U_c^u}{\partial p_u^c} = 0,\ \forall c \in C,\ \forall u \in \mathbb{U}_c. \tag{8}$$

Consequently, under the assumption of complete CSI availability at a BS, it can be easily found that the optimal contract bundles between each BS c and its serving users are given by $\{p_u^c, r_u^c\} = \{\frac{(t_u^c)^2}{2\mathcal{C}}, (\frac{t_u^c}{2\mathcal{C}})^2\}$.

B. FEASIBLE CONTRACT UNDER INCOMPLETE INFORMATION

In this section, we elaborate on the necessary and sufficient conditions to determine a feasible contract under the realistic scenario of incomplete CSI. The purpose of this analysis is to derive optimal contracts subsequently in Section III-C.

In a scenario of incomplete CSI, the BS has to ensure that the users are provided not only with a non-negative utility, but also with the maximum utility, when selecting the contract designed for their own type. The former refers to the Individual Rationality (IR) condition given by Eq. (6), while the latter corresponds to the Incentive Compatibility (IC) condition defined formally below.

Definition 2 (Incentive Compatibility (IC)): Each user must select the contract bundle $\{p_u^c, r_u^c\}$ that is designed specifically for their own type t_u^c, i.e.,

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq t_u^c \cdot e\big(r_{u'}^c\big) - p_{u'}^c,\ \forall c \in C,\ \forall u, u' \in \mathbb{U}_c,\ u \neq u'. \tag{9}$$

Hence, it can be easily inferred that when a contract satisfies the IR and IC constraints, the users have adequate incentives to truthfully reveal their private type, by choosing the contract bundle designed for them. Apart from ensuring the incentive compatibility, several additional conditions should hold true to render a contract feasible.

Proposition 1: For any feasible contract $\{p_u^c, r_u^c\}$, the following must hold true: $r_u^c > r_{u'}^c \iff t_u^c > t_{u'}^c$ and $r_u^c = r_{u'}^c \iff t_u^c = t_{u'}^c$, $\forall u, u' \in \mathbb{U}_c,\ u \neq u'$.

Proof: Initially, we prove the sufficiency of the proposition, i.e., $t_u^c > t_{u'}^c \implies r_u^c > r_{u'}^c$, by using the IC constraint in Eq. (9). Based on the IC constraint in Eq. (9), we have

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq t_u^c \cdot e\big(r_{u'}^c\big) - p_{u'}^c, \tag{10}$$

$$t_{u'}^c \cdot e\big(r_{u'}^c\big) - p_{u'}^c \geq t_{u'}^c \cdot e\big(r_u^c\big) - p_u^c. \tag{11}$$

Adding appropriately the corresponding terms of the inequalities of Eq. (10) and Eq. (11), we obtain

$$t_u^c \cdot e\big(r_u^c\big) + t_{u'}^c \cdot e\big(r_{u'}^c\big) \geq t_u^c \cdot e\big(r_{u'}^c\big) + t_{u'}^c \cdot e\big(r_u^c\big). \tag{12}$$

By performing simple factorization, inequality Eq. (12) is recasted into the following one:

$$\big(t_u^c - t_{u'}^c\big) \cdot \big[e\big(r_u^c\big) - e\big(r_{u'}^c\big)\big] \geq 0. \tag{13}$$

Given that $t_u^c > t_{u'}^c$ and $e(r_u^c)$ is a strictly increasing function with respect to r_u^c, it is concluded from Eq. (13) that $r_u^c > r_{u'}^c$.

Thereafter, we prove the necessity of the proposition, i.e., $r_u^c > r_{u'}^c \implies t_u^c > t_{u'}^c$. Since $r_u^c > r_{u'}^c$ and $e(r_u^c)$ is strictly increasing with r_u^c, it holds that $e(r_u^c) - e(r_{u'}^c) > 0$. Hence, from Eq. (13) we obtain that $t_u^c > t_{u'}^c$. As a result, it has been proven that $r_u^c > r_{u'}^c \iff t_u^c > t_{u'}^c$.

Following a similar procedure and argumentation, it can be easily proven that $r_u^c = r_{u'}^c \iff t_u^c = t_{u'}^c$, which completes the proof. ■

The rationale behind Proposition 1 is that a higher type t_u^c user, which represents a user of worse channel conditions, will receive a greater reward from the associated BS c in order to be incentivized to establish a connection and transmit its data.

Proposition 2 (Monotonicity): A user of higher type, i.e., $t_1^c < \cdots < t_u^c < \cdots < t_{|\mathbb{U}_c|}^c$, will receive a greater reward from the BS c, i.e., $r_1^c < \cdots < r_u^c < \cdots < r_{|\mathbb{U}_c|}^c$, as it will contribute a greater effort, i.e., $p_1^c < \cdots < p_u^c < \cdots < p_{|\mathbb{U}_c|}^c$.

Proof: Given our assumption that user types follow an ascending order $t_1^c < \cdots < t_u^c < \cdots < t_{|\mathbb{U}_c|}^c$, the first part of this proposition readily stems from Proposition 1. Subsequently, we prove that for any feasible contract $\{p_u^c, r_u^c\}$, the following holds true: $p_u^c > p_{u'}^c \iff r_u^c > r_{u'}^c$, $\forall u, u' \in \mathbb{U}_c,\ u \neq u'$.

First, we prove that if $p_u^c > p_{u'}^c$ then $r_u^c > r_{u'}^c$. According to the IC constraint in Eq. (9), we have $t_u^c \cdot e(r_u^c) - p_u^c \geq t_u^c \cdot e(r_{u'}^c) - p_{u'}^c \Leftrightarrow t_u^c \cdot (e(r_u^c) - e(r_{u'}^c)) \geq p_u^c - p_{u'}^c$. Since $p_u^c > p_{u'}^c$ and given that $e(r_u^c)$ is a strictly increasing function of r_u^c, it holds that $r_u^c > r_{u'}^c$. Similarly, in order to prove that if $r_u^c > r_{u'}^c$ then $p_u^c > p_{u'}^c$, we follow the IC constraint in

Eq. (9), and we have $t^c_{u'} \cdot e(r^c_{u'}) - p^c_{u'} \geq t^c_{u'} \cdot e(r^c_u) - p^c_u \Leftrightarrow p^c_u - p^c_{u'} \geq t^c_{u'} \cdot (e(r^c_u) - e(r^c_{u'}))$. Since $r^c_u > r^c_{u'}$ and $e(r^c_u)$ is a strictly increasing function with respect to r^c_u, we conclude that $p^c_u > p^c_{u'}$. This completes the proof of the monotonicity condition. ■

Proposition 3: A user of higher type, i.e., $t^c_1 < \cdots < t^c_u < \cdots < t^c_{|\mathbb{U}_c|}$, will receive higher utility to be incentivized by the BS c, i.e., $U^c_1 < \cdots < U^c_u < \cdots < U^c_{|\mathbb{U}_c|}$.

Proof: We examine two users $u, u' \in \mathbb{U}_c$, $u \neq u'$ of types $t^c_u > t^c_{u'}$. Based on the IC condition in (9), we have

$$t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq t^c_u \cdot e\big(r^c_{u'}\big) - p^c_{u'} > t^c_{u'} \cdot e\big(r^c_{u'}\big) - p^c_{u'}. \quad (14)$$

Then, it holds that $U^c_u > U^c_{u'}$ when $t^c_u > t^c_{u'}$ and thus, for $t^c_1 < \cdots < t^c_u < \cdots < t^c_{|\mathbb{U}_c|}$ we conclude that $U^c_1 < \cdots < U^c_u < \cdots < U^c_{|\mathbb{U}_c|}$. ■

Based on the above analyzed conditions that guarantee the feasibility of a contract under an incomplete information scenario, the optimization problem executed by each BS c can, then, be formulated. The objective of each BS is to maximize its utility in order to be able to collect and properly decode the users' transmitted signals (as dictated by the received SINR). At the same time, all associated users' personal constraints should be satisfied in order for the users to be willing to be served by the specific BS. Therefore, the following distributed optimization problem, which captures both the BS's and corresponding users' requirements, is formulated at each BS as follows:

$$\mathbf{P1:} \quad \max_{\{p^c_u, r^c_u\}_{\forall u \in \mathbb{U}_c}} \quad U_c = \sum_{u=1}^{|\mathbb{U}_c|} \big[\lambda^c_u \cdot \big(p^c_u - \mathcal{C} \cdot r^c_u\big)\big], \ \forall c \in C \quad (15a)$$

$$\mathbf{s.t.} \quad t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq 0, \ \forall u \in \mathbb{U}_c \quad (15b)$$

$$t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq t^c_u \cdot e\big(r^c_{u'}\big) - p^c_{u'}, \quad \forall u, u' \in \mathbb{U}_c, u \neq u' \quad (15c)$$

$$0 \leq r^c_1 < \cdots < r^c_u < \cdots < r^c_{|\mathbb{U}_c|} \quad (15d)$$

where Eq. (15b), (15c), and (15d) represent the aforementioned IR, IC, and monotonicity constraints, respectively. Considering the fact that problem **P1** is non-convex, the procedure described in Section III-C below, is carried out to reduce its constraints and obtain a tractable solution in an effective manner.

C. OPTIMAL CONTRACT UNDER INCOMPLETE INFORMATION

In the following, the explanation and detailed analysis of the IR and IC constraints reduction is pursued, for a contract under incomplete information.

Step 1 (IR Constraints Reduction): By considering the assumption about the user types ordering, i.e., $t^c_1 < \cdots < t^c_u < \cdots < t^c_{|\mathbb{U}_c|}$ and the IC condition in Eq. (9), we can write $t^c_u \cdot e(r^c_u) - p^c_u \geq t^c_u \cdot e(r^c_{u'}) - p^c_{u'} \geq t^c_u \cdot e(r^c_1) - p^c_1$. Also, given that $t^c_u > t^c_1$ and based on the IR condition in Eq. (6), we have

$$t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq t^c_u \cdot e\big(r^c_1\big) - p^c_1 > t^c_1 \cdot e\big(r^c_1\big) - p^c_1 \geq 0. \quad (16)$$

As a result, we deduce that if the IR constraint of the lowest user type t^c_1 is ensured (i.e., $t^c_1 \cdot e(r^c_1) - p^c_1 \geq 0$), all other IR constraints for the users with higher types will automatically be satisfied. Towards increasing BS's utility, the latter IR constraint can be considered alternatively as equality, i.e., $t^c_1 \cdot e(r^c_1) - p^c_1 = 0$. Hence, the $|\mathbb{U}_c|$ IR inequality constraints defined by problem **P1**, are reduced to one IR equality constraint.

Step 2 (IC Constraints Reduction): To accommodate the IC constraints reduction process, additional terminology is used about the IC constraints defined between different user types. In particular, the IC constraints between user types u and u', $u' \in \{1, \ldots, u-1\}$ are termed as Downward IC (DIC) constraints. Specifically, the DIC constraint between adjacent user types u and $u-1$ is referred to as local DIC constraints. Similarly, the IC constraints between user types u and u', $u' \in \{u+1, \ldots, |\mathbb{U}_c|\}$ are called Upward IC (UIC) constraints, while the UIC constraint between adjacent user types u and $u+1$ pertains to the local UIC constraints. Next, we will consecutively analyze on how to reduce both the DIC and the UIC constraints to conclude to a convex optimization problem.

Proposition 4: All the DIC constraints can be represented by the local DIC constraints.

Proof: We consider three adjacent user types, such that $t^c_{u-1} < t^c_u < t^c_{u+1}$. Then, the following two local DIC constraints can be derived:

$$t^c_{u+1} \cdot e\big(r^c_{u+1}\big) - p^c_{u+1} \geq t^c_{u+1} \cdot e\big(r^c_u\big) - p^c_u, \quad (17)$$

$$t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq t^c_u \cdot e\big(r^c_{u-1}\big) - p^c_{u-1}. \quad (18)$$

Based on Proposition 1, we have $r^c_u > r^c_{u'} \iff t^c_u > t^c_{u'}$. Also, for $r^c_u > r^c_{u-1} \overset{e\nearrow}{\iff} e(r^c_u) > e(r^c_{u-1}) \Leftrightarrow e(r^c_u) - e(r^c_{u-1}) > 0$. For $t^c_{u+1} > t^c_u$, the second inequality becomes

$$t^c_{u+1} \cdot \big[e\big(r^c_u\big) - e\big(r^c_{u-1}\big)\big] > t^c_u \cdot \big[e\big(r^c_u\big) - e\big(r^c_{u-1}\big)\big] \geq^{Eq.(18)} p^c_u - p^c_{u-1}. \quad (19)$$

From inequality Eq. (17) and with the use of Eq. (19) we have

$$\begin{aligned} t^c_{u+1} \cdot e\big(r^c_{u+1}\big) - p^c_{u+1} &\geq t^c_{u+1} \cdot e\big(r^c_u\big) - p^c_u \\ &\geq^{Eq.(19)} t^c_{u+1} \cdot e\big(r^c_{u-1}\big) - p^c_{u-1} \\ &\geq \cdots \\ &\geq t^c_{u+1} \cdot e\big(r^c_1\big) - p^c_1. \end{aligned} \quad (20)$$

Therefore, if the DIC constraint between user types $u+1$ and u holds, then it, also, holds for user types $u+1$ and $u-1$. This property can be recursively extended downward from user types $u-1$ to 1, as dictated by Eq. (20).

Evidently, if the local DIC constraints hold true, then all the DIC constraints are automatically satisfied. In other words, all the DIC constraints can be equivalently captured by

$$t^c_u \cdot e\big(r^c_u\big) - p^c_u \geq t^c_u \cdot e\big(r^c_{u-1}\big) - p^c_{u-1}. \quad (21)$$

■

Proposition 5: All the UIC constraints can be represented by the local DIC constraints.

Proof: An identical procedure with Proposition 4 is followed, and the local UIC constraints between three adjacent user types, such as $t_{u-1}^c < t_u^c < t_{u+1}^c$, are written as:

$$t_{u-1}^c \cdot e\big(r_{u-1}^c\big) - p_{u-1}^c \geq t_{u-1}^c \cdot e\big(r_u^c\big) - p_u^c, \tag{22}$$

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq t_u^c \cdot e\big(r_{u+1}^c\big) - p_{u+1}^c. \tag{23}$$

Based on Proposition 1, we have $r_u^c > r_{u'}^c \iff t_u^c > t_{u'}^c$. For $t_u^c > t_{u-1}^c$, the second inequality becomes

$$\begin{aligned} p_{u+1}^c - p_u^c &\geq^{Eq.(23)} t_u^c \cdot \big[e\big(r_{u+1}^c\big) - e\big(r_u^c\big)\big] \\ &> t_{u-1}^c \cdot \big[e\big(r_{u+1}^c\big) - e\big(r_u^c\big)\big]. \end{aligned} \tag{24}$$

From inequality Eq. (22) and with the use of Eq. (24) we have

$$\begin{aligned} t_{u-1}^c \cdot e\big(r_{u-1}^c\big) - p_{u-1}^c &\geq t_{u-1}^c \cdot e\big(r_u^c\big) - p_u^c \\ &\geq^{Eq.(24)} t_{u-1}^c \cdot e\big(r_{u+1}^c\big) - p_{u+1}^c \\ &\geq \cdots \\ &\geq t_{u-1}^c \cdot e\Big(r_{|\mathbb{U}_c|}^c\Big) - p_{|\mathbb{U}_c|}^c. \end{aligned} \tag{25}$$

Thus, if the UIC constraint between user types $u-1$ and u holds, then it, also, holds for user types $u-1$ and $u+1$. Similarly, this property can be extended upward from user type $u+1$ to $|\mathbb{U}_c|$, as dictated by Eq. (25).

Therefore, we have proved that if the local UIC constraints hold true, then all UIC constraints are automatically satisfied and can be captured by

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c \geq t_u^c \cdot e\big(r_{u+1}^c\big) - p_{u+1}^c. \tag{26}$$

To complete the proof, it is remarkable to observe that the local DIC constraint defined in Eq. (21), can easily imply the following local UIC constraint:

$$t_{u-1}^c \cdot e\big(r_{u-1}^c\big) - p_{u-1}^c \geq t_{u-1}^c \cdot e\big(r_u^c\big) - p_u^c. \tag{27}$$

In this respect, inequality Eq. (26) can be equivalently represented by Eq. (21) and thus, all UIC constraints are reduced to the local DIC constraint implied by Eq. (21). The latter constraint, i.e., Eq. (21), is considered as equality in order to each BS to achieve the maximum benefit from its serving users' effort. Hence, the $|\mathbb{U}_c| \cdot (|\mathbb{U}_c| - 1)$ IC inequality constraints defined by problem **P1**, are reduced to $|\mathbb{U}_c| - 1$ equality constraints, accordingly. ■

Based on the reduced IR and IC constraints, the optimization problem **P1** can be rewritten as follows:

$$\mathbf{P2:} \quad \max_{(r_u^c, p_u^c)_{\forall u \in \mathbb{U}_c}} \quad U_c = \sum_{u=1}^{|\mathbb{U}_c|} \big[\lambda_u^c \cdot \big(p_u^c - \mathcal{C} \cdot r_u^c\big)\big], \ \forall c \in C \tag{28a}$$

$$\text{s.t.} \quad t_1^c \cdot e\big(r_1^c\big) - p_1^c = 0, \ \forall u \in \mathbb{U}_c \tag{28b}$$

$$t_u^c \cdot e\big(r_u^c\big) - p_u^c = t_u^c \cdot e\big(r_{u-1}^c\big) - p_{u-1}^c, \ \forall u \in \mathbb{U}_c \tag{28c}$$

$$0 \leq r_1^c < \cdots < r_u^c < \cdots < r_{|\mathbb{U}_c|}^c. \tag{28d}$$

Algorithm 1 Unified User Association and Power Allocation

Initialize a_L, a_H, $\mu_{c,0}$, b and set $ite = 0$, $S_c = 0$, $F_c = 0$, $\forall c \in C$.
Initialize the action probabilities vector $\mathbf{Pr}_u^{(0)}$, $\forall u \in U$ with equal probabilities for the BSs, in the coverage area of which each user belongs, otherwise set $\mathrm{Pr}_u^{c,(0)} = 0$.
repeat
  for $u = 1$ to $|U|$ **do**
    Choose a BS to associate with based on the action probabilities vector $\mathbf{Pr}_u^{(ite)}$.
  end for
  for $c = 1$ to $|C|$ **do**
    Sort the user types in ascending order.
    Obtain the optimal contract bundles by solving **P2**.
    Broadcast the optimal contract bundles to all associated users $u \in \mathbb{U}_c$.
    for $u = 1$ to $|\mathbb{U}_c|$ **do**
      Select the contract designed for its own type.
      Execute the contract and transmit the data with uplink transmission power p_u^{c*}.
      Evaluate the achieved data rate R_u^c by broadcasting $\mathbf{x}_u^c$ and $\mathbf{y}_u^c$ reports to BS c.
    end for
    Calculate the holistic answer $x_{BTS,c}$ regarding the serving users' satisfaction based on Eq. (33).
    Update the counters S_c and F_c based on $x_{BTS,c}$, and thereafter, update the reputation μ_c based on Eq. (34).
    Calculate users' feedback signals $\mathcal{F}_c^{(ite)}$ as in Eq. (35).
    Broadcast the users' normalized feedback signals vector $\hat{\boldsymbol{\mathcal{F}}}_{\boldsymbol{u}}^{\boldsymbol{c},(\boldsymbol{ite})}$.
  end for
  for $u = 1$ to $|U|$ **do**
    Update the action probabilities vector $\mathbf{Pr}_u^{(ite+1)}$ based on Eq. (36) and Eq. (37).
  end for
until for all users $u \in U$ there is at least one action probability such that $Pr_u^{c,(ite+1)} \geq \varepsilon$, $\varepsilon \to 1$.

It is noted that the resulting optimization problem P2 is an equivalent transformation of the original P1 problem, thus resulting to the same outcome, i.e., the optimal contracts established among the users and the BSs. We can easily prove that **P2** is a convex programming problem by checking the Hessian matrix. Thus, **P2** can be solved by applying the Karush–Kuhn–Tucker (KKT) conditions. Accordingly, the optimal users' uplink transmission power vector $\mathbf{p}_c^* = [p_1^{c*}, \ldots, p_{|\mathbb{U}_c|}^{c*}]$ and BS's rewards vector $\mathbf{r}_c^* = [r_1^{c*}, \ldots, r_{|\mathbb{U}_c|}^{c*}]$, can be determined.

The contract design and optimization is handled by each BS of the heterogeneous network independently, and the detailed process is summarized as part of the Algorithm 1.

IV. AUTONOMOUS USER-TO-BS ASSOCIATION

In this section, we introduce a fully distributed and user-centric user-to-BS association framework, where the users of the underlying network topology, acting as learning automata, autonomously select the BS to be associated with and transmit their data. To enhance the users' satisfaction over the provided communication service, their decision is

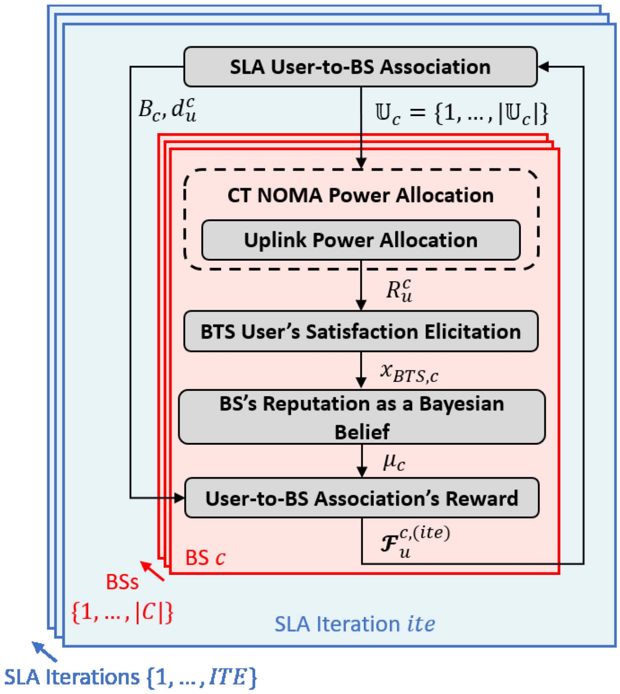


FIGURE 2. Overall resource orchestration flow and connection between the contract-theoretic and RL mechanisms.

probabilistically reinforced by considering the prospective BSs' network-related and social characteristics. Analytically, we introduce a mechanism based on the Bayesian Truth Serum concept in order to truthfully elicit the users' perceived satisfaction when served by a certain BS. The users are allowed to express their subjective opinion regarding their achieved data rate after the contract-theoretic uplink power allocation is performed. The overall users' satisfaction reports contribute to the extraction of an objective outcome pertinent to each BSs' service provisioning. This objective outcome is further utilized to formulate each BSs' reputation as a Bayesian Belief and provide it back to the users as a feedback for their BS selection. A high-level overview of the autonomous user-to-BS association is illustrated in Fig. 2. Additional details with respect to each individual mechanism, are provided in the subsequent sections. Finally, a comprehensive analysis on the algorithmic complexity of the unified user association and power allocation scheme complements the section.

A. USERS' SATISFACTION-RELATED INFORMATION TRUTHFULNESS BASED ON BAYESIAN TRUTH SERUM

Undoubtedly, the primary concern when designing resource allocation mechanisms for 5G networks, is to enhance users' perceived satisfaction. Aligning with the user-centric flavor of the overall resource orchestration procedure, we introduce an assessment phase, where the users report their personal assessment about the service provided by the BS that they are associated with, revealing in this way their corresponding perceived satisfaction. As a means of eliciting truthful assessments from the users in a vastly heterogeneous network, we adopt the method of Bayesian Truth Serum (BTS) [34]. The BTS method is strict Bayes-Nash incentive compatible for $|\mathbb{U}_c| \to \infty$ and allows us to extract a holistic objective evaluation from the users' subjective reviews, when the ground truth is totally unknown.

In particular, each user $u \in \mathbb{U}_c$ provides an answer regarding a binary question, i.e., "Are you satisfied from the experienced data rate when served by BS c?", by filing two reports:

- The information report $\mathbf{x}_u^c = [x_{u,c}^{NO}, x_{u,c}^{YES}]$, which is the user's personal answer to the aforementioned question. It holds true that $\sum_{i\in\{NO,YES\}} x_{u,c}^i = 1$, where $x_{u,c}^i \in \{0, 1\}, \forall i \in \{NO, YES\}$.
- The prediction report $\mathbf{y}_u^c = [y_{u,c}^{NO}, y_{u,c}^{YES}]$, where $y_{u,c}^{YES}$ corresponds to the prediction regarding the proportion of the users which answered YES, i.e., $x_{u,c}^{YES} = 1$, while $y_{u,c}^{NO}$ is the prediction regarding the users whose answer is NO, i.e., $x_{u,c}^{NO} = 1$. It holds true that $\sum_{i\in\{NO,YES\}} y_{u,c}^i = 1$, where $y_{u,c}^i \in [0, 1], \forall i \in \{NO, YES\}$.

Afterwards, the population endorsement frequencies $\bar{x}_c^i$ and the geometric mean of the users' predictions $\bar{y}_c^i$ for each of the available answers $i \in \{NO, YES\}$ can be calculated by

$$\bar{x}_c^i = \frac{1}{|\mathbb{U}_c|} \cdot \sum_{u=1}^{|\mathbb{U}_c|} x_{u,c}^i \tag{29}$$

and

$$\log\left(\bar{y}_c^i\right) = \frac{1}{|\mathbb{U}_c|} \cdot \sum_{u=1}^{|\mathbb{U}_c|} \log\left(y_{u,c}^i\right). \tag{30}$$

As a consequence, the BTS score $s_{BTS,u}^c(\mathbf{x}_u^c, \mathbf{y}_u^c)$ of each user u, which corresponds to the truthfulness of the user's committed answer, can be expressed by

$$s_{BTS,u}^c\left(\mathbf{x}_u^c, \mathbf{y}_u^c\right) = \sum_i x_{u,c}^i \cdot \log\left(\frac{\bar{x}_c^i}{\bar{y}_c^i}\right) + \alpha \cdot \sum_i \bar{x}_c^i \cdot \log\left(\frac{y_{u,c}^i}{\bar{x}_c^i}\right), \tag{31}$$

where the parameter $\alpha > 0$ controls the effect of the prediction score in the total BTS score. The first part of Eq. (31) is known as the information score, while the second part as the prediction score of the users. Regarding the information score, it increases with respect to how surprisingly common an answer is, i.e., how much greater the population endorsement frequency $\bar{x}_c^i$ is compared to the $\bar{y}_c^i$. It is noteworthy to mention that the surprisingly common criterion is based on the Bayesian Reasoning principle, according to which a user is of the opinion that the rest of the population will eventually underestimate its personal opinion, and hence, the user reports a higher prediction for its answer in order to further support it [35]. This is in line with the Bayesian argument, which concludes to the fact that a user's truthful opinion is more likely to be surprisingly common. The prediction score formulation is based on a penalty which is proportional to the Kullback-Leibler divergence between the actual population endorsement frequencies $\bar{x}_c^i$

of the answers and the corresponding user's predictions $y^i_{u,c}, \forall i \in \{NO, YES\}$ [36].

As a result, the optimal prediction score (i.e., the lowest penalty) is equal to 0, which is achieved when the user's prediction $y^i_{u,c}$ is equal to the real users' proportion $\overline{x}^i_c$ of the answer i, i.e., $log(\frac{y^i_{u,c}}{\overline{x}^i_c}) \overset{y^i_{u,c}=\overline{x}^i_c}{=} log(1) = 0$, also known as absolute accuracy. Thus, the average BTS score of each answer $i \in \{NO, YES\}$, can be written as

$$\overline{u}^i_c = \frac{1}{|\mathbb{U}_c| \cdot \overline{x}^i_c} \cdot \sum_{u=1}^{|\mathbb{U}_c|} x^i_{u,c} \cdot s^c_{BTS,u}(\mathbf{x}^c_u, \mathbf{y}^c_u). \tag{32}$$

The final holistic and most truthful answer $x_{BTS,c}$ regarding the satisfaction of the users belonging to the set $\mathbb{U}_c$ and served by the BS c, is the one with the highest average BTS score, as follows

$$x_{BTS,c} = \underset{i \in \{NO,YES\}}{\arg\max}\ \overline{u}^i_c. \tag{33}$$

B. BSS' REPUTATION AS A BAYESIAN BELIEF

After acquiring the users' personal assessment for a given BS association via the BTS method described in Section IV-A, we can eventually extrapolate the long-term social characteristics of the network, in terms of the users' perceived satisfaction, over several iterations of the user-to-BS association procedure. In our setting, to formulate the network's social characteristics we utilize the reputation μ_c for each BS c, which is a Bayesian model featuring adverse selection based on Bayesian updating of belief [37]. We assume that all users associated with a certain BS share the same prior belief distribution $\mu_{c,0} = \mu_0, \forall c \in C$, regarding the potential satisfaction that they might receive by getting associated with BS c. We consider that every BS c can offer either a satisfying or dissatisfying data rate with probabilities a_H and a_L, respectively, where it holds true that $0 < a_L < a_H < 1$. After the BTS evaluation takes place (Section IV-A), an up-vote or down-vote $\forall c \in C$ occurs, creating a history for every BS throughout the time horizon. We introduce S_c and F_c to indicate the number of times that BS c satisfied ($x_{BTS,c} = YES$) and dissatisfied ($x_{BTS,c} = NO$) the users being associated with it, correspondingly, up to the present time instance. Thus, each BS's c posterior belief distribution can be expressed as follows:

$$\mu_c = \frac{\mu_0 \cdot a_H^{(S_c)} \cdot (1-a_H)^{(F_c)}}{\mu_0 \cdot a_H^{(S_c)} \cdot (1-a_H)^{(F_c)} + (1-\mu_0) \cdot a_L^{(S_c)} \cdot (1-a_L)^{(F_c)}}. \tag{34}$$

To prevent situations, where large S_c and F_c exponents lead to unrepresentable numbers, the S_c and F_c counters are updated with a step equal to 0.5, each time a user is satisfied or dissatisfied, respectively.

By observing Eq. (34), we deduce that the reputation μ_c is correlated with the BS's c history of the users' BTS evaluations, i.e., S_c and F_c, since it increases when the former increases and decreases when the latter increases.

C. REINFORCEMENT LEARNING-ENABLED USER-TO-BS ASSOCIATION

Till this point, we have described in detail the operation of the inner part of the proposed resource orchestration framework, depicted in red in Fig. 2. In this section, we adduce the reinforcement learning approach based on the Stochastic Learning Automata (SLA) model [38], according to which each user selects the most beneficial BS to be associated with in a distributed and autonomous manner.

Inherently, the users within the considered wireless environment aim at minimizing their communication delay and thus, exhibit a preference towards associating with a BS in their close proximity. Furthermore, in our proposed solution, the users' association preference is affected by the BSs' bandwidth and reputation, leaning towards a resource conscious utilization of BS's available frequency resources, by preventing the drawn of excessive network traffic. Considering a user-to-BS association at a specific iteration ite of the SLA algorithm, each BS c within the network determines a vector of personalized feedback signals $\boldsymbol{\mathcal{F}}^{(ite)}_c = [\mathcal{F}^{c,(ite)}_1, \ldots, \mathcal{F}^{c,(ite)}_{|\mathbb{U}_c|}]$ for each serving user u, and broadcasts them to the respective user devices. Naturally, the users' personalized feedback signals reflect their benefit from communicating with the specific BS c and are given by

$$\mathcal{F}^{c,(ite)}_u = \frac{\mu_c \cdot B_c / \sum_{c=1}^{|C|} B_c}{d^c_u / \sum_{u=1}^{|\mathbb{U}_c|} d^c_u}, \tag{35}$$

where μ_c is the BS's long-term reputation over the preceding SLA iterations, while $B_c / \sum_{c=1}^{|C|} B_c$ and $d^c_u / \sum_{u=1}^{|\mathbb{U}_c|} d^c_u$ correspond to the normalized BS's available bandwidth and user's distance from the BS, respectively. Each user's personalized feedback signal of Eq. (35) is further normalized as $\hat{\mathcal{F}}^{c,(ite)}_u = \sqrt[4]{\mathcal{F}^{c,(ite)}_u / \sum_{u=1}^{|\mathbb{U}_c|} \mathcal{F}^{c,(ite)}_u}$, such that $\hat{\mathcal{F}}^{c,(ite)}_u \in [0, 1]$.

After each BS in the network broadcasts the users' normalized feedback signals, each user u acts as a stochastic learning automaton and updates its personal action probabilities vector $\mathbf{Pr}^{(ite)}_u = [Pr^{1,(ite)}_u, \ldots, Pr^{|C|,(ite)}_u]$ at the end of iteration ite. Specifically, for a user u associated with a BS $c \in C$ at iteration ite, the probability of selecting the same BS c in the subsequent iteration $ite+1$ is defined as

$$Pr^{c,(ite+1)}_u = Pr^{c,(ite)}_u + b \cdot \hat{\mathcal{F}}^{c,(ite)}_u \cdot \left(1 - Pr^{c,(ite)}_u\right), \tag{36}$$

while the probability of selecting a different BS $c' \in C$, $c' \neq c$ at the next iteration is determined by

$$Pr^{c',(ite+1)}_u = Pr^{c',(ite)}_u - b \cdot \hat{\mathcal{F}}^{c,(ite)}_u \cdot Pr^{c',(ite)}_u, \tag{37}$$

where $0 < b \leq 1$ is the SLA algorithm's learning rate.

Apparently, the users' update of their action probabilities concludes the overall resource orchestration procedure along one iteration of the proposed algorithm, initiating the subsequent algorithm's iteration with their updated BS selections, as demonstrated in Fig. 2. Throughout the time horizon, this iterative procedure enables the users to converge to the most beneficial selection of BS, as dictated by their normalized feedback signal. The optimization policy of the proposed reinforcement learning mechanism aims to optimize the long-term normalized personalized feedback received by each user, converging to a beneficial user-to-BS association, which is beneficial from the users' perspective. The convergence of the SLA algorithmic mechanism is achieved when for all users $u \in U$ there is at least one action probability such that $Pr_u^{c,(ite+1)} \geq \varepsilon$, $\varepsilon \rightarrow 1$ [39], [40]. The complete process and operation of the proposed unified user association and power allocation scheme is summarized in Algorithm 1.

D. COMPLEXITY ANALYSIS

First, to analyze the complexity of Algorithm 1, we investigate the contract-theoretic power allocation mechanism that is encapsulated in the overall user-to-BS association procedure. For a given iteration of the SLA algorithm, where all users in the network are associated with a prospective BS, the fully distributed optimization problem **P2**, presented at Section III-C, is executed by each BS. Hence, in our complexity analysis we presume that the contract-theoretic power allocation is performed in parallel by all BSs of the system.

The optimization problem **P2** can be solved via well known existing methods for solving constrained nonlinear optimization problems, and accordingly obtain the optimal contracts under the incomplete information scenario. For demonstration purposes, we utilize the Sequential Quadratic Programming (SQP) method [41], along with the *fmincon()* [42] function implemented by the MATLAB Optimization Toolbox to return the constrained nonlinear optimization problem's solution, the computational complexity of which is denoted as $\mathcal{O}(K)$ [43]. We also indicate as *ITE* the number of the iterations required by the SLA algorithm to converge. Consequently, the SLA algorithm's complexity is calculated as: $\mathcal{O}(ITE \cdot (K + |U| + |U| \cdot |C|))$, i.e., $\mathcal{O}(ITE \cdot (K + |U| \cdot |C|))$, taking into account that the complexities of the BSs' selections and the users' action probabilities' updates at every SLA iteration are $\mathcal{O}(|U|)$ and $\mathcal{O}(|U| \cdot |C|)$, respectively. Finally, due to the fact that both the complexities of the Bayesian Truth Serum and the reputations' updates are $\mathcal{O}(|U|)$, and since the rest of the SLA algorithm includes only algebraic calculations (of $\mathcal{O}(1)$ complexity), the overall complexity of Algorithm 1 is obatined as: $\mathcal{O}(ITE \cdot (K + |U| \cdot |C|))$.

V. PERFORMANCE EVALUATION

In this section, the performance and effectiveness of the proposed unified user association and power allocation scheme is demonstrated by performing a detailed numerical evaluation, via modeling and simulation. First, in Section V-A, we focus on validating the operation of the contract-theoretic mechanism in terms of the allocated optimal contract bundles and the obtained user and BS utilities. The specifics of the user-to-BS association procedure is further studied in Section V-B, where we explore and analyze the characteristics that pertain to the operation of the pure autonomous user-to-BS association mechanism. Having verified and analyzed the pure performance of both the contract-theoretic power allocation and the RL and BTS-based user association in detail, Section V-C concludes our evaluation with some comparative results obtained from their seamless joint operation. In particular, a comparative analysis over different heuristic user-to-BS association mechanisms and under different spatial user distributions is provided, that demonstrates and explains the superiority of the proposed RL-enabled orchestration scheme in increased density heterogeneous wireless deployments. For the latter, both uniform and non-uniform user distribution over the network topology scenarios are evaluated.

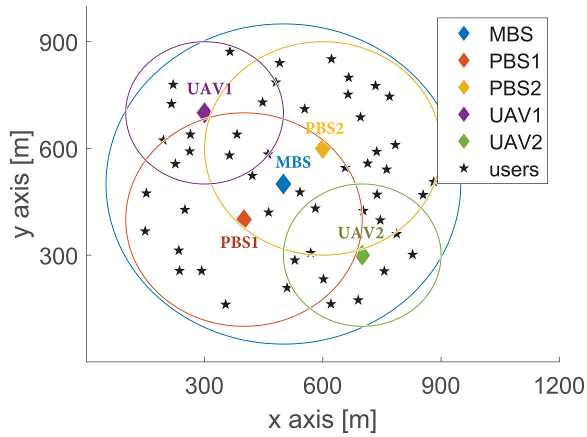


FIGURE 3. Wireless network simulation topology with uniform user distribution.

Throughout our evaluation, we consider a densely deployed portion of a macrocell of 450-meter radius, consisting of one Macro BS (MBS) located in the center of the cell, as well as two Pico BSs (PBSs) and two Unmanned Aerial Vehicles (UAVs), serving as NFPs aiming at temporarily alleviating the excessive network traffic. The coverage radius of the PBSs and the UAVs is set to 300 m and 200 m, respectively. In that manner, the coverage areas of all BSs overlap with each other, forming the hybrid aerial-terrestrial communications environment that is graphically represented in Fig. 3. In the following, unless otherwise explicitly stated, we consider that the simulated network topology serves 50 users uniformly distributed with maximum uplink transmission power equal to 23 dBm. The specific simulation parameters are given in Table 1. Our setting is perfectly aligned with the heterogeneous system baseline simulation model defined by the Third-Generation Partnership Project (3GPP) in [44] and is further extended to account for UAV-assisted communications. We model the channel conditions of the ground-to-air links between the users and the UAVs according to the reference free-space path loss model, as proposed in [45]. Regarding the contract theory-related parameters, we consider that each BS c estimates its serving users' channel conditions, i.e., their types, following a uniform distribution, such that $\lambda_u^c = 1/|\mathbb{U}_c|$, $\forall c \in C$, $\forall u \in \mathbb{U}_c$. Each BS's c unit cost is set equal to $\mathcal{C} = 0.65$, while each user's $u \in U$ reward factor is $\rho = 10^{-18}$. In the following results, for demonstration purposes, unless otherwise explicitly stated,

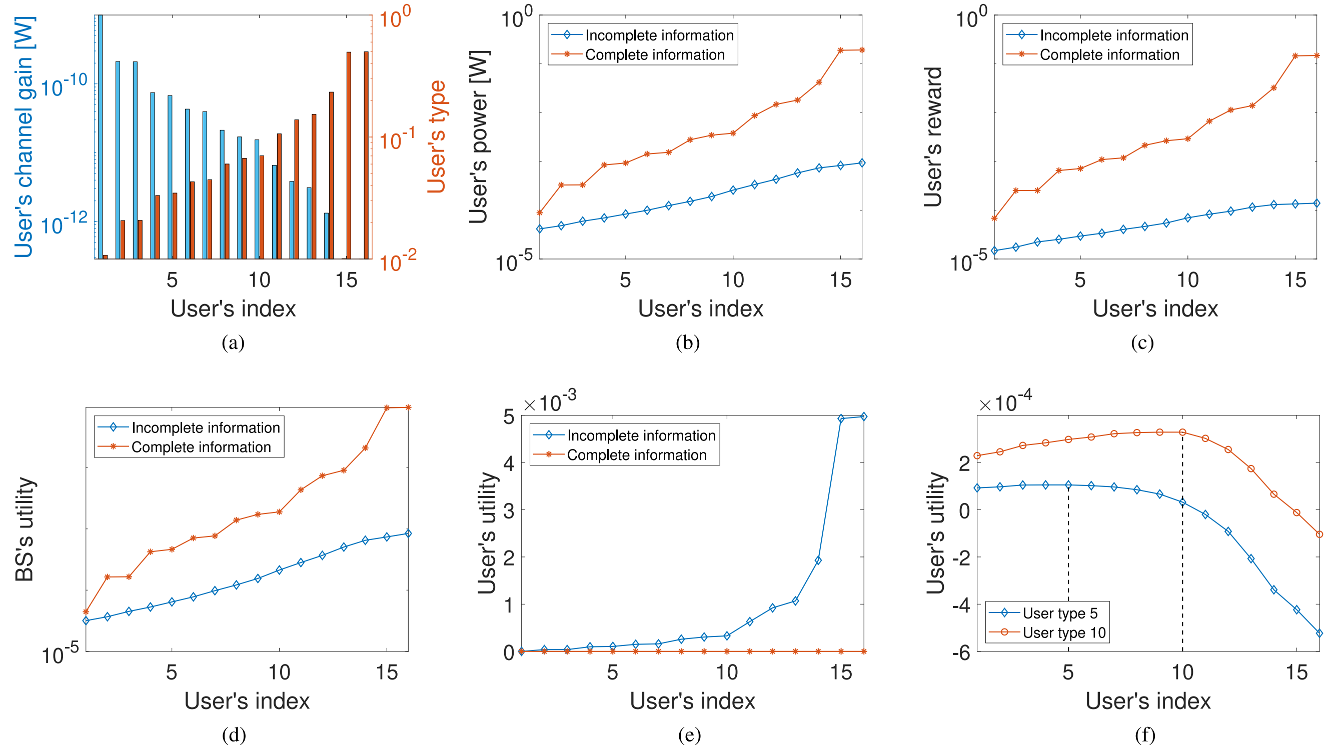


FIGURE 4. Pure contract-theoretic power allocation evaluation for MBS under complete and incomplete CSI scenarios.

the learning rate parameter of the SLA algorithm is set equal to $b = 0.7$. It is noted that the users and the UAVs remain stationary throughout a decision period, while the inclusion of the mobility aspect within the examined problem is part of our current and future work.

A. CONTRACT-THEORETIC MECHANISM EVALUATION

In this section, the proper functioning and operation of the proposed contract-theoretic power allocation mechanism is examined under both complete and incomplete CSI scenarios. A complete execution of the unified user association and power allocation scheme is performed and the contract-theoretic results for the optimal user-to-BS association, where the algorithm converged, are obtained. In Fig. 4(a)-4(f), we indicatively analyze and present the results of the contract-theoretic power allocation achieved for the MBS and its 16 associated users in total. Similar results and analysis are obtained for any BS of the considered wireless network topology.

In particular, Fig. 4(a) presents the users' channel gain and their corresponding type as a function of the user's index. The results confirm that as the user index increases, the communication's channel conditions worsen, i.e., $g_1^{MBS} > \cdots > g_u^{MBS} > \cdots > g_{16}^{MBS}$, while the user's type increases, i.e., $t_1^{MBS} < \cdots < t_u^{MBS} < \cdots < t_{16}^{MBS}$. This behavior is imposed by the inversely proportional relationship between these two considered variables, i.e., $t_u^c = (G_u^c \cdot \sum_{u=1}^{|\mathbb{U}_c|}\{1/G_u^c\})^{-1/2}$ that is used to achieve a fair power allocation obeying to NOMA principles, as it is demonstrated later in this section.

The obtained optimal contract bundles are depicted in Fig. 4(b)-4(c) as a function of the user index, for both the cases of complete and incomplete CSI, illustrating the user's invested transmission power (Fig. 4(b)) and the corresponding MBS's provided reward (Fig. 4(c)). Moreover, the MBS's attained utility U_{MBS}^u by each user's $u \in \mathbb{U}_{MBS}$ provided effort (i.e., uplink transmission power) and each user's utility U_u^{MBS}, when associated with the MBS, are presented in Fig. 4(d) and Fig. 4(e), respectively, as a function of the user index under the complete and incomplete CSI scenarios. The results show that a user of higher type, who experiences worse channel conditions (Fig. 4(a)), transmits its data with higher power (Fig. 4(b)), ensuring the user fairness encouraged by power-domain NOMA. Thus, the MBS rewards the users of higher type, with a higher reward (Fig. 4(c)), and consequently, those users achieve higher utility compared to users of lower type (Fig. 4(e)). Apart from the increased users' achieved utility, MBS's utility is, also, increasing with respect to the users' type (Fig. 4(d)).

Furthermore, comparing the results in Fig. 4(b)-4(c) between the scenarios of complete and incomplete CSI, it is inferred that the obtained contract bundles follow a similar behavior with respect to the users' types in both scenarios. Thus, the results verify the accuracy of the performed resource allocation in situations where there is absence of complete CSI from the MBS's behalf. Nevertheless, the motivation behind the MBS's contract bundle allocation strategy along the different CSI scenarios, primarily differentiates the achieved users' and MBS's attained utilities, as revealed by

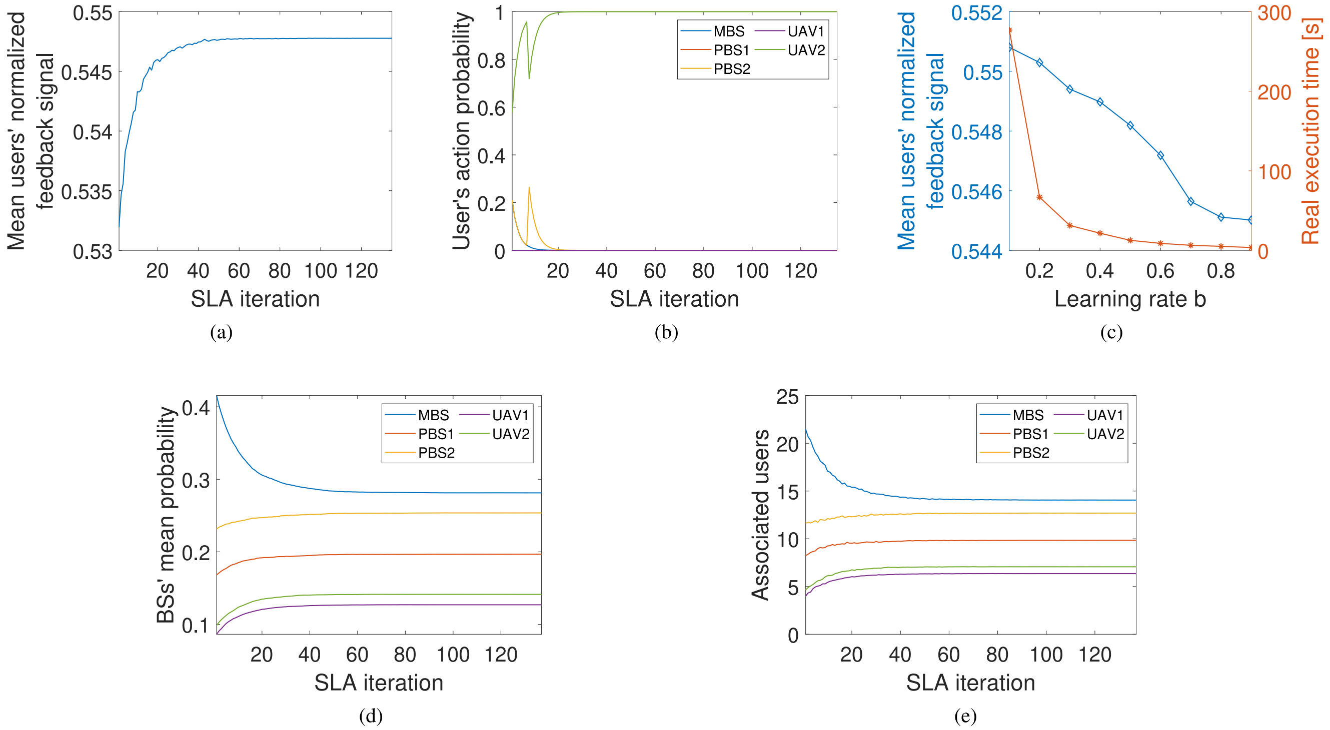


FIGURE 5. Pure reinforcement learning-enabled association mechanism's evaluation.

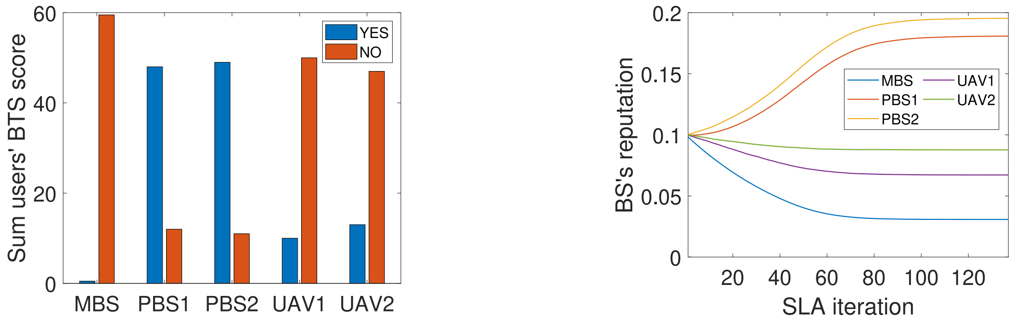
the results in Fig. 4(d)-4(e). Specifically, knowing a priori the users' types, the MBS fully exploits the users' effort to maximize its personal utility, providing them back with the minimum possible reward that marginally ensures the users' acceptance of the contract, i.e., the satisfaction of their rationality constraints. Consequently, based on Fig. 4(d), the MBS's utility is higher under the complete CSI scenario, while at the same time, the utility achieved by all associated users is equal to zero (Fig. 4(e)).

To complement the evaluation of the contract feasibility under incomplete CSI, in Fig. 4(f), the utilities of two selected users with indexes 5 and 10 are examined over each contract bundle designed by the MBS (represented in the horizontal axis by the user index) to showcase that the contract bundles are incentive compatible. The results confirm the IC condition satisfaction, in the sense that each user being part of the proposed contract-theoretic agreement is provided with the adequate incentives to select the contract bundle designed for its own type and experienced channel characteristics. By selecting the most appropriate contract for their type, the users on the one hand steer the power allocation procedure, while on the other hand they incidentally reveal their type and thus, their CSI, contributing to the establishment of a common CSI knowledge with the MBS.

B. USER-TO-BS ASSOCIATION MECHANISM EVALUATION

In this section, we aim to elucidate the operational characteristics of the pure autonomous user-to-BS association mechanism proposed in this article. First, we target the evaluation of the RL-enabled association mechanism's external operation, i.e., its convergence behavior and achieved association outcome (Fig. 5(a)-5(e)). Afterwards, we emphasize on the internal features that guide the RL algorithm's and thus, the users' behavior throughout the association process. In this regard, the impact of the users' satisfaction-related information elicitation via the BTS method to the overall association procedure is analyzed (Fig. 6(a)-6(b)). The results introduced for the rest of Section V, unless otherwise explicitly stated, have been averaged over a number of 500 executions.

Initially, we study the convergence behavior of the proposed RL association mechanism based on the SLA model. In Fig. 5(a), the cumulative mean users' normalized feedback signal (Eq. (35)) is illustrated with respect to the SLA algorithm's iterations. The results reveal that after approximately 50 SLA iterations, the mean users' normalized feedback signal converges to its maximum value and henceforth, the users persist in their BS association selection. Indeed, the convergence of the users' feedback signal directly implies the convergence of the users' personal action probabilities vector, determined by Eq. (36)-(37), which dictates the probability of selecting a specific BS. This observation is further verified by Fig. 5(b), which demonstrates the action probabilities for one indicative user as a function of the SLA algorithm's iterations. Taking into consideration that the user under investigation is located within the coverage areas of the MBS, the PBS2 and the UAV2, an equal number



(a) (b)

FIGURE 6. The impact of BSs' history of users' BTS evaluations to the evolution of BSs' reputation over the SLA iterations.

of approximately 50 SLA iterations is required in order for the user's action probabilities to converge. Eventually, the specific user's most beneficial BS association is the one, for which the user's action probability converges at 1, i.e., UAV2 in this case.

Concerning the overall RL-enabled association algorithm's speed of convergence, which actually refers to the speed of convergence of Algorithm 1, we conduct a Monte Carlo simulation over all possible values of the SLA model's learning rate parameter $b \in [0.1, 0.9]$ and derive the resulting real execution time in [s], as well as the achieved mean users' normalized feedback signal (Fig. 5(c)). As the value of the learning rate parameter b decreases, the exploration of the possible user-to-BS association alternatives is becoming exhaustive. Consequently, the algorithm identifies user-to-BS associations that lead to improved users' benefit, denoted by the increasing trend of the mean users' normalized feedback signal. Thus, the exploration of higher benefit user-to-BS associations is performed with the cost of increased real execution time. The learning rate parameter b can appropriately be controlled in realistic applications accounting for the trade-off among the users' benefit and the time critical decision making.

To gain more insight about the convergence behavior of the RL mechanism across the overall wireless network simulation topology, as presented in Fig. 3, we calculate the mean BSs' probabilities of being selected as a function of the SLA iterations (Fig. 5(d)). The corresponding results are next, correlated with the emerged total number of users associated with each BS versus the SLA iterations (Fig. 5(e)). Owing to the fact that the MBS provides complete coverage to the considered wireless network topology, the mean MBS's probability of being selected is higher compared to the rest of the BSs during the first SLA iterations. As a result, the total number of users associated with the MBS is, also, higher during the first SLA iterations. As the algorithm's execution evolves, part of the users that where initially assigned with the MBS are uniformly assigned with other BSs after assessing the tradeoff between user-to-BS distance and BSs' bandwidth availability.

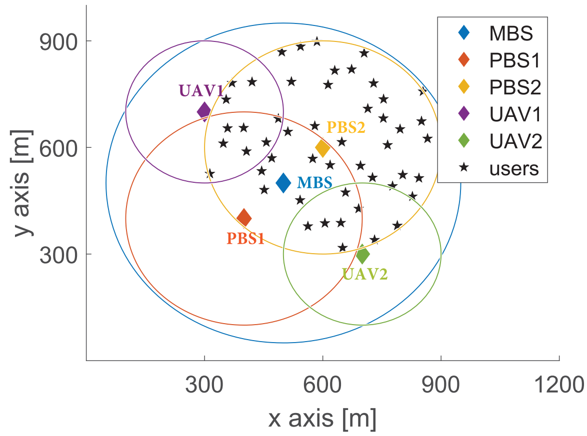
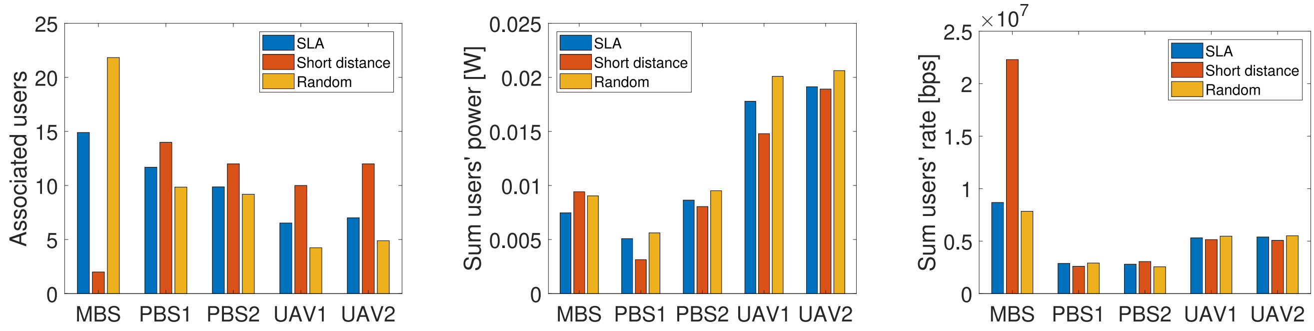


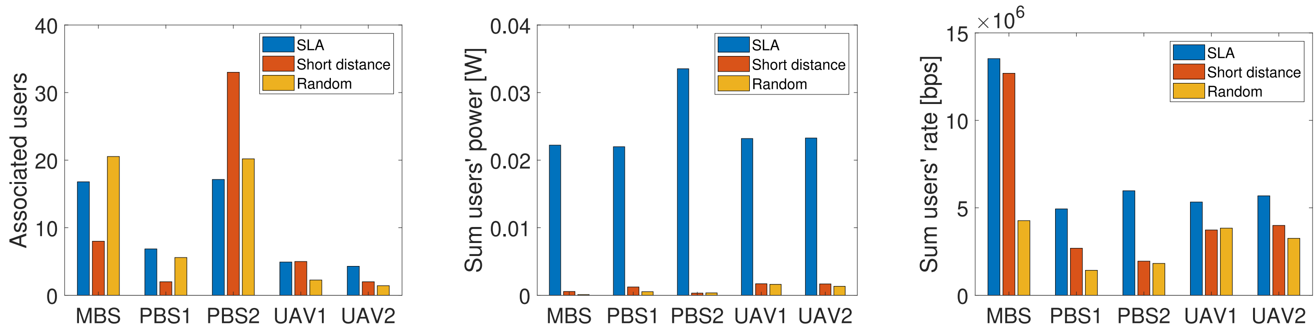
FIGURE 7. Wireless network simulation topology with non-uniform user distribution.

Targeting at the interpretation of the RL-enabled association mechanism's internal features that steer such a user-to-BS association, we provide an analysis regarding the evolution of the BSs' reputation based on their history of users' BTS evaluations. Specifically, Fig. 6(a) presents the sum users' BTS score for each BS (i.e., sum of "YES" and "NO" answers) after the algorithm's convergence, which is actually captured by the counters S_c and F_c. Accordingly, Fig. 6(b) depicts each BSs' reputation with respect to the SLA iterations. At this point, it should be recalled that for a total number of 120 SLA iterations, the sum BTS scores of "YES" and "NO" answers add up to 60, due to the S_c and F_c counters' increment with a step equal to 0.5. Following our analysis and focusing on the MBS, the dominance of its associated far-distanced users, whose achieved data rate is not satisfying enough, leads to an equal to zero "YES" BTS score (Fig. 6(a)), justifying the degradation of the MBS's reputation over the SLA iterations (Fig. 6(b)). In the contrary, the short-distanced users communicating with the PBSs, whose available bandwidth is adequate enough to provide them with a satisfying service, vote in favor of the PBSs, causing an increase in the reputation of the latter over the SLA iterations. However, this observation does not apply for the users served by the UAVs, mainly due to the restricted UAVs' available bandwidth.



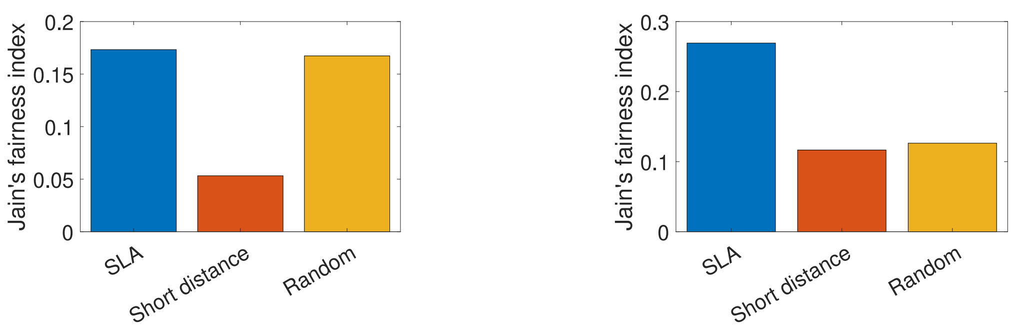
(a) (b) (c)

FIGURE 8. Comparative evaluation between SLA, short distance and random association per BS, in terms of (a) total number of associated users, (b) sum of users' powers and (c) sum of users' rates, under uniform users' distribution.



(a) (b) (c)

FIGURE 9. Comparative evaluation between SLA, short distance and random association per BS, in terms of (a) total number of associated users, (b) sum of users' powers and (c) sum of users' rates, under non-uniform users' distribution.



(a) (b)

FIGURE 10. Comparative evaluation between SLA, short distance and random association in terms of the overall achieved Jain's fairness index under (a) uniform users' distribution and (b) non-uniform users' distribution.

C. COMPARATIVE EVALUATION

This section is devoted to demonstrating the effectiveness of the devised unified scheme towards the achieved users' fairness in the share of the available resources. Specifically, our proposed RL-based user-to-BS association mechanism is compared against other heuristic mechanisms, i.e., short-distance and random associations. Following the short-distance associations, the users are associated with the BS in their closest proximity, whereas based on the random association the users are randomly associated with a BS, given that they belong to its coverage area. After the association phase, for fairness in the comparison, the contract-theoretic power allocation designed in this article is also applied to the derived short-distance and random user-to-BS associations. Apart from the different association mechanisms used for benchmarking purposes, we adduce an additional comparative analysis considering different spatial user distributions. Specifically, the results presented encompass the cases of both uniform (Fig. 3) and non-uniform (Fig. 7) user distribution. Both wireless network deployments consist of the same number and type of BSs, as well as the same number of users.

Fig. 8(a)-8(c) present the resulting total number of associated users to each BS, as well as the sum of users' transmission power and achieved data rate per BS, after the application of each unified alternative scheme under the uniform user distribution case. Similar results are derived for the case of non-uniform user distribution in Fig. 9(a)-9(c). A comparison between the resulting fairness in terms of the users' achieved data rates under each unified alternative scheme is introduced in Fig. 10(a)-10(b). To evaluate the users' fairness, we consider the converged user-to-BS associations between the users $u \in U$ and BSs $c \in C$ and derive the system's overall fairness applying the fairness criterion known as Jain's index [46], defined as follows

$$J = \frac{\left(\sum_{u=1}^{|U|} R_u^c\right)^2}{|U| \cdot \sum_{u=1}^{|U|} \left(R_u^c\right)^2}, \ J \in [0, 1]. \tag{38}$$

Nominally, Jain's index constitutes an independent of scale criterion that measures how fair and even the performed share of throughput is in distributed computer systems. The higher the index, the fairer the resource allocation is. Hence, the maximum index value is obtained in the case when all users receive the same share of resources, which in our analysis corresponds to the equal share of the data rates.

Considering the uniform user distribution case, there exists a direct discrimination between the unified scheme employing short-distance compared to the SLA and random-based association mechanisms, stemming from the uneven user-to-BS association induced when applying short-distance (Fig. 8(a)). In more detail, due to the shortest intermediate distance between the users and the PBSs and UAVs, a significantly high number of users is ultimately served by the PBSs and UAVs, while an extremely low number of users is associated with the MBS. As a result, the low number of users served by the MBS achieves an extremely high data rate compared to the rest of the users' served by the small BSs of restricted bandwidth resources (Fig. 8(c)), whereas a subtle differentiation is observed in terms of sum users' consumed power across the different unified schemes (Fig. 8(b)). The unfair share in the users' data rate is further corroborated by the resulting Jain's fairness index, depicted in Fig. 10(a). Indeed, the users' fairness when applying the short-distance association mechanism is very low, while the SLA-based and random-based unified schemes present similar behavior under the uniform user distribution.

Subsequently, we investigate the case of non-uniform user distribution (Fig. 7). Emphasizing on the operation of the random-based unified scheme, we observe that there are no adequate incentives for the random association mechanism to uniformly associate the users with different BSs other than the PBS2 and MBS (Fig. 9(a)), in the proximity of which the majority of the users are located (as observed in Fig. 7). Consequently, the remarkably high number of users associated with the PBS2 experience lower data rates compared to the rest of the users (Fig. 9(c)). Also, the large number of users associated with the MBS also achieve low data rates, which are balanced due to the high bandwidth availability that characterizes the MBS. This result is further confirmed by the obtained Jain's fairness index presented in Fig. 10(b). The SLA-enabled user-to-BS association mechanism concludes to a more even and fair share in the users' achieved data rates, by associating them in a uniform manner with different BSs. Also, the results reveal that the users' association with longer-distanced BSs from the users is performed with the cost of increased sum of users' consumed powers (Fig. 9(b)).

VI. CONCLUSION & FUTURE WORK

In this work, a unified resource orchestration framework jointly performing the user-to-BS association and the resulting optimal power allocation in heterogeneous wireless networks is introduced, based on the principles of contract theory and reinforcement learning. Specifically, a reinforcement learning mechanism based on the theory of the stochastic learning automata is adopted to enable the users to select the base station to be associated with, by considering the feedback of the surrounding communications environment and the users' subjective opinion regarding the quality of service provided by the base stations. Then, a contract-theoretic mechanism is proposed that captures the interactions between the base stations and the corresponding users that are served by them, following the labor economics principle. A maximization problem of each base station's utility function is formulated, while jointly considering the optimality of each user's utility function, given its characteristics and type, concluding to the optimal uplink transmission power of each user. A low complexity unified user association and power allocation algorithm is devised. Finally, detailed numerical results are provided, demonstrating the operation and benefits of the proposed user association and power allocation framework.

Part of our current and future work refers to the resource management problem in a multi-tier architecture, targeting the end-to-end communication path. Such a setting, includes not only the radio access part which has been the focus of our work here, but also the backhauling communication link from the various types of cells to the core network. The challenge is that the overall resource management can be treated under the proposed contract-theoretic framework, where the employers' and employees' roles are interchangeable between successive tiers. Such an approach could capture in a formal manner the impact and interdependencies of the resource management decisions and constraints among the different tiers, towards delivering end-to-end resource orchestration. Finally, it is of high research interest and part of our future work, to further investigate and exploit the applicability of the proposed framework in the recently emerging reconfigurable intelligent surface-based systems, which are envisioned to be part of the upcoming 6G wireless networks [47], [48].

REFERENCES

[1] "Cisco annual Internet report (2018–2023)," Cisco, San Jose, CA, USA, White Paper, Mar. 2020.

[2] D. Muirhead, M. A. Imran, and K. Arshad, “A survey of the challenges, opportunities and use of multiple antennas in current and future 5G small cell base stations,” *IEEE Access*, vol. 4, pp. 2952–2964, 2016.

[3] D. Sikeridis, E. E. Tsiropoulou, M. Devetsikiotis, and S. Papavassiliou, “Wireless powered public safety IoT: A UAV-assisted adaptive-learning approach towards energy efficiency,” *J. Netw. Comput. Appl.*, vol. 123, pp. 69–79, Dec. 2018.

[4] B. Li, Z. Fei, and Y. Zhang, “UAV communications for 5G and beyond: Recent advances and future trends,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2241–2263, Apr. 2019.

[5] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, “Risk-aware resource control with flexible 5G access technology interfaces,” in *Proc. IEEE 20th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, Washington, DC, USA, 2019, pp. 1–9.

[6] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, “Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends,” *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015, doi: 10.1109/MCOM.2015.7263349.

[7] Z. Chen, Z. Ding, X. Dai, and R. Zhang, “An optimization perspective of the superiority of NOMA compared to conventional OMA,” *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5191–5202, Oct. 2017.

[8] M. Aldababsa, M. Toka, S. Gökçeli, G. Karabulut Kurt, and O. Kucur, “A tutorial on nonorthogonal multiple access for 5G and beyond,” *Wireless Commun. Mobile Comput.*, vol. 2018, 2018, Art. no. 9713450. [Online]. Available: https://doi.org/10.1155/2018/9713450

[9] G. C. Alexandropoulos and C. B. Papadias, “A reconfigurable iterative algorithm for the K-user MIMO interference channel,” *Signal Process.*, vol. 93, no. 12, pp. 3353–3362, 2013.

[10] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[11] Z. Ding, F. Adachi, and H. V. Poor, “The application of MIMO to non-orthogonal multiple access,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[12] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, “A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[13] A. Asheralieva and Y. Miyanaga, “Optimal contract design for joint user association and intercell interference mitigation in heterogeneous LTE-A networks with asymmetric information,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5284–5300, Jun. 2017.

[14] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, “On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users,” *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[15] Z. Ding, P. Fan, and H. V. Poor, “Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[16] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, “A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[17] S. Timotheou and I. Krikidis, “Fairness for non-orthogonal multiple access in 5G systems,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[18] W. Bao, H. Chen, Y. Li, and B. Vucetic, “Joint rate control and power allocation for non-orthogonal multiple access systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2798–2811, Dec. 2017.

[19] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, “Spectrum allocation and power control for non-orthogonal multiple access in hetnets,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5825–5837, Sep. 2017.

[20] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, “Joint uplink base station association and power control for small-cell networks with non-orthogonal multiple access,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5567–5582, Sep. 2017.

[21] S. Zhang and G. Kang, “User association and power control for energy efficiency maximization in M2M-enabled uplink heterogeneous networks with NOMA,” *Sensors*, vol. 19, no. 23, p. 5307, Dec 2019.

[22] M. S. Elbamby, M. Bennis, W. Saad, M. Debbah, and M. Latva-Aho, “Resource optimization and power allocation in in-band full duplex-enabled non-orthogonal multiple access networks,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2860–2873, Dec. 2017.

[23] A. Celik, M.-C. Tsai, R. M. Radaydeh, F. S. Al-Qahtani, and M.-S. Alouini, “Distributed user clustering and resource allocation for imperfect NOMA in heterogeneous networks,” *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7211–7227, Oct. 2019.

[24] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, “On the performance of non-orthogonal multiple access systems with partial channel information,” *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[25] P. Xu and K. Cumanan, “Optimal power allocation scheme for non-orthogonal multiple access with α-fairness,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2357–2369, Oct. 2017.

[26] Z. Wei, D. W. K. Ng, and J. Yuan, “Power-efficient resource allocation for MC-NOMA with statistical channel state information,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, USA, 2016, pp. 1–7.

[27] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, “Joint user scheduling and power allocation optimization for energy-efficient noma systems with imperfect CSI,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.

[28] Z. Wei, D. W. K. Ng, J. Yuan, and H. Wang, “Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information,” *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.

[29] P. Bolton and M. Dewatripont, *Contract Theory*. Cambridge, MA, USA: MIT, 2005.

[30] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, “Contract-based incentive mechanisms for device-to-device communications in cellular networks,” *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2144–2155, Oct. 2015.

[31] W. Lu, S. Hu, X. Liu, C. He, and Y. Gong, “Incentive mechanism based cooperative spectrum sharing for ofdm cognitive IoT network,” *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 662–672, Apr.–Jun. 2020.

[32] R. Tang, J. Cheng, and Z. Cao, “Contract-based incentive mechanism for cooperative NOMA systems,” *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 172–175, Jan. 2019.

[33] Z. Hasan and V. K. Bhargava, “Relay selection for OFDM wireless systems under asymmetric information: A contract-theory based approach,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3824–3837, Aug. 2013.

[34] D. Prelec, “A Bayesian truth serum for subjective data,” *science*, vol. 306, no. 5695, pp. 462–466, 2004.

[35] P. Sedlmeier and G. Gigerenzer, “Teaching Bayesian reasoning in less than two hours,” *J. Exp. Psychol. Gen.*, vol. 130, no. 3, pp. 380–400, 2001.

[36] T. Van Erven and P. Harremos, “Rényi divergence and Kullback–Leibler divergence,” *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[37] L. M. B. Cabral, *The Economics of Trust and Reputation: A Primer*, New York Univ., New York, NY, USA, and CEPR, London, U.K., 2005.

[38] Y. Xu, J. Wang, and Q. Wu, “Distributed learning of equilibria with incomplete, dynamic, and uncertain information in wireless communication networks,” in *Game Theory Framework Applied to Wireless Communication Networks*. Hershey, PA, USA: IGI Global, 2016, pp. 63–86.

[39] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, “Dynamic provider selection & power resource management in competitive wireless communication markets,” *Mobile Netw. Appl.*, vol. 23, no. 1, pp. 86–99, 2018.

[40] C. Huang, R. Mo, and C. Yuen, “Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Aug. 2020.

[41] J. Nocedal and S. Wright, “Sequential quadratic programming,” in *Numerical Optimization* (Springer Series in Operations Research and Financial Engineering). New York, NY, USA: Springer, 2006, pp. 529–562. [Online]. Available: https://doi.org/10.1007/978-0-387-40065-5_18

[42] T. Coleman, M. A. Branch, and A. Grace, *Optimization Toolbox For Use with MATLAB: User's Guide, Version 2*. Natick, MA, USA: MathWorks Inc., 1990.
[43] P. A. Apostolopoulos, E. E. Tsiropoulou, and S. Papavassiliou, "Risk-aware data offloading in multi-server multi-access edge computing environment," *IEEE/ACM Trans. Netw.*, vol. 28, no. 3, pp. 1405–1418, Jun. 2020.
[44] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, 3GPP Standard 3GPP TR 36.814 V9.2.0, Mar. 2017.
[45] W. Khawaja, I. Guvenc, D. W. Matolak, U. Fiebig, and N. Schneckenburger, "A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2361–2391, 3rd Quart., 2019.
[46] R. Jain, D. M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Sep. 1998. [Online]. Available: https://arxiv.org/abs/cs/9809099.
[47] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, early access, Jul. 8, 2020, doi: 10.1109/MWC.001.1900534.
[48] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

MARIA DIAMANTI received the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki in 2018. She is currently pursuing the Ph.D. degree and an Assistant Researcher with the School of Electrical and Computer Engineering, National Technical University of Athens. Her research interests lie in the areas of 5G wireless networks, resource management and optimization, contract theory, and reinforcement learning.

GEORGIOS FRAGKOS (Graduate Student Member, IEEE) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens in 2018. He is currently pursuing the Ph.D. degree and a Research Assistant with the Department of Electrical and Computer Engineering, University of New Mexico. His main research interests include reinforcement learning, game theory, optimization, contract theory, and blockchain. He was selected by the IEEE Albuquerque Section to received the IEEE Outstanding Graduate Engineering Student Award 2020 for outstanding academic performance in the field of computing and communications as well as by the School of Engineering, University of New Mexico, for the Outstanding Graduate Award 2020.

EIRINI ELENI TSIROPOULOU (Member, IEEE) is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of New Mexico. Her main research interests lie in the area of cyber-physical social systems and wireless heterogeneous networks, with emphasis on network modeling and optimization, resource orchestration in interdependent systems, reinforcement learning, game theory, network economics, and Internet of Things. Four of her papers received the Best Paper Award at IEEE WCNC in 2012, ADHOCNETS in 2015, IEEE/IFIP WMNC 2019, and INFOCOM 2019 by the IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling. She received the NSF CRII Award in 2019 and the Early Career Award by the IEEE Communications Society Internet Technical Committee in 2019. She was selected by the IEEE Communication Society–N2Women-as one of the top ten Rising Stars of 2017 in the communications and networking field.

SYMEON PAPAVASSILIOU (Senior Member, IEEE) is currently a Professor with the School of ECE, National Technical University of Athens. From 1995 to 1999, he was a Senior Technical Staff Member with AT&T Laboratories, Middletown, NJ, USA. In August 1999, he joined the ECE Department, New Jersey Institute of Technology, USA, where he was an Associate Professor until 2004. He has an established record of publications in his field of expertise, with more than 300 technical journal and conference published papers. His main research interests lie in the area of computer communication networks, with emphasis on the analysis, optimization, and performance evaluation of mobile and distributed systems, wireless networks, and complex systems. He received the Best Paper Award in IEEE INFOCOM 94, the AT&T Division Recognition and Achievement Award in 1997, the U.S. National Science Foundation Career Award in 2003, the Best Paper Award in IEEE WCNC 2012, the Excellence in Research Grant in Greece in 2012, the Best Paper Awards in ADHOCNETS 2015, ICT 2016, and IEEE/IFIP WMNC 2019, as well as the 2019 IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling Best Paper Award (for his INFOCOM 2019 Paper). He also served on the board of the Greek National Regulatory Authority on Telecommunications and Posts from 2006 to 2009.