Gambling on Reconfigurable Intelligent Surfaces

Stefan Schwarz, *Senior Member, IEEE*

Abstract—We consider multi-operator wireless networks where broadband reconfigurable intelligent surfaces (RISs) effectively cover the transmission bands of all operators. These RISs are supplied by a dedicated provider and dynamically leased on-demand to individual operators to support their transmissions. When an operator takes control of a RIS, it can adjust its phase-response to meet the requirements of its users. This sets the stage for a competitive scenario where operators vie for control of RISs. To address this competition, we introduce an auction format designed to efficiently allocate RISs to operators. Furthermore, we develop a multi-agent reinforcement learning environment to optimize operators' bidding strategies, demonstrating its superiority over the heuristic dominant strategy of greedy bidding.

Index Terms—Reconfigurable intelligent surfaces, multi-operator mobile networks, auctions, reinforcement learning.

I. INTRODUCTION

IN RECENT years, RISs have demonstrated their ability to enhance the performance of wireless networks across various aspects. They have the capability to improve the capacity, coverage, and energy efficiency of wireless networks [1], [2], [3]. This becomes particularly crucial in mitigating shadowing effects at higher carrier frequencies. Additionally, RISs can enhance the secrecy of wireless transmissions [4] and support wireless localization [5].

In this study, our focus is on incorporating RISs into multi-operator wireless networks. Given their capacity for achieving wide-band transmission [6], RISs can effectively cover the frequency bands of multiple operators. This presents a coexistence challenge, as a RIS response setup suitable for one operator may compromise the performance of another. This challenge is also highlighted in [7], where the so-called *bandwidth-of-influence*, i.e. the frequency band over which a RIS significantly impacts incident signals, of several RIS-types is experimentally characterized, showing strong influence of the applied technology. Multi-operator RIS deployments are also compromised by inter-operator pilot contamination [8].

To overcome this coexistence challenge, we consider a network model where RISs are not owned by individual operators but by a dedicated RIS provider/controller. This provider sells on-demand leases of RISs in a free market to individual operators, supporting transmissions within their networks. This dynamic introduces a unique competitive landscape, which has not been investigated before, where operators actively compete for control over RISs. We propose the implementation

of an auction format tailored to efficiently allocate RISs among operators and develop a multi-agent reinforcement learning (RL) framework to dynamically optimize operators' bidding strategies. Through our experiments, we substantiate the effectiveness of this approach, showcasing its superiority over the heuristic dominant strategy of greedy bidding. This not only enhances the overall efficiency of RIS allocation but also underscores the adaptability achievable through the integration of RL into the auction process for multi-operator RIS-assisted wireless networks. The proposed auction format could be realized within the RIS-enabled smart wireless environment architecture of [7]. RL has previously proven effective in optimizing the phase-response of RIS alongside beamforming weights for a base station within a single-operator setup [9]. Conversely, our study demonstrates that RL can effectively direct the allocation of RISs in a multi-operator environment.

Notation: The complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ is $\mathcal{CN}\left(\boldsymbol{\mu}, \mathbf{C}\right)$. The uniform distribution over the interval $[a, b]$ is $\mathcal{U}(a, b)$. The transpose of vector $\mathbf{x}$ is $\mathbf{x}^{\mathsf{T}}$ and the i-th element is $\mathbf{x}[i]$. The empty set is $\emptyset$. The size of set $\mathcal{X}$ is $|\mathcal{X}|$. The indicator of a is $\mathbb{1}(a)$, i.e. $\mathbb{1}(a) = 1$ if $a \neq 0$ and zero otherwise. The phase of complex number z is $\arg(z)$. The expected value of random variable r is $\mathbb{E}(r)$.

II. SYSTEM MODEL

We consider the mobile networks of N_{O} operators, each serving the same geographic region through their separate and non-interfering bands centered around the carrier frequency f_c. Operator o serves $N_{\mathrm{U}}^{(o)}$ users via $N_{\mathrm{B}}^{(o)}$ base stations. We assume that users and base stations are equipped with single antennas. We consider orthogonal frequency division multiple access (OFDMA) transmissions at each base station, so that transmissions are only impaired by inter-cell-interference. These assumptions may appear simplistic; however, they are fundamentally unessential for the auction problem described later, and should be regarded as illustrative, serving the purpose of simplifying the formulation while maintaining relevance.

Under these assumptions, we obtain the following frequency-flat per-subcarrier single-input single-output (SISO) down-link input-output relationship of user u served by operator o

$$y_u^{(o)} = h_{u,d_u}^{(o)} x_{d_u}^{(o)} + \sum_{i=1, i \neq d_u}^{N_{\mathrm{B}}^{(o)}} h_{u,i}^{(o)} x_i^{(o)} + n_u^{(o)}, \quad (1)$$

where d_u denotes the index of the base station serving user u, $h_{u,\ell}^{(o)}$ is the channel between the user and base station ℓ, and $n_u^{(o)} \sim \mathcal{CN}\left(0, \sigma_n^2\right)$ is noise.

Within the considered region, N_{R} RISs assist the transmissions between base stations and users. We consider the most basic idealistic model for the interaction of RISs

with wireless signals through frequency-flat diagonal phase-response matrices. While more accurate models exist [10], their incorporation might potentially obscure the main focus in this work. With appropriate modifications of the RIS response and the RIS-assisted channels defined below, the proposed setup can also be generalized to more advanced RIS technologies, such as intelligent omni-surfaces [11], stacked intelligent metasurfaces [12] and non-diagonal RIS; we leave this for future work.

The RISs are owned by an independent RIS provider, who offers on-demand leases of the RISs to the highest bidding operator. We assume that the RISs are sufficiently broadband to effectively cover the transmission bands of all operators. If an operator takes control of a RIS, it can set the phase-response of the RIS to maximize the performance of its users. In particular, considering RIS r consisting of M discrete reconfigurable elements, the RIS response is governed by a diagonal matrix $\boldsymbol{\Phi}_r = \mathrm{diag}\left(e^{\mathrm{j}\phi_1^{(r)}}, \ldots, e^{\mathrm{j}\phi_M^{(r)}}\right) \in \mathbb{C}^{M \times M}$. Although the operators transmit in separate, non-overlapping frequency bands, this still leads to a mutual coupling effect between them. In particular, the optimization of the RIS response for one operator can negatively impact the performance of another operator. For instance, it may result in increased inter-cell interference for the non-controlling operator or create destructive multi-path interference for users' channels.

The users' channels are comprised of a direct component $h_{u,\ell}^{(o,d)}$ superimposed by a RIS-assisted component $h_{u,\ell}^{(o,a)}$

$$h_{u,\ell}^{(o)} = h_{u,\ell}^{(o,d)} + h_{u,\ell}^{(o,a)}. \tag{2}$$

The RIS-assisted component can further be written as

$$h_{u,\ell}^{(o,a)} = \sum_{r=1}^{N_{\mathrm{R}}} \left(\mathbf{h}_{u,r}^{(o)}\right)^{\mathsf{T}} \boldsymbol{\Phi}_r \mathbf{h}_{r,\ell}^{(o)}, \tag{3}$$

with $\mathbf{h}_{u,r}^{(o)}, \mathbf{h}_{r,\ell}^{(o)} \in \mathbb{C}^{M \times 1}$ denoting the vector-valued user-to-RIS and RIS-to-base station channels, respectively.

The users' signal to interference and noise ratios (SINRs) $\beta_u^{(o)}$ and achievable rates $r_u^{(o)}$ are

$$\beta_u^{(o)} = \frac{\left|h_{u,d_u}^{(o)}\right|^2 P_{d_u}^{(o)}}{\sigma_n^2 + \sum_{\substack{i=1, \\ i \neq d_u}}^{N_{\mathrm{B}}^{(o)}} \left|h_{u,i}^{(o)}\right|^2 P_i^{(o)}}, \quad r_u^{(o)} = \log_2\left(1 + \beta_u^{(o)}\right), \tag{4}$$

with $P_\ell^{(o)}$ denoting the transmit power of base station ℓ of operator o. For simplicity we consider $P_\ell^{(o)} = P, \forall \ell, o$.

III. AUCTION-BASED RIS ALLOCATION

We now develop an auction-based allocation of RISs to operators. To conduct this auction, operators must be able to estimate the utility of a RIS allocation, i.e. how much a given allocation improves the performance of their networks. This estimation cannot be based on perfect channel knowledge, since these channels can only be accurately estimated once the RISs have been assigned and pilot signals have been transmitted. To obtain a coarse utility estimate, we first derive SINR and rate expressions that are based only on macroscopic channel properties which are relatively easy to observe. Based on these utility estimates, we then develop a low complexity simultaneously ascending auction format in Section III-B.

A. Utility Estimation

a) Utility function: Consider bidding for a subset $\mathcal{R} \subseteq \{1, \ldots, N_{\mathrm{R}}\}$ of RISs. To determine the appropriate bid amount for this specific subset, each operator must assess the utility or valuation of this subset. Hence, we establish a utility function by calculating the sum of exponentiated rates for the operator

$$U^{(o)}(\mathcal{R}) = \frac{\sum_{u=1}^{N_{\mathrm{U}}^{(o)}} \left(\bar{r}_u^{(o)}(\mathcal{R})\right)^{1/\alpha}}{\sum_{u=1}^{N_{\mathrm{U}}^{(o)}} \left(\bar{r}_u^{(o)}(\emptyset)\right)^{1/\alpha}} - 1, \tag{5}$$

where $\bar{r}_u^{(o)}(\mathcal{R})$ is an estimate of the rate achieved by user u when RISs $\mathcal{R}$ are controlled by the operator. We use the percentage improvement compared to allocating no RISs to the operator. The parameter $\alpha \in (0, \infty)$ allows to gauge the fairness of user rates. When α approaches 0, the operator prioritizes higher user rates, whereas as α approaches infinity, the emphasis shifts towards equalizing the rates of all users.

b) Channel model: To illustrate the estimation of the rate $\bar{r}_u^{(o)}(\mathcal{R})$, we adopt a specific simple geometry-based wireless channel model. In particular, consider a vector-valued channel $\mathbf{h} \in \mathbb{C}^{M \times 1}$ of an arbitrary link between two nodes (users, base stations, RISs); for SISO channels we just set $M = 1$

$$\mathbf{h} = \gamma \left(\underbrace{\sqrt{\frac{K}{1+K}} e^{\mathrm{j}\varphi} \mathbf{a}(\theta)}_{k} + \underbrace{\sqrt{\frac{1}{1+K}} \mathbf{g}}_{\bar{k}}\right). \tag{6}$$

Here, γ denotes the distance-dependent macroscopic path loss. It also depends on the propagation conditions of the link, i.e. whether it is in line-of-sight (LOS) or non-line-of-sight (NLOS); in the simulations these factors are determined for each link individually. The Rician K-factor serves to linearly combine a directional path $e^{\mathrm{j}\varphi} \mathbf{a}(\theta)$ with a random scattering component $\mathbf{g} \sim \mathcal{CN}(0, \mathbf{I}_M)$. Here, $\mathbf{a}(\theta)$ is the RIS response vector w.r.t. a plane-wave in angular direction θ and φ is a propagation distance/delay dependent phase-shift. For links related to non-serving base stations, we assume that φ follows a uniform distribution $\varphi \sim \mathcal{U}(0, 2\pi)$, whereas, for channels associated with the serving base station, we set $\varphi = 0$ assuming perfect synchronization. We consider different K-factors depending on the LOS/NLOS conditions of the link. This model is applied to all channels in (2),(3).

c) SINR estimation: To estimate the SINR and rate of user u, we replace the instantaneous channel gains $|h_{u,\ell}^{(o)}|^2$ in (4) with their expected values. For this, we assume that $\boldsymbol{\Phi}_r$ is optimized for the user if the RIS is controlled by the corresponding operator $r \in \mathcal{R}$, whereas it is random $\phi_i^{(r)} \sim \mathcal{U}(0, 2\pi)$ otherwise. The optimized RIS response is

$$\phi_i^{(r)} = -\left(\arg\left(\mathbf{a}\left(\theta_{u,r}^{(o)}\right)[i]\right) + \arg\left(\mathbf{a}\left(\theta_{r,d_u}^{(o)}\right)[i]\right)\right), \tag{7}$$

which only accounts for the directional part in (6), because the scattering part varies too quickly to adjust the RIS for it.

For channels $\mathbf{h}_{r,\ell}^{(o)}$ between base stations and RISs, we assume that the directional path is dominant, i.e. the K-factor is so large that the Gaussian scattering component can be neglected ($k_{r,\ell}^{(o)} = 1, \bar{k}_{r,\ell}^{(o)} = 0$). This is justified considering that RISs are intentionally placed by the RIS provider to support the operators' base stations. Under these assumptions,

the expected channel gain of the intended signal becomes

$$\mathbb{E}\left(\left|h_{u,d_u}^{(o)}\right|^2\right) = p_c + p_i + p_u, \quad p_u = M\sum_{r\notin\mathcal{R}}\left(\gamma_{u,r}^{(o)}\gamma_{r,d_u}^{(o)}\right)^2,$$

$$p_c = \left(\gamma_{u,d_u}^{(o,d)}k_{u,d_u}^{(o,d)} + M\sum_{r\in\mathcal{R}}k_{u,r}^{(o)}\gamma_{u,r}^{(o)}\gamma_{r,d_u}^{(o)}\right)^2,$$

$$p_i = \left(\gamma_{u,d_u}^{(o,d)}\bar{k}_{u,d_u}^{(o,d)}\right)^2 + M\sum_{r\in\mathcal{R}}\left(\bar{k}_{u,r}^{(o)}\gamma_{u,r}^{(o)}\gamma_{r,d_u}^{(o)}\right)^2.$$

(8)

Here, p_c contains the coherently combined signals received over the directional components of the direct and RIS-assisted channels. Term p_i contains the incoherently combined signals received over the Gaussian scattering components. These are incoherently combined, because the RIS responses $\mathbf{\Phi}_r$ of controlled RISs $r \in \mathcal{R}$ only compensate for the phase shifts of the directional components. Finally, p_u is the incoherent combination of signals received from RISs not controlled by the operator. These are incoherently combined, because their RIS response matrices $\mathbf{\Phi}_r, r \notin \mathcal{R}$ are random.

Equivalent expressions are derived for channels of interfering signals $|h_{u,i}^{(o)}|^2, i \neq d_u$. A difference arises in the coherently combined signal power p_c. This is due to the fact that the RIS responses $\mathbf{\Phi}_r$, optimized in (7) for the serving base station $\mathbf{a}(\theta_{r,d_u}^{(o)})$, generally do not align with the channels of interferers $\mathbf{a}(\theta_{r,i}^{(o)})$. This leads to a reduction of the interfering signals by $\left|\mathbf{a}(\theta_{r,d_u}^{(o)})^\mathsf{T}\mathbf{a}(\theta_{r,i}^{(o)})\right|$. All necessary terms for computing average signal powers can thus be derived from macroscopic channel properties. Substituting into (4) results in $\bar{r}_u^{(o)}(\mathcal{R})$.

B. Auction Format

The allocation of RISs to operators can be handled using a combinatorial auction, similar to spectrum auctions. However, our goal is to conduct this auction dynamically during live operation of the operators' networks, depending on demand. We cannot execute a complex combinatorial auction using a direct-revelation mechanism for valuations of all potential RIS subset allocations (e.g., employing the Vickrey-Clarke-Groves (VCG) mechanism [13]). This would demand excessive computation and communication overhead to exchange valuations with the auctioneer. Note that the valuation/utility (5) of a subset $\mathcal{R}$ is not simply the sum of the utilities of the individual RISs.

We therefore consider a low-complexity indirect auction mechanism, in particular a simultaneously ascending "Japanese" forward auction [14]. In round t of the auction, the auctioneer (RIS provider) sets a uniform price $p_t > p_{t-1}$ for each RIS, starting from an initially low price p_0. In each round, the price is increased by a fixed increment $\Delta_p = p_t - p_{t-1}$. The bidders (operators) submit their bids in the form of binary vectors $\mathbf{b}^{(o)} \in \{0,1\}^{N_{\text{RIS}}}$, indicating their willingness to pay the current price for specific RIS units. If a particular RIS unit receives a bid from only one operator, the auctioneer accepts the bid, and the RIS is allocated to that operator. If there are no further bids for a particular RIS, it remains unassigned and its response is set randomly. The auction concludes when all RISs are assigned or when there are no more bids. Furthermore, we implement a consistent bidding activity rule, meaning that operators are not allowed to bid for a particular RIS in round

t if they did not place a bid for the same RIS in round $t - 1$ (rule-defying bids are ignored by the auctioneer). This rule facilitates the identification of the preferences amongst operators.

Alternatively a descending auction could be used, where the price starts high and falls until a bidder accepts it. However, ascending auctions are more bidder-friendly as every participant can observe the interest of other bidders from the beginning. Through our experiments, we have noticed that the RL agent learns more effectively using the adopted ascending format.

IV. BIDDING STRATEGIES

Consider round t of the auction: we denote by $\mathcal{R}_{t-1}^{(o)}$ the set of RISs that have been allocated to operator o in previous rounds. The set of remaining RISs is $\mathcal{R}_t = \{1,\ldots,N_\text{R}\} \setminus \bigcup_{o=1}^{N_o}\mathcal{R}_{t-1}^{(o)}$. Each bidder has to decide for which of the remaining RISs in $\mathcal{R}_t$ it is willing to pay the current price p_t. Therefore, each bidder has to estimate the value of adding RIS $r \in \mathcal{R}_t$ to its already allocated RISs $\mathcal{R}_{t-1}^{(o)}$. In principle, this requires determining the average utility of incorporating RIS r alongside all combinations of the remaining RISs in $\mathcal{R}_t$. This is necessary because the operator lacks knowledge about which other RISs it might secure. However, for a larger number of RISs, the complexity of this combinatorial approach becomes impractical. Consequently, we simplify the calculation by assessing the value of acquiring RIS r with the assumption that it would be the sole RIS secured by the bidder

$$V_t^{(o)}(r) = U^{(o)}\left(\mathcal{R}_{t-1}^{(o)} \cup r\right) - U^{(o)}\left(\mathcal{R}_{t-1}^{(o)}\right). \quad (9)$$

A. Heuristic Bidding

In the given auction setup, each bidder has a dominant strategy, namely to stay in the auction as long as the valuation of a RIS unit is higher than the price. To compare the percentage value in (9) to the price p_t, we introduce a constant $c_V^{(o)}$ that represents the maximum price the operator is willing to pay for 100 % improvement. Using this approach, we identify the set of RISs that may be worth bidding on

$$\varrho_t^{(o)} = \left\{r \in \mathcal{R}_t \middle| c_V^{(o)}V_t^{(o)}(r) \geq p_t\right\}. \quad (10)$$

However, if the operator bids on all of them, it risks potentially surpassing its available budget. Let $B_t^{(o)}$ represent the remaining budget of operator o at round t, calculated as the initial budget $B_0^{(o)}$ at the start of the auction minus the costs paid for $\mathcal{R}_{t-1}^{(o)}$. A conservative strategy is to bid only on the $\lfloor B_t^{(o)}/p_t \rfloor$ most valuable RISs in $\varrho_t^{(o)}$. Without a strict budget cap, the operator may bid on more items, knowing it may not secure all of them. Nevertheless, in our simulations we apply a strict budget cap and have found that the conservative approach yields the highest reward as defined below.

B. RL-Based Bidding

Alternatively to this greedy approach, we can train RL agents for each operator to learn optimized bidding strategies. In our simulations, we investigate using the same agent for each operator individually, but there is also the possibility of training a RL agent to compete against other strategies. Below we specify the state, observation and action spaces, and the

rewards utilized for RL training. Note that the observations are based only on the information available to each operator individually, without any exchange of information between them.

a) States: The state of our auction environment at time t is determined by the following variables

$$\mathcal{S}_t = \left(p_t, \mathcal{R}_t, \left\{ \mathcal{R}_{t-1}^{(o)}, V_t^{(o)}(r), B_t^{(o)} | \forall o, r \right\} \right) \qquad (11)$$

b) Observations: The partial observation of operator/agent o at time t contains

$$\mathcal{O}_t^{(o)} = \left(p_t, B_t^{(o)}, \left\{ V_t^{(o)}(r) | \forall r \right\} \right). \qquad (12)$$

To maintain a fixed-size observation space, a requirement for existing algorithms, we assign a value $V_t^{(o)}(r)$ of 0 to RIS units that are no longer available. This applies to RISs that have already been allocated in previous rounds, as well as, to RISs that the operator did not bid on in previous rounds, enforcing the mentioned activity rule. This approach also allows using the same trained agent in environments with varying numbers of RIS units. We achieve this by defining spaces large enough to cover a maximum number of RISs and setting values to zero if the actual number of RISs is less than the maximum allowed. However, it is advisable not to unnecessarily increase the complexity of the RL algorithm with larger spaces.

c) Actions: Based on $\mathcal{O}_t^{(o)}$, each agent makes bidding decisions using the binary action vector $\mathbf{b}^{(o)} \in \{0,1\}^{N_{\text{RIS}}}$.

d) Rewards: Assume that the agent wins the set $w_t^{(o)} \subseteq \mathcal{R}_t$ of RISs in round t. The reward is then composed of three contributions $r_t^{(o)} = r_1^{(o)} + r_2^{(o)} + r_3^{(o)}$:

- Reward according to the value of allocated RISs

$$r_1^{(o)} = c_V^{(o)} V_t^{(o)}\left(w_t^{(o)} \right) - p_t \left| w_t^{(o)} \right|. \qquad (13)$$

- Penalize invalid bids on RISs with value 0

$$r_2^{(o)} = -c_P^{(o)} \sum_{r=1}^{N_r} \mathbf{b}^{(o)}[r] \left(1 - \mathbb{1}\left(V_t^{(o)}(r) \right) \right), \qquad (14)$$

where $c_P^{(o)}$ denotes a tunable punishment factor.

- Penalize bids that cause exceeding the available budget

$$r_3^{(o)} = -c_P^{(o)} \left(\left| w_t^{(o)} \right| - \lfloor B_t^{(o)}/p_t \rfloor \right). \qquad (15)$$

It is not possible to strictly enforce staying within the available budget with RL. Therefore $c_P^{(o)}$ must be chosen sufficiently large to keep such occurrences to a minimum.

e) Implementation: [1] We used the proximal policy optimization algorithm [15] of Stable-Baselines3 2.1.0 [16] to build our RL agent, using its default hyper-parameters. The multi-agent RL auction environment was realized using Gymnasium 0.29.0 [17] and PettingZoo 1.24.1 [18] combined with Supersuit 3.9.0 [19] to create a vector environment for multi-agent training. In such environments, it is necessary to normalize the continuous state and observation spaces to finite ranges. This requires some experimentation with the environment, since the possible range of values in (9) depends on the environment geometry, and the number of RISs N_{R} and RIS elements M.

---

[1]Code is available at https://github.com/StefanSchwarzTUW.

|	
Carrier frequency	$f_c = 26\,\text{GHz}$
Number of base stations	$N_{\text{B}}^{(o)} = 3$
Number of users	$N_{\text{U}}^{(o)} = 10$
Number of RISs	$N_{\text{R}} = 20$
Number of RIS elements	$M = 1000$
Transmit power per subcarrier	$P = 100\,\text{mW}$
Subcarrier bandwidth	$15\,\text{kHz}$
AWGN noise power spectral density	-174 dBm/Hz
Noise figure	$6\,\text{dB}$
Path-loss exponent under LOS (NLOS)	2 (3.75)
K-factor under LOS (NLOS)	100 (3)
Distance-dependent LOS-probability	$p_{\text{LOS}}(d) = e^{-d/50}$
Shadow fading variance	$10\,\text{dB}$
Auction initial price	$p_0 = 0$
Auction price increment	$\Delta_p = 0.05$
Budget (if not mentioned otherwise)	$B_0^{(o)} = 1$
Utility function parameter	$\alpha = 1$
Value constant	$c_V^{(o)} = 1$
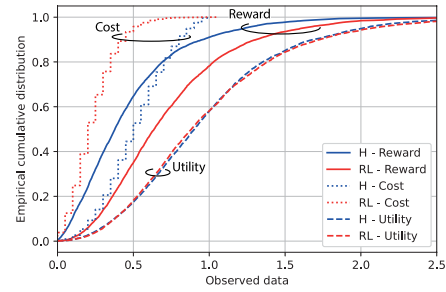
Fig. 1. Performance comparison of heuristic (H) and RL-based bidding.

V. SIMULATIONS

We consider transmissions in millimeter wave wireless networks of $N_{\text{O}} = 2$ operators. We consider a region of interest of $100\,\text{m}^2$ where we randomly place users. The base stations and RISs are arranged in a regular grid, with an additional layer of randomness introduced through a Gaussian distortion with a standard deviation of 20 meters. We assume that RISs are intentionally placed such that there is a LOS path between RISs and base stations (e.g., on top of buildings). We also assume that the direct paths between base stations and users are blocked, such that NLOS propagation conditions apply. For the channels between RISs and users, we consider a distance-dependent LOS probability as specified in Table I.

The RL agent undergoes training through episodes, each aligning with a distinct auction. The episode length is thereby variable, depending on the number of steps to complete the auction. In every episode, a fresh environment is created, complete with randomized positions and wireless channel realizations. Invalid actions are penalized by a punishment factor $c_P^{(o)}$ that is ten times larger than the largest possible value $V_t^{(o)}$. The agent is trained for $3 \cdot 10^5$ auction steps.

In Fig. 1, we observe that the RL agent outperforms the heuristic approach and achieves a higher reward. Breaking it down into costs and utility achieved by the RIS allocation, we can see that both approaches essentially achieve the same utility; however, the RL agent incurs significantly lower costs. This means that the agent implicitly learns to coordinate the bids of operators: fundamentally, the RL agent only bids on
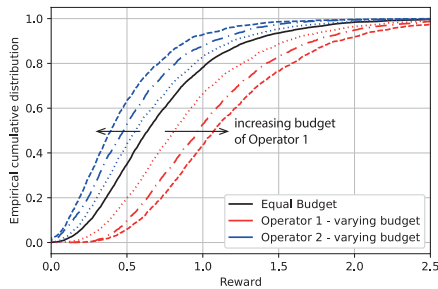
Fig. 2. Impact of available budget. The budget of operator 1 varies within $B_0^{(1)} \in \{1, 2, 4, 8\}$ while the budget of operator 2 is fixed to $B_0^{(2)} = 1$.
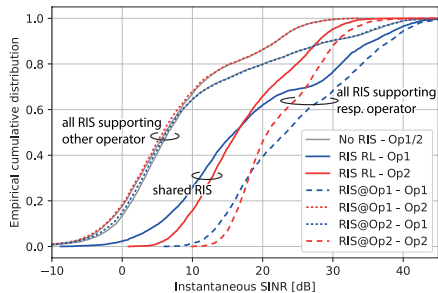

Fig. 3. Distribution of instantaneous SINRs for fixed random positions of network elements and varying microscopic fading channel conditions.

the most valuable RISs, while not driving the price of other RISs up. The heuristic, on the other hand, drives the price of all RISs up until one operator drops out. It is worth noting that each operator uses its own isolated instance of the RL agent, observing only its values and not those of others. As a result, there is no direct collaboration between the operators.

Next, we vary the budget of operator 1, while keeping the budget of operator 2 constant. This enables operator 1 to acquire a larger number of RISs, thereby improving its reward; see Fir. 2. The achieved result, coming from an agent trained with a budget set to one, implies that adjusting the budget does not require retraining the agent. This is also true for scenarios with varying numbers of users and base stations, although not explicitly shown in the simulations due to space constraints. The agent, trained solely on observing the RIS allocation value, does not rely on knowledge of specific values for N_U and N_B.

The results so far are based on the utility estimates derived in Sec. III-A. Next, we therefore demonstrate the behavior in terms of the instantaneous SINR (4) for fixed positions and random microscopic fading channels. In Fig. 3, we show the distribution of the users' SINRs for various RIS allocations. Gray curves depict performance in the absence of RISs, essentially mirroring the operators' performance when RISs are present but remain unallocated to them (dotted curves). The notation *RIS@OpK - OpL* means that all RISs are assigned to operator K and we observe the performance of operator L. Dashed curves show the performance when all RISs support the respective operator,[2] while solid lines show the performance achieved by the RL agent. There is a substantial enhancement in performance when compared to scenarios without available RISs or when they are not explicitly allocated to the operators. Therefore, strategically

sharing RISs between operators can significantly improve the performance of both networks at the same time.

VI. CONCLUSION

In this study, we delved into multi-operator wireless networks supported by RISs supplied by a dedicated RIS provider. Introducing an auction format for this RIS market, we illustrated the applicability of RL in learning effective bidding strategies. The outcomes from our simulations showcased significant enhancements in SINR and data rates, underscoring improved performance for all participating network operators.

REFERENCES

[1] M. A. Msleh, F. Heliot, and R. Tafazolli, "Ergodic capacity analysis of reconfigurable intelligent surface assisted MIMO systems over Rayleigh-Rician channels," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 75–79, Jan. 2023.

[2] S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 269–273, Jan. 2021.

[3] Q. N. Le, V.-D. Nguyen, O. A. Dobre, and R. Zhao, "Energy efficiency maximization in RIS-aided cell-free network with limited backhaul," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 1974–1978, Jun. 2021.

[4] M. H. Khoshafa, T. M. N. Ngatched, and M. H. Ahmed, "Reconfigurable intelligent surfaces-aided physical layer security enhancement in D2D underlay communications," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1443–1447, May 2021.

[5] M. Luan, B. Wang, Y. Zhao, Z. Feng, and F. Hu, "Phase design and near-field target localization for RIS-assisted regional localization system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1766–1777, Feb. 2022.

[6] Q. Hu, H. Yang, X. Zeng, and X. Y. Zhang, "Wideband reconfigurable intelligent surface using dual-resonance element," *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 10, pp. 2422–2426, Jun. 2023.

[7] G. C. Alexandropoulos et al., "RIS-enabled smart wireless environments: Deployment scenarios, network architecture, bandwidth and area of influence," *EURASIP J. Wireless Commun. Netw.*, vol. 2023, no. 1, pp. 1–43, Oct. 2023.

[8] D. Gürgünoğlu, E. Björnson, and G. Fodor, "Combating inter-operator pilot contamination in reconfigurable intelligent surfaces assisted multi-operator networks," 2023, *arXiv:2311.01151*.

[9] W. Xu, J. An, C. Huang, L. Gan, and C. Yuen, "Deep reinforcement learning based on location-aware imitation environment for RIS-aided mmWave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1493–1497, Jul. 2022.

[10] M. Di Renzo, F. H. Danufane, and S. Tretyakov, "Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization," *Proc. IEEE*, vol. 110, no. 9, pp. 1164–1209, Sep. 2022.

[11] H. Zhang et al., "Intelligent omni-surfaces for full-dimensional wireless communications: Principles, technology, and implementation," *IEEE Commun. Mag.*, vol. 60, no. 2, pp. 39–45, Feb. 2022.

[12] J. An et al., "Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2380–2396, Aug. 2023.

[13] N. Nisan and A. Ronen, "Algorithmic mechanism design," *Games Econ. Behav.*, vol. 35, no. 1, pp. 166–196, 2001.

[14] T. Roughgarden, *Twenty Lectures Algorithmic Game Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

[16] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable reinforcement learning implementations," *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-1364.html

[17] M. Towers et al. (Mar. 2023). *Gymnasium*. [Online]. Available: https://zenodo.org/record/8127025

[18] J. K. Terry et al., "PettingZoo: Gym for multi-agent reinforcement learning," 2021, *arXiv:2009.14471*.

[19] J. K. Terry, B. Black, and A. Hari, "SuperSuit: Simple microwrappers for reinforcement learning environments," 2020, *arXiv:2008.08932*.

---

[2]Simultaneously achieving both dashed curves is not possible since RISs can only be assigned to a single operator.