Combinatorial Multi-armed Bandit Algorithms for Real-time Energy Trading in Green C-RAN

Wan Nur Suryani Firuz Wan Ariffin, Xinruo Zhang and Mohammad Reza Nakhai
Centre for Telecommunications Research, King’s College London, Strand, UK, WC2R 2LS
E-mail: [k1206546, xinruo.zhang, reza.nakhai]@kcl.ac.uk

Abstract—Without a proper observation of the energy demand of the receiving terminals, the retailer may be obliged to purchase additional energy from the real-time market and may take the risk of losing profit. This paper proposes two combinatorial multi-armed bandit (CMAB) strategies in green cloud radio access network (C-RAN) with simultaneous wireless information and power transfer under the assumption that no initial knowledge of forthcoming energy demand and renewable energy supply are known to the central processor. The aim of the proposed strategies is to find the set of optimal sizes of the energy packages to be purchased from the day-ahead market by observing the instantaneous energy demand and learning from the behaviour of cooperative energy trading, so that the total cost of the retailer can be minimized. Two novel iterative algorithms, namely, ForCMAB energy trading and RevCMAB energy trading are introduced to search for the optimal set of energy packages in ascending and descending order of package sizes, respectively. Simulation results indicate that CMAB approach in our proposed strategies offers the significant advantage in terms of reducing overall energy cost of the retailer, as compared to other schemes without learning-based optimization.

I. INTRODUCTION

Cloud radio access network (C-RAN) has illustrated its considerable advantages in inter-cell interference (ICI) mitigation, system throughput improvement and reducing the capital expenditure and operating expense (OPEX) of the network operator, i.e., the retailer. Recently, the integration of C-RAN and simultaneous wireless information and power transfer (SWIPT), where the signals transmitted from remote radio heads (RRHs) can be exploited by the battery limited energy receiving terminals (ETs) for self-sustainability, has attracted the attention of researchers [1] and [2]. However, energy cost has become a major OPEX due to dramatic rise of energy consumption induced by the high density of RRHs deployment [3]. Consequently, the retailer may be obliged to purchase additional energy from the grid in the real-time market to cover extra energy requirements and may take a risk of losing profit. Subsequently, equipping the RRHs with renewable energy harvesting devices that can generate local renewable energy from environmental sources for green communications has been considered as a promising technique to benefit both the environment and the retailer. With the implementation of two-way energy trading with the grid, the retailer can maximally benefit from utilising their locally generated renewable energy and selling the excessive energy back to the grid [4]. Provided that all the base stations (BSs) are equipped with renewable energy harvesters and implemented with two-way energy trading, [4] proposes a joint energy trading and full cooperation scheme in coordinated multipoint (CoMP) network, where the data of all the users is available at the central processor (CP) and will be distributed to all BSs for cooperative transmission via backhaul links. However, the data circulation between the CP and the BSs requires huge backhaul signalling overhead when full coordination is enabled [5]. This scheme, nevertheless, takes no consideration of backhaul capacity restrictions, which may be infeasible for practical capacity-constrained backhaul links.

Motivated by the literature that sparse beamforming problem is commonly formulated as a $\ell_0$-norm optimization problem and handled with reweighted $\ell_1$-norm method introduced in [6], the authors in [1], [7]–[10] propose dynamic sparse beamforming designs subject to quality of service (QoS) constraints for capacity-limited backhaul links in C-RAN. The authors in [1] integrate the aforementioned works with SWIPT concept and study the resource allocation algorithm. However, none of them take into account the renewable energy sources that can be further extended to the joint management of the resource allocation and energy trading for green communication. The joint cooperative resource management and energy trading problem in downlink green C-RAN are tackled in our previous works [2] and [11]. The former proposes a real-time BS energy management strategy in a green CoMP network with SWIPT, where the CP jointly minimizes the energy consumption and energy trading, in order to reduce the overall energy cost of the retailer. The latter studies three different cooperative real-time energy trading strategies in C-RAN and applies a sparse beamforming technique to find the optimal trade-off between the degree of partial cooperation among the RRHs in serving the receiving terminals and the total energy cost of the retailer. However, in both papers, we assume that the retailer purchases a set of fixed sizes of energy packages for the RRHs from the day-ahead market without considering the actual amount of local renewable energy generation at RRHs as well as the changes in energy requirements of the information receiving terminals (ITs) and ETs.

In this paper, we further extend our previous works to a learning-based practical approach and assume that the CP has no initial knowledge of forthcoming power budget and energy consumption at the individual RRHs. With various sizes of energy packages which are offered in the day-ahead market, the new responsibility of the CP is to determine the set of optimal sizes of the energy packages to be purchased for the RRHs on the basis of actual energy supply and demand, to further minimize the total energy cost of the retailer. We model this problem as a combinatorial multi-armed bandit (CMAB) problem and introduce two algorithms, namely, ForCMAB Energy Trading and RevCMAB Energy Trading, to observe the
energy demand and search for the set of optimal sizes of the energy packages to be purchased from the day-ahead market.

The rest of this paper is organized as follows. Section II introduces the system model. In section III, a real-time cooperative energy trading strategy is formulated and then transformed into numerically tractable form. In section IV, two combinatorial multi-armed bandit algorithms are proposed. Numerical simulation results are analyzed in section V. Finally, section VI summarizes the paper. Notations: \( w, w', W, (\cdot)^H \) and \( tr(\cdot) \), respectively, represent a scalar \( w \), a vector \( w \), a matrix \( W \), the complex conjugate transpose operators and the trace operators. \( W \geq 0 \) denotes that \( W \) is a positive semidefinite matrix and \( C_{n \times m} \) indicates the sets of \( n \)-by-\( m \) dimensional complex matrices. \( \| \cdot \|_p \) denotes the \( \ell_p \)-norm of a vector and \( \| \cdot \|_0 \) indicates the number of non-zero entries in the vector. Note that, the normalized energy unit, i.e., \( J s^{-1} \), is adopted in this paper, therefore, the terms ‘energy’ and ‘power’ are mutually convertible.

II. SYSTEM MODEL

We consider a downlink C-RAN with SWIPT from \( N \) RRHs, equipped with \( M \) antennas, towards \( K_i \) active single-antenna ITs and \( K_i \) active single-antenna ETs, respectively, over a shared bandwidth. The CP is the central processing unit that coordinates all the cooperative strategies for the RRHs based on perfect knowledge of channel state information and distributes data of all the ITs to the corresponding RRHs via backhaul links. Let \( \mathcal{L}_b = \{1, \ldots, N\}, \mathcal{L}_c = \{1, \ldots, K_c\}, \mathcal{L}_{[\text{idt}]} = \{1, \ldots, K_{[\text{idt}]}\} \) and \( \mathcal{L}_i = \{1, \ldots, K_i\} \) indicate, respectively, the set of indexes of the RRHs, the active ETs, the idle ETs and the active ITs. We further assume that the energy transmission between the CP and RRHs is accomplished via dedicated power lines.

A. Energy Management model

We assume that at least one renewable energy harvesting device, e.g., wind turbine and/or solar panel, is installed at each RRH in order to generate local renewable energy from environmental sources. Furthermore, no RRH is equipped with frequently rechargeable storage devices and the RRHs are obliged to transmit the excessive power back to the CP for sale. In practice, the renewable energy generation is unequal due to efficiency of different types of renewable energy harvesting devices and various RRHs locations. Let \( E_n, B_n^{[\text{ahead}]}, B_n^{[\text{real}]}, S_n \) be defined, respectively, as the amount of renewable energy generated at the \( n \)-th RRH, the amount of energy that has already been purchased from the grid in the day-ahead market, the amount of energy that is necessary to be purchased from the real-time market and the amount of excessive energy sold back to the grid via the retailer. Furthermore, let \( P_n^{[\text{IT}]} \) and \( P_n^{[\text{circuit}]} \) indicate the total transmit power at the \( n \)-th RRH and the hardware circuit power consumption at the \( n \)-th RRH, respectively. Then, the total energy consumption at the \( n \)-th RRH, i.e., \( P_n^{[\text{total}]} \), is upper-bounded by the total available energy at the \( n \)-th RRH, i.e.,

\[
P_n^{[\text{total}]} = P_n^{[\text{IT}]} + P_n^{[\text{circuit}]} \leq E_n + B_n^{[\text{ahead}]} + B_n^{[\text{real}]} - S_n.
\]

In practice, the price of generating per unit renewable energy, denoted by \( \pi_{\text{[new]}} \), is much cheaper than the price of purchasing per unit energy from the day-ahead market, represented by \( \pi_{\text{[ahead]}} \). From the supply and demand perspective, we assume that the retailer purchases additional energy from the real-time market at a higher price \( \pi_{\text{[real]}} \) and sells the unconsumed energy back to the grid at a reduced price \( \pi_{\text{[sell]}} \), i.e., \( \pi_{\text{[real]}} \geq \pi_{\text{[ahead]}} \geq \pi_{\text{[sell]}} \geq \pi_{\text{[new]}} \).

B. Downlink Transmission Model

We denote the overall beamforming vector from all the RRHs towards the \( i \)-th IT, \( i \in \mathcal{L}_c \), as \( w_i = [w_i^H, \ldots, w_i^H]^H \in C^{MN \times 1} \), where \( w_{ni} \in C^{M \times 1} \) is the beamformer from the \( n \)-th RRH towards the \( i \)-th IT. Let \( v_e = [v_e^H, \ldots, v_e^H]^H \in C^{MN \times 1} \) represent the overall beamforming vector from all the RRHs to the \( e \)-th active ET and \( h_{ni} \in C^{M \times 1} \) indicate the channel vector between the \( n \)-th RRH and the \( i \)-th IT. We denote the overall channel vector between all the RRHs and the \( i \)-th IT by \( h_i = [h_i^1, \ldots, h_i^N]^H \in C^{MN \times 1} \). The received signals at the \( i \)-th IT, \( i \in \mathcal{L}_c \), is then given by

\[
y_i = h_i^H w_i s_i^{[\text{IT}]} + \sum_{j \in \mathcal{L}_b, j \neq i} h_i^H w_j s_j^{[\text{IT}]} + \sum_{e \in \mathcal{L}_c} h_i^H v_e s_e^{[\text{ET}]} + n_i,
\]

where the terms at the right hand side of (2), respectively, represent the information-carrying signal intended to the \( i \)-th IT, the inter-user interference caused by all other non-desired information beams, the interference caused by the energy-carrying signals intended to all active ETs and the additive white Gaussian noise with variance of \( \sigma_i^2 \) at the \( i \)-th IT. Since no information is carried by the energy-carrying signals, they can be any arbitrary random signals generated at RRHs. We assume that the transmitted symbols, i.e., \( s_i^{[\text{IT}]} \), \( s_j^{[\text{IT}]} \) and \( s_e^{[\text{ET}]} \), are independent and identically distributed and their transmission energy is normalized to one. Without loss of generality, we also let the noise variances \( \sigma_i^2 \) be identical at all receiving terminals. Then, the signal-to-interference-plus-noise ratio (SINR) at the \( i \)-th IT, \( i \in \mathcal{L}_c \), is formulated as

\[
\text{SINR}_{i}^{[\text{IT}]} = \frac{|h_i^H w_i|^2}{\sum_{j \in \mathcal{L}_b, j \neq i} |h_i^H w_j|^2 + \sum_{e \in \mathcal{L}_c} |h_i^H v_e|^2 + \sigma_i^2}.
\]

The backhaul capacity consumption for the \( n \)-th RRH can be expressed as \( C_n^{[\text{backhaul}]} = \sum_{i \in \mathcal{L}_c} ||w_{ni}||_0^2 \) \( R_i \) \( \forall n \in \mathcal{L}_b \), where \( R_i = \log_2(1 + \gamma_i) \) is the achievable data rate (bit/s/Hz) by the \( i \)-th IT and \( ||w_{ni}||_0^2 \) is an indicator function that illustrates the scheduling choices of the individual ITs, i.e., the \( i \)-th IT has as

\[
||w_{ni}||_0^2 = \begin{cases} 0, & \text{if } ||w_{ni}||_0^2 = 0, \\ 1, & \text{if } ||w_{ni}||_0^2 \neq 0. \end{cases}
\]

Here, \( ||w_{ni}||_0^2 = 0 \) indicates that the \( i \)-th IT is not served by/scheduled to the \( n \)-th RRH and, hence, the backhaul link between the CP and the \( n \)-th RRH is not used for joint data transmission to the \( i \)-th IT. The total energy harvested by the \( e \)-th active ET, \( e \in \mathcal{L}_c \), can be defined as

\[
\mathcal{G}_e^{[\text{ET}]} = \eta \left( |g_e^H v_e|^2 + \sum_{j \in \mathcal{L}_b, j \neq e} |g_e^H v_j|^2 + \sum_{i \in \mathcal{L}_c} |g_e^H w_i|^2 \right),
\]

where \( 0 \leq \eta \leq 1 \) indicates the conversion efficiency from the harvested RF energy to the electrical energy, \( g_e = \)
\([g^H_n, \ldots, g^{H}_{b}]^H \in \mathbb{C}^{MN \times 1}\) represents the overall channel vector between all the RRHs and the \(e\)-th active ET. To further improve the energy efficiency, we adopt the ET authorization algorithm proposed by [2] that can be implemented in the CP. The CP will authorize the \(n\)-th RRH, \(n \in \mathcal{L}_n\) to transmit an amount of energy towards the \(e\)-th active ET, \(e \in \mathcal{L}_e\) which is located within its \(n\)-th hexagonal energy serving area, i.e., \(\|v_{le}\|_{2}^2 = 1\) and \(\|v_{me}\|_{2}^2 = 0\), \(m \in \mathcal{L}_b, m \neq n\). Note that only one RRH is serving the \(e\)-th active ET and all the beamformers from the other RRHs to the \(e\)-th ET are set to zero. On the contrary, the \(z\)-th ET which is located outside of any hexagonal energy serving area will be set as an idle ET, \(z \in \mathcal{L}_e^{[idle]}\) and all the beamformers towards the idle ET are set to be zero, \(\|v_{ze}\|_{2}^2 = 0, \forall e \in \mathcal{L}_b\). The total amount of energy that can be harvested from surroundings by the \(z\)-th idle ET, \(z \in \mathcal{L}_e^{[idle]}\), is given by

\[
\eta_z = \sum_{i \in \mathcal{L}_i} |f_z^H w_i|^2 + \sum_{e \in \mathcal{L}_e} |f_z^H v_e|^2,
\]

where \(f_z = [f^H_1, \ldots, f^H_N]^H \in \mathbb{C}^{MN \times 1}\) denotes the overall channel vector between all the RRHs and the \(z\)-th idle ET.

### III. REAL-TIME ENERGY TRADING IN C-RAN
#### A. Problem Formulation

We introduce a joint resource management and energy trading formulation to strike an optimum balance amongst backhaul capacity restrictions, total transmit power and total energy cost as

\[
\min_{w_{ni}, v_{ne}, B_{n}^{[real]}, B_{n}^{[ideal]}} \mathcal{P}^{[coop]} + \sum_{n \in \mathcal{L}_b} P_{n}^{[Tx]} + \sum_{n \in \mathcal{L}_b} \left\{ B_{n}^{[real]} \right\}
\]

s.t.

C1: \(\text{SINR}^{[IT]} \geq \gamma_i, \forall i \in \mathcal{L}_i\),

C2: \(\sum_{e \in \mathcal{L}_e} g_{e}^{[ET]} \geq \gamma_{e}^{[min]}, \forall e \in \mathcal{L}_e\),

C3: \(\sum_{e \in \mathcal{L}_e} g_{e}^{[ET-idle]} \geq \gamma_{e}^{[id]} \forall e \in \mathcal{L}_e^{[idle]}\),

C4: \(P_{n}^{[Tx]} + P_{n}^{[circuit]} \leq E_n + B_{n}^{[ahead]} + B_{n}^{[ideal]} - S_n, \forall n \in \mathcal{L}_b\),

C5: \(P_{n}^{[Tx]} \leq P_{n}^{[max]} \forall n \in \mathcal{L}_b\),

C6: \(B_{n}^{[ahead]} \leq C_{n}^{[b-limit]} \forall n \in \mathcal{L}_b\),

C7: \(\sum_{n \in \mathcal{L}_b} B_{n}^{[ahead]} + \sum_{n \in \mathcal{L}_b} P_{n}^{[ideal]} \leq P_{CP}^{[max]} - P_{CP}^{[circuit]}\),

C8: \(B_{n}^{[real]} \geq 0, \forall n \in \mathcal{L}_b\),

where \(\mathcal{P}^{[coop]} = (\sum_{i \in \mathcal{L}_i} \|w_{i}\|_{2}^2 + \cdots + \sum_{i \in \mathcal{L}_i} \|w_{Ni}\|_{2}^2) + (\sum_{e \in \mathcal{L}_e} \|v_{le}\|_{2}^2 + \cdots + \sum_{e \in \mathcal{L}_e} \|v_{Ne}\|_{2}^2)\) and \(P_{n}^{[Tx]} = \sum_{i \in \mathcal{L}_i} \|w_{ni}\|_{2}^2 + \sum_{e \in \mathcal{L}_e} \|v_{ne}\|_{2}^2 \) \(n \in \mathcal{L}_b\). C1 denotes the minimum SINR requirements \(\gamma_i\) for the \(i\)-th ITs. \(P_{n}^{[min]}\) in C2 represents the minimum energy transmission requirements by the active ETs while \(P_{z}^{[ideal]}\) in C3 is the requirements of minimum energy harvested from the surroundings by the idle ETs. C4 indicates that the total transmit power of each RRH is constrained by its power budget and the non-transmission power consumption as per (1). C5 denotes that the total transmit power should not exceed the maximum transmit power allowance \(P_{n}^{[max]}\) at the \(n\)-th RRH. C6 denotes the backhaul link capacity restrictions for the individual RRHs. C7 specifies the constraint for the total power supplied by the CP to the RRHs, where \(P_{CP}^{[circuit]}\) and \(P_{CP}^{[max]}\) are the hardware circuit power consumption and the maximum power provision at the CP, respectively.

#### B. Reweighted \(\ell_1\)-norm and semidefinite programming (SDP)

We overcome the problem of intractability of the \(\ell_0\)-norm term in the objective function and constraints C6 by approximating them as sum weighted power and \(C_{n}^{[b-limit]} \approx \sum_{i \in \mathcal{L}_i} \|\xi_{ni}\|_{w_{ni}}^2 \|_{2} R_{i} = \sum_{i \in \mathcal{L}_i} \xi_{ni} \|w_{ni}\|_{2} R_{i} \), respectively, via the reweighted \(\ell_1\)-norm method [7]. Defining \(H_e = h_{1}h_{e}^H, G_e = g_{e}g_{e}^H, F_{e} = f_{e}f_{e}^H\) and the rank-one semidefinite matrix \(W_{i} = w_{i}w_{i}^H, V_{e} = v_{e}v_{e}^H\), and relaxing the rank-one constraints via SDR approach, the problem in (7) can be reformulated as

\[
\min_{w_{n}, v_{e}, B_{n}^{[real]}, B_{n}^{[ideal]}} \sum_{n \in \mathcal{L}_b} \left( \sum_{i \in \mathcal{L}_i} \xi_{ni} \text{tr}(W_{i}D_{n}) + \sum_{e \in \mathcal{L}_e} \kappa_{ne} \text{tr}(V_{e}D_{n}) \right) + \left( \sum_{i \in \mathcal{L}_i} \text{tr}(W_{i}) + \sum_{e \in \mathcal{L}_e} \text{tr}(V_{e}) \right) + \sum_{n \in \mathcal{L}_b} \left\{ B_{n}^{[real]} \right\}
\]

s.t.

C1: \(\text{tr}(H_{i}W_{i}) \geq \gamma_{i} \sum_{j \in \mathcal{L}_i, j \neq i} \text{tr}(H_{i}W_{j}) + \gamma_{i} \sigma_{i}^2, \forall i \in \mathcal{L}_i\),

C2: \(\text{tr}(G_{e}V_{e}) + \sum_{j \in \mathcal{L}_e, j \neq e} \text{tr}(G_{e}V_{j}) + \sum_{i \in \mathcal{L}_i} \text{tr}(G_{e}W_{i}) \geq F_{e}^{[min]} \eta^{-1}, \forall e \in \mathcal{L}_e\),

C3: \(\sum_{i \in \mathcal{L}_i} \text{tr}(F_{i}W_{i}) + \sum_{e \in \mathcal{L}_e} \text{tr}(F_{e}V_{e}) \geq \rho_{i}^{[ideal]} \eta^{-1}, \forall i \in \mathcal{L}_i\),

C4: \(\sum_{i \in \mathcal{L}_i} \text{tr}(W_{i}D_{n}) + \sum_{e \in \mathcal{L}_e} \text{tr}(V_{e}D_{n}) \leq [E_{n} - S_{n} + B_{n}^{[ahead]} + B_{n}^{[ideal]} - P_{CP}^{[circuit]}], \forall n \in \mathcal{L}_b\),

C5: \(\sum_{i \in \mathcal{L}_i} \text{tr}(W_{i}D_{n}) + \sum_{e \in \mathcal{L}_e} \text{tr}(V_{e}D_{n}) \leq P_{n}^{[max]}, \forall n \in \mathcal{L}_b\),

C6: \(\sum_{i \in \mathcal{L}_i} \xi_{ni} \text{tr}(W_{i}D_{n})R_{i} \leq C_{n}^{[b-limit]} \forall n \in \mathcal{L}_b\),

C7 - C9: \(\forall n \in \mathcal{L}_b\),

C10: \(W_{i} \succeq 0, \forall i \in \mathcal{L}_i, \quad C11: V_{e} \succeq 0, \forall e \in \mathcal{L}_e\),

where \(D_{n} \geq 0\) is a block diagonal matrix, given by

\[
D_{n} = \\
\begin{pmatrix}
(n-1)M & M & (N-n)M \\
M & (N-n)M \\
(M \cdots 0, 0, \cdots 0) \\
\end{pmatrix}, \forall n \in \mathcal{L}_b.
\]

Algorithm 1 summarizes the steps of sparse beamforming design for a given set of energy package purchased from day-ahead market, where the problem in (8) is repeatedly solved using the iteratively updated weight factor \(\xi_{ni}\) and \(\kappa_{ne}\). The cooperative links between the RRHs and the active receiving terminals are iteratively removed on the basis of the power budgets and backhaul link capacity restrictions at the individual RRHs. Consequently, the RRHs with low transmit power in the
$k$-th iteration result in high weight factors, which will further force the transmit power to be reduced in the $(k+1)$-th iteration until the solution sparsity is attained.

Algorithm 1 Reweighted $\ell_1$-norm method.

1: Initialize: constant $\mu \to 0$, iteration count $k = 0$, weight factor $\xi_n(0) = 1$, $\kappa_{ne}(0) = 1$, maximum number of iterations $k_{max}$.
2: While $\xi_n$ and $\kappa_{ne}$ are not converged or $t \neq k_{max}$ do
3: Find the optimal $\mathbf{W}_n(k)$ and $\mathbf{V}_n(k)$ by solving (8);
4: Update the weight factor $\xi_n(k+1)$ as follows,
$$\xi_n(k+1) = \frac{1}{B_n(k)D_n+\mu}, \forall n \in L_b, i \in L_i$;  
5: Update the weight factor $\kappa_{ne}(k+1)$ as follows,
$$\kappa_{ne}(k+1) = \frac{1}{B_n(k)D_n+\mu}, \forall n \in L_b, e \in L_e$;  
6: Increment the iteration number $k = k + 1$;
7: End while

IV. COMBINATORIAL MULTI-ARMED BANDIT FOR REAL-TIME ENERGY TRADING IN C-RAN

The multi-armed bandit (MAB) problem models a slot machine attempts to maximize the accumulated reward by iteratively optimizing the decisions among a set of arms based on existing knowledge, known as exploitation, while simultaneously acquiring new knowledge by observing the associated reward, known as exploration [12].

In this paper, we employ an abstract idea of MAB for the CP to learn from the behaviour of the energy trading in C-RAN iteratively, where each arm corresponds to the size of an energy package per RRH to be purchased by the retailer from the day-ahead market prior to the actual time of energy demand. Therefore, the new responsibility of the CP is to find the set of optimal sizes of the energy packages to be purchased from the day-ahead market without the initial knowledge of forthcoming instantaneous energy demand, in order to minimize the total energy cost of the retailer. We assume that there is a total instantaneous energy demand, in order to minimize the total day-ahead market without the initial knowledge of forthcoming instantaneous energy demand, until the solution sparsity is attained.

In this section, we introduce a forward CMAB (ForCMAB) Energy Trading algorithm to find the optimal combination of arms in an ascending order of packages sizes in a given number of trials $T$, where $T = t_{exp} + t_{max}$ is the total of exploration trials, $t_{exp}$ and exploitation trials, $t_{max}$. The steps are summarized in Algorithm 2, where we assume a set of index of the sizes of energy packages offered by the grid during the day-ahead market $E^\text{total}_n = \{E^1, \ldots, E^J\}$ with $E^1 < E^2 < \ldots < E^J$ is an arithmetic progression (AP) with common difference of $C$. Furthermore, we define package size $x = xC$, where $x$ is a constant number that has been chosen by the CP.

Algorithm 2 ForCMAB Energy trading

1: Initialize: $A_0^{[set]} = \{0_1, \cdots, 0_N\}$, exploration count $t = 0$, exploitation count $u = 0$, maximum number of trials $T$.
2: Run Algorithm 1;
3: CP calculates $B_n^{[total]}(0)$ and defines the initial individual reward $R(B_n^{[ahead]}(0)) = 0$, accumulated reward $R(A_0^{[set]}) = 0$ and total accumulated reward over $T$
$$\sum_{x=1}^{T} R(A_x^{[set]}) = 0, \forall n \in L_b$.
4: While $R(A_{t-1}^{[set]}) \geq R(A_t^{[set]})$ do
5: Increment the iteration number $t = t + 1$.
6: Exploration of new combinatorial arms, $A_t^{[set]}$ by solving (8);
7: CP calculates $B_n^{[total]}(t)$, $R(B_n^{[ahead]}(t))$ and $R(A_t^{[set]})$, \forall $n \in L_b$.
8: If the individual reward for the $n$-th RRH
$$R(B_n^{[ahead]}(t)) \geq R(B_n^{[ahead]}(t-1)) \text{ and } B_n^{[ahead]}(t) \neq E^J, \forall n \in L_b$.
9: Then update $B_n^{[ahead]}(t+1) = B_n^{[ahead]}(t)+\text{package size}$,
10: Else update $B_n^{[ahead]}(t+1) = B_n^{[ahead]}(t), \forall n \in L_b$.
11: End if
12: Update $A_{t+1}^{[set]} = \{B_1^{[ahead]}(t+1), \cdots, B_N^{[ahead]}(t+1)\}$.
13: End while
14: Set $A_t^{[set]} = A_{t+1}^{[set]} = \{B_1^{[ahead]}(t+1), \cdots, B_N^{[ahead]}(t+1)\}$.
15: Update $t_{exp} = t$.
16: While $u \neq T$ do
17: Increment the iteration number $u = u + 1$.
18: Exploration of the arms with the optimal reward $A_t^{[set]} = A_t^{[set]}$, run Algorithm 1 and solve (8).
19: CP calculates $B_n^{[total]}(u)$, $R(B_n^{[ahead]}(u))$ and $R(A_t^{[set]})$.
20: End while
21: Calculate total accumulated reward over $T$;
$$\sum_{x=1}^{T} R(A_x^{[set]}) = \sum_{t=1}^{t_{exp}} R(A_t^{[set]}) + \sum_{u=t_{exp}+1}^{T} R(A_u^{[set]})$$

Note that in Algorithm 2, the CP starts the operation with $A_0^{[set]} = \{0_1, \cdots, 0_N\}$ to check if the available local renewable energy is sufficient to satisfy the QoS of the ETs and ITs. If satisfied, the CP will decide not to buy any packages from
the day-ahead market. At each trial \( t \), the CP decides to purchase \( N \) combination of sizes of the energy packages from the day-ahead market for \( N \) RRHs based on the individual reward obtained from the current trial \( t \) and the previous trial \((t-1)\). The iterations of CMAB search are continued until the aggregated reward is maximized. In contrast to the strategy used in the former algorithm, we introduce a reverse CMAB (RevCMAB) Energy Trading algorithm, i.e., Algorithm 3 to find the optimal combination of the energy packages in descending order of packages sizes to be purchased from the day-ahead market, so that the maximum individual RRH reward can be achieved, and thus, the total energy cost of the retailer is minimized.

**Algorithm 3 RevCMAB Energy trading**

1: Initialize: a set of maximum sizes of energy packages offered by the day-ahead market \( \mathcal{A}_{\text{set}} = \{ \mathcal{E}_1, \ldots, \mathcal{E}_N \} \).

2: STEP 2 - STEP 7 in Algorithm 2

3: If the individual reward for the \( n \)-th RRH

\[ R(B_{\text{head}}^{\text{ahead}}(t)) \geq R(B_{\text{head}}^{\text{ahead}}(t-1)) \] and

\[ B_{\text{head}}^{\text{ahead}}(t) \neq 0, \quad \forall n \in \mathcal{L}_b \] then update \( B_{\text{head}}^{\text{ahead}}(t+1) = B_{\text{head}}^{\text{ahead}}(t) \) - package\[\text{size} \],

else update \( B_{\text{head}}^{\text{ahead}}(t+1) = B_{\text{head}}^{\text{ahead}}(t), \quad \forall n \in \mathcal{L}_b \).

4: end if

5: STEP 12 - STEP 21 in Algorithm 2

**V. SIMULATION RESULTS**

![Multi-user Downlink SWIPT C-RAN Simulation Topology](image)

We consider a downlink C-RAN consists of 3 adjacent RRHs SWIPT towards 30 single-antenna ITs and 6 single-antenna ETs. Each RRH is equipped with 8 antennas and located 500m away from each other, as shown in Fig 1. The renewable energy generation at each RRH is given in Fig 2. The total energy cost of the retailer is minimized.

![Total energy cost versus number of trials at \( \gamma = 20 \text{dB} \)](image)
As it can be observed from the figure, the learning speed of the ForCMAB Energy Trading algorithm is much higher than RevCMAB Energy Trading algorithm for package size = 200 mW. The comparison of the total energy cost of the retailer versus various SINR targets for different strategies at the 24-th trial is illustrated in Fig. 3. It can be observed from the figure that after 24 times of trials, both of our proposed ForCMAB Energy Trading and RevCMAB Energy Trading algorithms outperform the strategy proposed in [2] and [11] in terms of total energy cost of the retailer. Furthermore, the performance gap increases with the increasing SINR requirements. Fig. 4 presents in details a set of the optimal energy packages chosen by the CP to be purchased from the day-ahead market at the 24-th trial. It is noticeable that instead of using a fixed set of energy packages, both of our proposed strategies purchase a set of optimal energy packages from day-ahead market on the basis of actual energy generation and the energy requirements at the individual RRHs. In addition, even though both of the proposed strategies have similar performance in terms of total energy cost of the retailer at $\gamma = 20 \text{ dB}$, the proposed RevCMAB Energy Trading algorithm tends to purchase higher amount of energy packages from the day-ahead market for the individual RRHs.

VI. CONCLUSION

This paper proposes two CMAB algorithms, namely, ForCMAB Energy Trading and RevCMAB Energy Trading, to observe the instantaneous energy demand and learn from the behaviour of cooperative energy trading in the green C-RAN with SWIPT. At each trial, the sparse beamforming technique is employed to find the optimal trade-off between the cooperative transmission and the total energy cost of the retailer. Assuming that the RRHs have no initial knowledge of forthcoming energy consumption and renewable energy production in real-time energy trading, we employed the CMAB learning process to search for the optimal set of energy packages to be purchased from the day-ahead market in a limited number of trials, to further reduce the total energy cost of the retailer. Our simulation results confirm that in terms of reducing the total energy cost of the retailer, both proposed strategies outperform two recently proposed strategies without CMAB approach in green C-RAN.

REFERENCES


