Zhang, L., Feng, G., Qin, S., Sun, Y. and Cao, B. (2022) Access control for ambient backscatter enhanced wireless internet of things. *IEEE Transactions on Wireless Communications*. (Early Online Publication)

This is the Author Accepted Manuscript.

http://eprints.gla.ac.uk/265742/

Deposited on: 21 February 2022

Access Control for Ambient Backscatter Enhanced Wireless Internet of Things

Long Zhang, Gang Feng, *Senior Member, IEEE*, Shuang Qin, *Member, IEEE*,
Yao Sun, *Senior Member, IEEE*, and Bin Cao, *Senior Member, IEEE*

Abstract—**Beyond fifth-generation (B5G) and future networks face the challenges of spectral, energy and cost efficiency for large-scale machine-type communications. Recently, emerging ambient backscatter communication (AmBC) technology provides a promising paradigm for the development of green Internet of Things (IoT) networks in B5G era. Unlike existing work on AmBC which mostly focuses on physical layer with relatively ideal model, i.e., classic three-nodes model composed of radio frequency (RF), backscatter device (BD) and IoT device, this paper studies the access control strategy, including coefficient design and device association, from the perspective of networking. Assuming whether channel information is available a-priori, we propose online and offline access control strategies respectively. For offline access control strategy, we leverage the difference of two convex functions approximation (DCA) and dual decomposition to transform the non-concave optimization problem into the concave one, and design a distributed access control strategy called DCA-S. Furthermore, for the case that channel information is assumed to be unknown in advance due to the dynamics of primary and backscatter networks, we design a combinatorial multi-armed bandit (CMAB) access control strategy (CMAB-S). Numerical results show that the proposed DCA-S and CMAB-S can achieve significant performance improvement of the system in both cases of available and unavailable channel information compared with benchmark schemes.**

Index Terms—**Ambient Backscatter Network, Internet of Things, Coefficient Design, Device Association.**

I. INTRODUCTION

The vision of beyond fifth-generation (B5G) and future 6G wireless networks is expected to achieve ubiquitous connectivity, with aim of supporting much higher spectral/energy/cost efficiency, higher data rate (up to 1Tbps), lower end-to-end latency ($<$ 1ms), higher connection density ($\geq 10^7$ devices/km^2), etc. [1]. In this context, a new forecast is that there will be $50 \sim 75.4$ billion connected devices by 2025

and 79.4 zettabytes of data [2], driven by the tremendous evolution of internet of things (IoT) and various content-based services. However, current IoT is still at its infancy stage, due to the extremely severe challenges brought by large-scale deployment and wide-range commercialization of IoT systems. In particular, the connectivity of tens of billions of IoT devices poses extremely high stress on spectrum, energy and cost efficiency, which hinders the effective support of massive IoT applications. Besides, the limited battery capacity of IoT devices is another challenge to be tackled. In order to support the large expansion of connected IoT devices for B5G/6G era, it is highly desirable to improve spectrum utilization and power efficiency for wireless communication systems.

Backscatter communication is envisioned as a promising technology for low-power wireless communication systems. As a well-known application, radio frequency identification (RFID) system uses passive and low-cost devices for communications, where a dedicated carrier generator such as a reader is required to generate radio frequency (RF) signals to power the tags. Without using dedicated spectrum resources, ambient backscatter communication (AmBC) [3] was proposed to enable wireless devices to communicate using ambient RF signals as the only source of power, such as using orthogonal-frequency-division-multiplexing (OFDM) signals [4]. AmBC is expected to effectively solve spectral, energy and cost efficiency problems for large-scale machine-type communications, so that the communications between IoT devices can be realized almost everywhere and anytime. Fig. 1 shows the three-nodes backscatter communication model including backscatter device (BD), IoT device and ambient RF source. By harvesting energy from primary transmissions such as WiFi and cellular systems, BDs reflect the surrounding ambient signal by adjusting RF impedance switch under the backscatter controller, so that data is backscattered to an IoT device.

In recent years, most related work [5]–[10] focuses on the fundamentals of backscatter communications in physical layer, while considering relatively ideal single-RF based topology such as three-nodes model. However, the limitation of single-RF source could not be sufficient to support high throughput and wide coverage for massive IoT devices, especially considering the dynamics of direct and backscatter signals. Consequently, multi-RF AmBC can be exploited to improve the usability of ambient RF signals and enable flexible network deployment by using appropriate ambient RF sources in practical scenarios. Instead of focusing on the performance assurance of single backscatter network, an access control scheme for both primary and backscatter communications is
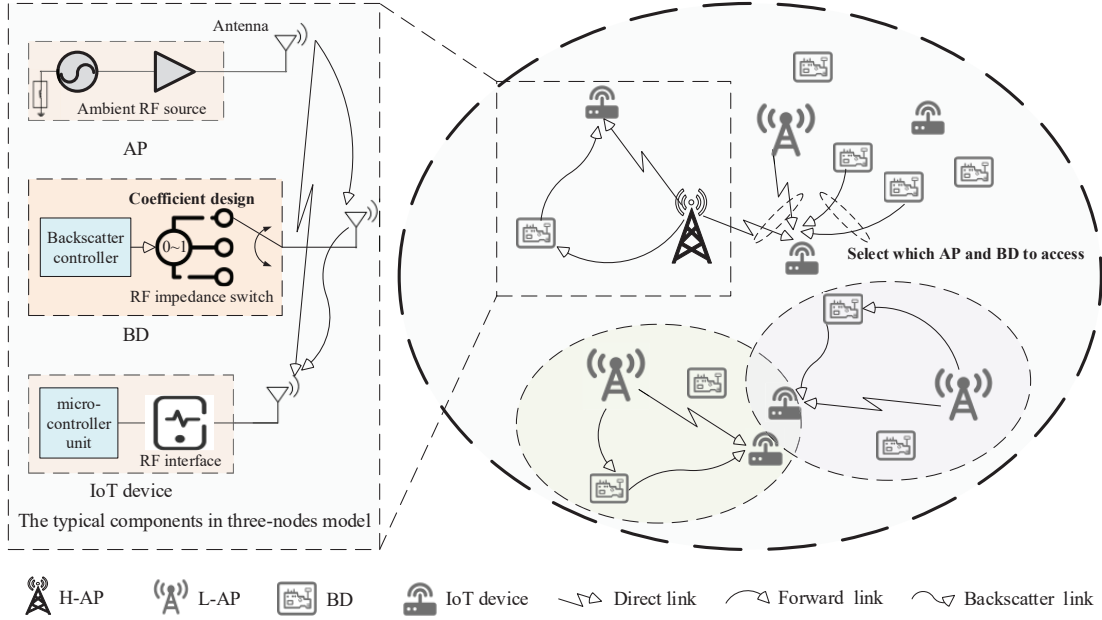
Fig. 1: Example of the access control strategy in multi-RF backscatter networks.

essential to increase the transmission efficiency and reduce mutual interference of the system.

Due to the different characteristics of ambient RF sources and devices, the access control scheme needs to effectively tackle the problems of BD selection, user association, RF admissibility, coefficient design, which are the main design challenges. As shown in Fig. 1, an AmBC underlaying primary networks may have multiple access points (APs) with different capabilities, e.g., high power-AP (H-AP) and low power-AP (L-AP). As a result, the access control scheme is needed to decide which AP and BD to associate with to establish primary and backscatter transmissions, so as to achieve satisfactory overall network performance. Meanwhile, the backscatter controller adjusts the coefficient of individual BDs according to RF impedance switch. On the one hand, the associated AP affects both the QoSs of primary and backscatter transmissions due to spectrum multiplexing. On the other hand, appropriate coefficient design and device association can reduce mutual interference. In this way, the achievable transmission capacity of the primary and backscatter systems can be maximized by properly exploiting the RF source and BD.

Inspired by these observations, this paper focuses on the access control strategy in multi-RF backscatter networks from the perspective of networking. The main contributions in this work are summarized as follows.

- First, in order to characterize the throughput maximization problem, we establish the primary and backscatter transmission models based on a time-slotted manner. We formulate the access control strategy problem with the objective of maximizing the weighted throughput sum of primary and backscatter networks under QoS constraints. To the best of our knowledge, this is the first work that studies coefficient design and device association in multi-RF scenario of AmBC.
- Next, we consider an offline optimization approach to

solve the access control problem when the estimate of channel state information (CSI) is assumed to be available a-priori based on pilot signals. In order to convert the original problem into a mathematically tractable form, we first exploit the special structure of the optimization problem and slack the decision variables reasonably. Then we use the classic difference of two convex functions approximation (DCA) method to transform the original non-concave problem into the concave one. In order to reduce the computational complexity, the original problem is further decomposed into several subproblems with lower dimensions by using dual decomposition theory, and a distributed DCA-based access control strategy (DCA-S) is proposed.

- In addition, assuming that the priori CSI is unavailable in advance, due to the dynamics of both the primary and backscatter networks, we propose an online access control strategy based on learning optimization. With the long-term optimization objective, the access control strategy is formulated as a combinatorial multi-armed bandit (CMAB) problem, and a modified upper confidence bound (UCB) algorithm called CMAB-based access control strategy (CMAB-S) is designed while taking into account one-shot offline solution.
- Finally, we provide numerical results to validate the performance of the proposed DCA-S and CMAB-S. Compared with benchmarks, DCA-S and CMAB-S can achieve significant improvement in average throughput and the number of admissible IoT devices for both cases of available and unavailable priori channel information.

The rest of this paper is organized as follows. In Section II, we illustrate the system model and problem formulation for AmBC enabled IoT networks. Section III presents the offline DCA-S access control strategy. Then we propose a CMAB-

S access control strategy based on learning optimization in section IV. In Section V, we evaluate the performance and discuss the results. Section VI reviews the related work, and finally, Section VII concludes the whole paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a backscatter communication system consisting of an H-AP and K L-APs, in which L IoT devices and M BDs are randomly distributed within the maximum service distance (MSD) [5]. Within the MSD, each IoT device can communicate with the associated AP with different orthogonal frequency resources [6]. In addition, BD can transmit information to IoT devices by modulating its data bits over ambient signals. For example, IoT devices can select the optimal WiFi RF source based on the location, signal strength, etc., and cooperatively receive data from WiFi and wearable sensors for body area network applications. Let the set of APs be $\mathcal{K} = \{1, 2, ..., K\}$, the set of BDs be $\mathcal{M} = \{1, 2, ..., M\}$, the set of IoT devices be $\mathcal{L} = \{1, 2, ..., L\}$, and the set of antennas be $\mathcal{A} = \{1, 2, ..., A\}$. Each IoT device is equipped with A $(A \geq 1)$ antennas to improve backscatter signal gain by using antenna diversity [7]. Note that the high-throughput backscatter transmission can be achieved when the number of the received antennas satisfies $A > K$ [8]. For simplicity, flat-fading channel model is considered in this paper, where all channels remain constant during each time slot, but they can independently vary in different intervals [7] [11]. During each time slot, the direct link from AP k to the a-th antenna of IoT device l is denoted by $f_{k,l,a}$, the forward link from AP k to BD m is denoted by $h_{k,m}$, and the backscatter link from BD m to the a-th antenna of IoT device l is denoted by $g_{m,l,a}$, where $k \in \mathcal{K}$, $m \in \mathcal{M}$, $l \in \mathcal{L}$, and $a \in \mathcal{A}$.

For the co-existence of primary and backscatter communications, each IoT device receives backscattered signals in separate time slots [12], which is typical time division multiple access (TDMA) mode. Within a time slot, all BDs adjacent to the currently communicating IoT device can share the same time slot and spectrum for backscatter transmission. Despite the low transmission efficiency, TDMA can avoid interference among BDs and save bandwidth resources.

A. Primary and Backscatter Transmissions

In time slot n, the direct link signal received by IoT device l from AP k is $r_{k,l}^{\mathrm{d}}(n) = \sum_{a \in \mathcal{A}} f_{k,l,a} \sqrt{P_{k,l,a}^{\mathrm{d}}} s_{k,l}(n)$, where $s_{k,l}(n)$ denotes the signal transmitted from AP k to IoT device l. Suppose that $s_{k,l}(n)$ is the unit power signal and follows the standard circular symmetric complex gaussian (CSCG) distribution [5], i.e., $s_{k,l}(n) \sim \mathcal{CN}(0,1)$. $P_{k,l,a}^{\mathrm{d}} = \frac{P_k^{\mathrm{s}} G_k^{\mathrm{s}} G_{l,a}^{\mathrm{d}} \lambda^2}{(4\pi)^2 (L_{k,l}^{\mathrm{d}})^{v_1}}$ is the average received signal power from AP k to the a-th antenna of IoT device l, where P_k^{s} is the transmit power of AP k, G_k^{s} is the transmitter antenna gain, $G_{l,a}^{\mathrm{d}}$ is the receiver antenna gain, λ is the wavelength of the signal, v_1 is the path loss exponent, and $L_{k,l}^{\mathrm{d}}$ is the distance from BS k to IoT device l [12].

Because of the double fading effect, the signal of backscatter link is typically weaker than that of the direct link. In other words, the transmission rate of backscatter links is lower than that of direct links, which is suitable for low-rate transmission and cos-sensitive scenarios. Suppose that the symbol period T_c of BD covers W $(W \geq 1)$ symbol periods T_s of AP, i.e., $T_c = WT_s$. Therefore, during a primary transmission period, the backscatter signal received by IoT device l from BD m in time slot n can be expressed as $r_{m,l}^{\mathrm{c}}(n) = \alpha_{k,m,l} h_{k,m} \sum_{a \in \mathcal{A}} g_{m,l,a} \sqrt{P_{m,l,a}^{\mathrm{c}}} s_{k,l}(n) c_{m,l}$, where $n = 1, 2, ..., W$, $\alpha_{k,m,l}$ is the backscatter coefficient of BD m, $P_{m,l,a}^{\mathrm{c}} = \frac{P_{k,m}^{\mathrm{b}} G_m^{\mathrm{b}} G_{l,a}^{\mathrm{d}} \lambda^2}{(4\pi)^2 (L_{m,l}^{\mathrm{c}})^{v_3}}$ is the average received signal power from BD m to the a-th antenna of IoT device l, $P_{k,m}^{\mathrm{b}} = \frac{P_k^{\mathrm{s}} G_k^{\mathrm{s}} G_m^{\mathrm{b}} \lambda^2}{(4\pi)^2 (L_{k,m}^{\mathrm{b}})^{v_2}}$ is the average received signal power from AP k to the BD m, G_m^{b} is the antenna gain of BD m, $L_{k,m}^{\mathrm{b}}$ is the distance from AP k to BD m, $L_{m,l}^{\mathrm{c}}$ is the distance from BD m to IoT device l, v_2 and v_3 are the path loss exponents. Let $c_{m,l}$ be the message transmitted from BD m to IoT device l, and $c_{m,l} \in \{0, 1\}$. When $c_{m,l} = 1$, backscatter state is triggered and bit "1" is transmitted from BD m to IoT device l. Otherwise $c_{m,l} = 0$, there is no information for backscatter transmission.

In time slot n, the signal summation of the primary transmission $r_{k,l}^{\mathrm{d}}(n)$ and backscatter transmission $r_{m,l}^{\mathrm{c}}(n)$ at the a-th antenna can be expressed as

$$r_{l,a}(n) = \sqrt{P_{k,l,a}^{\mathrm{d}}} \left(f_{k,l,a} + \alpha_{k,m,l} h_{k,m} g_{m,l,a} \sqrt{\ell} c_{m,l} \right) s_{k,l}(n) + u_{l,a}(n), \tag{1}$$

where $n = 1, 2, ..., W$, $\ell = \frac{(G_m^{\mathrm{b}})^2 \lambda^2 (L_{k,l}^{\mathrm{d}})^{v_1}}{(4\pi)^2 (L_{k,m}^{\mathrm{b}})^{v_2} (L_{m,l}^{\mathrm{c}})^{v_3}}$, and $u_{l,a}(n)$ is the complex gaussian noise.

Based on the decoding strategy and successful interference cancellation (SIC) technology [7], the IoT device first decodes $s_{k,l}(n)$ by treating backscatter signals as interference. Meanwhile, the interference from other APs is not considered in the system, which is because the co-channel interference caused by channel reuse can be efficiently mitigated by operating adjacent APs in different frequencies. We denote the decoding matrix as $\Omega = [\Omega_1; \Omega_2]$, where $\Omega_1 = [\omega_{1,1}, \ldots, \omega_{1,A}]$ is the decoding vector of primary transmission and $\Omega_2 = [\omega_{2,1}, \ldots, \omega_{2,A}]$ is the decoding vector of backscatter transmission. Multiplying by different decoding matrices such as MRC, MMSE and ZF, the primary and backscatter signals can be decoded at the receiver separately, that is, $\Omega \times [r_{l,1}(n); \ldots; r_{l,A}(n)]$. Note that the detailed decoding strategy can be referred to [7].

For primary transmission, the signal to interference plus noise ratio (SINR) $\gamma_{k,m,l}^{\mathrm{d}}$ is given by $\gamma_{k,m,l}^{\mathrm{d}} = \frac{\sum_{a \in \mathcal{A}} \omega_{1,a} |f_{k,l,a}|^2 P_{k,l,a}^{\mathrm{d}}}{\alpha_{k,m,l} \frac{1}{W} \sum_{n=1}^{W} \sum_{a \in \mathcal{A}} \omega_{1,a} |h_{k,m}|^2 \ell |g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}} + \sigma_1^2}$, where $\sigma_1^2 = \sum_{a \in \mathcal{A}} \omega_{1,a} u_{l,a}(n)$. According to the definition in [8] [11] [13], the primary transmission throughput from BS k to IoT device l can be expressed as

$$R_{k,m,l}^{\mathrm{d}} = \log\left(1 + \gamma_{k,m,l}^{\mathrm{d}}\right). \tag{2}$$

After $s_{k,l}(n)$ is removed from $r_l(n)$, the SNR for decoding backscatter signal can be expressed as

$$\gamma_{k,m,l}^{\text{b}} = \frac{\sum_{n=1}^{W} \alpha_{k,m,l} \sum_{a \in \mathcal{A}} \omega_{2,a} |h_{k,m}|^2 \ell |g_{m,l,a}|^2 P_{k,l,a}^{\text{d}}}{\sigma_2^2}, \quad \text{where}$$

$\sigma_2^2 = \sum_{a \in \mathcal{A}} \omega_{2,a} u_{l,a}(n)$. Since the symbol period of T_c covers W successive symbol periods T_s of AP, the SINR used to decode BD symbol can be regarded as increasing W times, and the symbol rate reduces $1/W$ times [8] [13]. Hence, the achievable backscatter throughput from BD m to IoT device l using AP k as source can be expressed as

$$R_{k,m,l}^{\text{b}} = \frac{1}{W} \log\left(1 + \gamma_{k,m,l}^{\text{b}}\right). \tag{3}$$

B. Problem Formulation

Our design problem of this paper can be stated as: maximizing the weighted sum of primary and backscatter throughput subject to specific QoS constraints. Based on the primary and backscatter throughput in Eqs. (2) and (3), we can formulate the access control problem as

$$\max \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} u_{k,l} v_{m,l} \left[\rho \bar{R}_{k,m,l}^{\text{d}} + (1 - \rho) \bar{R}_{k,m,l}^{\text{b}} \right], \tag{4}$$

$$\text{s.t.} \quad C_1 : \sum_{k \in \mathcal{K}} u_{k,l} = 1, \ \forall l \in \mathcal{L}, \tag{4-1}$$

$$C_2 : \sum_{l \in \mathcal{L}} v_{m,l} = 1, \ \forall k \in \mathcal{K}, \tag{4-2}$$

$$C_3 : \alpha^{\min} \leq \alpha_{k,m,l} \leq \alpha^{\max}, \ \forall m \in \mathcal{M}, \tag{4-3}$$

$$C_4 : \sum_{k \in \mathcal{K}} \sum_{m \in \mathcal{M}} u_{k,l} v_{m,l} R_{k,m,l}^{\text{d}} \geq R_l^{\text{th}}, \ \forall l \in \mathcal{L}, \tag{4-4}$$

$$C_5 : u_{k,l}, v_{m,l} \in \{0,1\}, \ \forall (k,m,l) \in \mathcal{K} \times \mathcal{M} \times \mathcal{L}, \tag{4-5}$$

where ρ is the weight factor, $\rho \in [0,1]$. $\bar{R}_{k,m,l}^{\text{d}} = R_{k,m,l}^{\text{d}}/R_{\text{d}}^{nor}$ and $\bar{R}_{k,m,l}^{\text{b}} = R_{k,m,l}^{\text{b}}/R_{\text{b}}^{nor}$ are the normalized throughput in order to fill a gap of backscatter and primary throughput, i.e., R_{d}^{nor} and R_{b}^{nor} are constant. $u_{k,l}$ and $v_{m,l}$ are 0-1 decision variables used to pair AP k, BD m and IoT device l, where $u_{k,l} = 1$ and $v_{m,l} = 1$ if the l-th IoT device is associated with the k-th AP and m-th BD and $u_{k,l} = 0$ and $v_{m,l} = 0$ otherwise. Constraint (4-1) means that each IoT device can only access one AP in one time slot. Constraint (4-2) guarantees the backscatter transmission via IoT device l and BD m in one time slot. Constraint (4-3) guarantees the feasible regime of backscatter coefficient. Constraint (4-4) ensures that the QoS requirements of the primary transmission must meet a minimum threshold R_l^{th}. Constraint (4-5) states that the access variables are 0-1 integer.

III. ACCESS CONTROL STRATEGY BASED ON OFFLINE OPTIMIZATION

A. The Offline Access Control Strategy

We propose to apply alternative iteration method to jointly optimize coefficient design and device association in an offline manner. In details, our solution to the joint coefficient design and device association problems is described as follows.

- **Coefficient design**: Give a specific device association solution, the optimization of coefficient design is still mathematically intractable due to the non-concave structure of the problem. With this property, we first rewrite this non-concave optimization problem into the difference of two concave functions. After that, exploiting this special structure, the DCA method is used to transform the non-concave optimization problem into the concave one. Furthermore, the large-scale optimization problem can be decomposed into a master problem and several subproblems, and the optimal coefficient design can be solved using dual decomposition theory.
- **Device association**: In order to guarantee the QoS requirements of primary transmission, we should determine the set of accessible IoT devices and BDs that satisfies the minimum rate requirement R_l^{th}. For the fixed backscatter coefficient, we adopt relaxation and decoupling methods to convert device association subproblems into a mathematically tractable problem. As a result, the integer programming is transformed into the difference of two concave functions, and a DCA-heuristic method is developed to solve device association problem to obtain a suboptimal solution.

Based on the solutions to subproblems, the master problem updates dual variables for each subproblem. Finally, the iteration continues until a convergence condition is satisfied.

B. DCA for Coefficient Design Problem

We can observe that the objective function (4) and constraint (4-4) can be rewritten as the difference between two concave functions with respect to $f(\cdot)$ and $g(\cdot)$, which is equivalent to the original coefficient problem. Given fixed $u_{k,l}$ and $v_{m,l}$, the coefficient design problem can be rewrite as follows.

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & f(\boldsymbol{\alpha}) - g(\boldsymbol{\alpha}) \\ = & \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} \left(\underbrace{f_1(\alpha_{k,m,l}) + f_2(\alpha_{k,m,l})}_{f(\alpha_{k,m,l})} - g(\alpha_{k,m,l}) \right), \end{aligned} \tag{5}$$

$$\text{s.t.} \ C_4 : f_1(\boldsymbol{\alpha}) - g(\boldsymbol{\alpha}) =$$
$$\sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} \left(f_1(\alpha_{k,m,l}) - g(\alpha_{k,m,l}) \right) \geq R_l^{\text{th}}, \ \text{and} \ C_3, \tag{5-1}$$

where $\boldsymbol{\alpha} = \{\alpha_{k,m,l}, \forall (k,m,l) \in \mathcal{K} \times \mathcal{M} \times \mathcal{L}\}$ is the vector of backscatter coefficient, and $f_1(\alpha_{k,m,l})$, $f_2(\alpha_{k,m,l})$ and $g(\alpha_{k,m,l})$ are shown in Eq. (6).

We can find that the objective function in (5) is indeed the difference of two concave functions $f(\boldsymbol{\alpha})$ and $g(\boldsymbol{\alpha})$. However, the objective function $f(\boldsymbol{\alpha}) - g(\boldsymbol{\alpha})$ is non-concave. Therefore, it is intractable to apply conventional convex or quasi-concave algorithms to solve this problem. To optimize backscatter coefficient in (5), we exploit the special structure of the optimization objective, and propose to use DCA method to solve such a non-concave problem.

The main idea of DCA is as follows. For the difference of two concave functions, the minuend $g(\cdot)$ is approximated by

$$\begin{cases} f_1\left(\alpha_{k,m,l}\right) = w_{k,m,l}\rho\log\left(\alpha_{k,m,l}\frac{1}{W}\sum_{n=1}^{W}\sum_{a\in\mathcal{A}}\omega_{1,a}|h_{k,m}|^2\ell|g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}} + \sum_{a\in\mathcal{A}}\omega_{1,a}|f_{k,l,a}|^2 P_{k,l,a}^{\mathrm{d}} + \sigma_1^2\right), \\ f_2\left(\alpha_{k,m,l}\right) = w_{k,m,l}\left(1-\rho\right)\frac{1}{W}\left[\log\left(\alpha_{k,m,l}\sum_{n=1}^{W}\sum_{a\in\mathcal{A}}\omega_{2,a}|h_{k,m}|^2\ell|g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}} + \sigma^2\right) - \log\left(\sigma_2^2\right)\right], \\ g\left(\alpha_{k,m,l}\right) = w_{k,m,l}\rho\log\left(\alpha_{k,m,l}\frac{1}{W}\sum_{n=1}^{W}\sum_{a\in\mathcal{A}}\omega_{1,a}|h_{k,m}|^2\ell|g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}} + \sigma_1^2\right), \end{cases} \tag{6}$$

its first-order Taylor expansion around some feasible points, so as to solve the resultant concave function and obtain the global optimal solution. For clarity and conciseness, we first give the following definitions:

$$\begin{cases} Q_{k,l}^1 = \sum_{a\in\mathcal{A}}\omega_{1,a}|f_{k,l,a}|^2 P_{k,l,a}^{\mathrm{d}}, \\ Q_{k,m,l}^2 = \frac{1}{W}\sum_{n=1}^{W}\sum_{a\in\mathcal{A}}\omega_{1,a}|h_{k,m}|^2\ell|g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}}, \\ Q_{k,m,l}^3 = \sum_{n=1}^{W}\sum_{a\in\mathcal{A}}\omega_{2,a}|h_{k,m}|^2\ell|g_{m,l,a}|^2 P_{k,l,a}^{\mathrm{d}}. \end{cases} \tag{7}$$

Then, in the s-th iteration, coefficient design problem in Eq. (5) can be transformed into a DCA function around feasible point $\boldsymbol{\alpha}\left[s\right]$, which can be regarded as an upper bound of $g\left(\boldsymbol{\alpha}\left[s\right]\right)$ at $\boldsymbol{\alpha}\left[s\right]$, expressed as.

$$\max_{\boldsymbol{\alpha}} f\left(\boldsymbol{\alpha}\right) - g\left(\boldsymbol{\alpha}\left[s\right]\right) - \nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right) \tag{8}$$
$$\text{s.t. } \widetilde{C}_4 : f_1\left(\boldsymbol{\alpha}\right) - g\left(\boldsymbol{\alpha}\left[s\right]\right) - \nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right) \geq R_l^{\mathrm{th}},$$
$$C_3, \text{ and } C_5, \tag{8-1}$$

where ∇ is the gradient, and $\nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right)$ can be expressed by

$$\nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right) =$$
$$\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}}\left[\frac{u_{k,l}v_{m,l}\rho Q_{k,m,l}^2\left(\alpha_{k,m,l} - \alpha_{k,m,l}\left[s\right]\right)}{\left(\alpha_{k,m,l}\left[s\right]Q_{k,m,l}^2 + \sigma_1^2\right)\ln 2}\right].$$

We can observe that objective function in Eq. (8) is a concave function with respect to $\alpha_{k,m,l}$. Therefore, effective convex optimization methods can be adopted to solve the above optimization problem. In the following, we elaborate on how to solve the optimization problem in Eq. (8) using Lagrange duality and decomposition theory. First, we transfer constraint $\widetilde{C}_4$ into the objective function with Lagrange multipliers and define the Lagrange as follows.

$$L\left(\boldsymbol{\alpha}, \boldsymbol{\lambda}\right) = f\left(\boldsymbol{\alpha}\right) - g\left(\boldsymbol{\alpha}\left[s\right]\right) - \nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right)$$
$$+ \sum_{l=1}^{L}\lambda_l\left(f_1\left(\boldsymbol{\alpha}\right) - g_1\left(\boldsymbol{\alpha}\left[s\right]\right) - \nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}\left[s\right]\right) - R_l^{\mathrm{th}}\right), \tag{9}$$

where $\boldsymbol{\lambda} = \{\lambda_l, \forall l\in\mathcal{L}\}$ is the Lagrange multiplier associated with constraint $\widetilde{C}_4$. Then, the Lagrange dual function can be expressed as

$$\min_{\boldsymbol{\lambda}}\max_{\boldsymbol{\alpha}} L\left(\boldsymbol{\alpha}, \boldsymbol{\lambda}\right), \text{ s.t. } \lambda_l \geq 0, \ \forall l\in\mathcal{L}, \text{ and } C_3. \tag{10}$$

The Lagrange dual can be solved by alternately optimizing $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$. Obviously, such optimization method requires a centralized controller for all APs, BDs and IoT devices.

However, due to the requirements of low cost and low power consumption, limited resources and delay requirements make it difficult to support such centralized solution. Therefore, we resort to distributed algorithms instead of directly solving the Lagrange dual. Next, we exploit the decomposability structure in Eq. (9) and adopt dual decomposition theory to decompose the original large problem into distributively solvable master problem and subproblems.

Since $L\left(\boldsymbol{\alpha}, \boldsymbol{\lambda}\right)$ does not include cross-term of $\boldsymbol{\lambda}$, the Lagrange dual function in Eq. (10) can be equivalently expressed as the sum of subproblems $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$ (defined in Eq. (11)) by exchanging the computation order, as shown below.

$$\min_{\lambda_l}\max_{\alpha_{k,m,l}}\sum_{l\in\mathcal{L}}\left(L_l\left(\alpha_{k,m,l}, \lambda_l\right) + \lambda_l R_l^{\mathrm{th}}\right), \tag{12}$$
$$\text{s.t. } \lambda_l \geq 0, \ \forall l\in\mathcal{L}, \text{ and } C_3. \tag{12-1}$$

In this way, the optimization problem defined by Eq. (12) can be separated into two loops of optimization. In the inner loop, we have L subproblems, i.e., $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$. In the outer loop, the dual variable λ_l can be updated by minimizing the Lagrange dual function in Eq. (12).

*1) **The inner loop:*** Given fixed λ_l, subproblem $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$ can be expressed as

$$\max_{\alpha_{k,m,l}} L_l\left(\alpha_{k,m,l}, \lambda_l\right), \text{ s.t. } C_3. \tag{13}$$

Using dual decomposition, we can solve each subproblem $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$ separately and get the following solution $\alpha_{k,m,l}\left[s+1\right]$:

$$\alpha_{k,m,l}\left[s+1\right] = \arg\max_{C_3} L_l\left(\alpha_{k,m,l}, \lambda_l\right). \tag{14}$$

Due to the strict concavity of subproblem $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$, $\alpha_{k,m,l}\left[s+1\right]$ can be efficiently solved by available convex optimization solvers [14] [15].

*2) **The outer loop:*** Based on the obtained solutions to the subproblems, dual variable $\boldsymbol{\lambda}$ can be obtained by solving the following master dual problem:

$$\min_{\lambda_l} D\left(\lambda_l\right) = \sum_{l\in\mathcal{L}}\left(L_l^{max}\left(\alpha_{k,m,l}, \lambda_l\right) + \lambda_l R_l^{\mathrm{th}}\right), \tag{15}$$
$$\text{s.t. } \lambda_l \geq 0, \ \forall l\in\mathcal{L}, \tag{15-1}$$

where $L_l^{max}\left(\alpha_{k,m,l}, \lambda_l\right)$ is the maximum value of the Lagrangian solved in Eq. (13) for a given λ_l. Since subproblem $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$ is convex, the strong dual holds and the duality gap is zero. Therefore, the minimum value of the master problem is equal to the maximum value of the sum subproblems.

Leveraging subgradient method, dual variables λ_l can be updated for each subgradient of $L_l\left(\alpha_{k,m,l}, \lambda_l\right)$, as shown

$$L_l\left(\alpha_{k,m,l},\lambda_l\right)=\sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\Big[f\left(\alpha_{k,m,l}\right)-g\left(\alpha_{k,m,l}\left[s\right]\right)-\nabla g\left(\alpha_{k,m,l}\left[s\right]\right)\left(\alpha_{k,m,l}-\alpha_{k,m,l}\left[s\right]\right)\Big]$$
$$+\lambda_l\left(\sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\left[f_1\left(\alpha_{k,m,l}\right)-g\left(\alpha_{k,m,l}\left[s\right]\right)-\nabla g\left(\alpha_{k,m,l}\left[s\right]\right)\left(\alpha_{k,m,l}-\alpha_{k,m,l}\left[s\right]\right)\right]\right). \tag{11}$$

below.

$$\lambda_l\left[s+1\right]=\Big[\lambda_l\left[s\right]-\delta\left[s\right]\Big(f_1\left(\boldsymbol{\alpha}\right)-g\left(\boldsymbol{\alpha}\left[s\right]\right)-$$
$$\nabla g\left(\alpha_{k,m,l}\left[s\right]\right)\left(\alpha_{k,m,l}-\alpha_{k,m,l}\left[s\right]\right)-R_l^{\text{th}}\Big)\Big]^+, \tag{16}$$

where $\delta\left[s\right]>0$ is a sufficiently small positive step-size, and $\left[\cdot\right]^+$ guarantees that the dual variable $\lambda_l\left[s+1\right]$ is in a nonnegative regime.

In DCA function, $g\left(\boldsymbol{\alpha}\right)$ is upper bounded by its first-order Taylor expansion around feasible solution $\boldsymbol{\alpha}\left[s\right]$. Hence, DCA function provides a well approximated lower bound minimization for the original problem, i.e., $h\left(\boldsymbol{\alpha}\right)=f\left(\boldsymbol{\alpha}\right)-g\left(\boldsymbol{\alpha}\right)\geq f\left(\boldsymbol{\alpha}\right)-g\left(\boldsymbol{\alpha}\left[s\right]\right)-\nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha}-\boldsymbol{\alpha}\left[s\right]\right)=\widetilde{h}\left(\boldsymbol{\alpha}\right)$. Since constraints are compact, the following condition is satisfied: $h\left(\boldsymbol{\alpha}\left[s+1\right]\right)\geq f\left(\boldsymbol{\alpha}\left[s\right]\right)-g\left(\boldsymbol{\alpha}\left[s\right]\right)-\nabla g\left(\boldsymbol{\alpha}\left[s\right]\right)\left(\boldsymbol{\alpha}\left[s+1\right]-\boldsymbol{\alpha}\left[s\right]\right)\geq h\left(\boldsymbol{\alpha}\left[s\right]\right)$. According to Cauchy theorem, sequence $\boldsymbol{\alpha}\left(s\right)$ always converges with the increase of s [14].

For coefficient design problem, we can obtain a KKT optimal solution. According to [16], at least a KKT solution can be obtained if the objective function is differentiable and can be written as the DCA structure $f\left(\boldsymbol{\alpha}\right)-g\left(\boldsymbol{\alpha}\right)$ under a convex feasible set. Furthermore, we can conclude that the global optimal solution can be obtained if $f\left(\boldsymbol{\alpha}\right)-g\left(\boldsymbol{\alpha}\right)$ does not exceed one peak (valley) point, which ensures that the DCA algorithm does not fall into a local optimal solution. For $u_{k,l}v_{m,l}=1$, $f\left(\alpha_{k,m,l}\right)-g\left(\alpha_{k,m,l}\right)$ can be equivalently expressed as: $R(\alpha_{k,m,l})=\rho\log\left(1+1/(\alpha_{k,m,l}Q^5_{k,m,l}+Q^6_{k,m,l})\right)+\left(1-\rho\right)/W\log\left(1+\alpha_{k,m,l}Q^4_{k,m,l}\right)$, where $Q^4_{k,m,l}=Q^3_{k,m,l}/\sigma^2_2$, $Q^5_{k,m,l}=Q^2_{k,m,l}/Q^1_{k,l}$, and $Q^6_{k,m,l}=\sigma^2_1/Q^1_{k,m,l}$. Obviously, we can observe that the first term $\log\left(1+1/(\alpha_{k,m,l}Q^5_{k,m,l}+Q^6_{k,m,l})\right)$ is a monotonically decreasing function, and the second term $\log\left(1+\alpha_{k,m,l}Q^4_{k,m,l}\right)$ is a monotonically increasing function. Therefore, the function $R(\alpha_{k,m,l})$ has at most one peak (valley) point, which is similar with the property of the function $\log\left(1+x\right)+\log\left(1+1/x\right)$.

C. DCA for Device Association Problem

Intuitively, device association problem can be formulated as a combinatorial integer programming involving 0-1 variables, and the complexity of straightforward brute-force search over the sets of APs $\mathcal{K}$, BDs $\mathcal{M}$ and IoT devices $\mathcal{L}$ is obviously infeasible in large-scale deployment. Inspired by the hidden DCA structure [17], our strategy is to adopt relaxation and

DCA methods to convert the original problem (4) into a mathematically tractable form. It is worth noting that this relation reduces the complexity of combinatorial integer programming. In this way, we can formulate device association problem as a DCA function and develop a DCA-based heuristic algorithm to solve the optimization problem efficiently.

By relaxing binary variables, the constraint C_5 can be written as its equivalent form:

$$\begin{cases}\widetilde{C}_5:u_{k,l},v_{m,l}\in\left[0,1\right],\forall\left(k,m,l\right)\in\mathcal{K}\times\mathcal{M}\times\mathcal{L},\\ C_6:\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}}\left(u_{k,l}-u^2_{k,l}\right)\leq0,\\ C_7:\sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}}\left(v_{m,l}-v^2_{m,l}\right)\leq0,\end{cases} \tag{17}$$

where C_6 and C_7 are auxiliary constraints, used to restrict the value of continuous variable in domain $\{0,1\}$ [17] [18].

Since $u_{k,l}\in\{0,1\}$ is equivalent to $\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}}\left(u_{k,l}-u^2_{k,l}\right)=0$, the binary set $u_{k,l}\in\{0,1\}$ is the difference of two convex sets $\left[0,1\right]$ and $C_6:\sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}}\left(u_{k,l}-u^2_{k,l}\right)\leq0$, i.e., $\{0,1\}=\left[0,1\right]/\{u_{k,l}\in C_6\}$. Similarly, the binary set $v_{m,l}\in\{0,1\}$ is the difference of two convex sets $\left[0,1\right]$ and $C_7:\sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}}\left(v_{m,l}-v^2_{m,l}\right)\leq0$ [17].

In other words, the constraints $\widetilde{C}_5$, C_6 and C_7 satisfy the DCA function. Furthermore, the cross-term $u_{k,l}v_{m,l}$ can be equivalently replaced by $u_{k,l}v_{m,l}=\frac{1}{2}\left(u_{k,l}+v_{m,l}\right)^2-\frac{1}{2}\left(u^2_{k,l}+v^2_{m,l}\right)$. Given fixed $\alpha_{k,m,l}$ and substituting $u_{k,l}v_{m,l}=\frac{1}{2}\left(u_{k,l}+v_{m,l}\right)^2-\frac{1}{2}\left(u^2_{k,l}+v^2_{m,l}\right)$ into optimization problem in Eq. (4), device association problem can be expressed as the difference between two convex functions with respect to $u_{k,l}$ and $v_{m,l}$. Meanwhile, we can observe that device association problem becomes separable due to the absence of coupling constraints. Therefore, device association problem can be decomposed into L equivalent subproblems to be solved separately. For the l-th subproblem, we have

$$\max_{u_{k,l},v_{m,l}}\quad\sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\frac{1}{2}\left(u_{k,l}+v_{m,l}\right)^2R\left(\alpha_{k,m,l}\left[s\right]\right)-$$
$$\sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\frac{1}{2}\left(u^2_{k,l}+v^2_{m,l}\right)R\left(\alpha_{k,m,l}\left[s\right]\right), \tag{18}$$
$$\text{s.t. }\widetilde{C}_5,\ C_6,\ \text{and }C_7, \tag{18-1}$$

where $R\left(\alpha_{k,m,l}\left[s\right]\right)=\rho\bar{R}^{\text{d}}_{k,m,l}+\left(1-\rho\right)\bar{R}^{\text{b}}_{k,m,l}$. To handle the nonconvex constraints C_6 and C_7, we can integrate the constraints into the optimization objective by associating two Lagrange multipliers μ_1 and μ_2. Hence, the Lagrange dual

function is given by

$$\min_{\mu_1,\mu_2} \max_{u_{k,l},v_{m,l}} L\left(u_{k,l},v_{m,l},\mu_1,\mu_2\right) =$$

$$P_l\left(u_{k,l},v_{m,l}\right) - Q_l\left(u_{k,l},v_{m,l}\right), \quad (19)$$

$$\text{s.t. } \widetilde{C}_5, \ \mu_1 \text{ and } \mu_2 \geq 0, \quad (19\text{-}1)$$

where

$$P_l\left(u_{k,l},v_{m,l}\right) = \sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\frac{1}{2}\left(u_{k,l}+v_{m,l}\right)^2 R\left(\alpha_{k,m,l}\left[s\right]\right)-$$

$$\mu_1\left(\sum_{k\in\mathcal{K}}u_{k,l}+\left(\sum_{k\in\mathcal{K}}u_{k,l}\right)^2\right)-$$

$$\mu_2\left(\sum_{m\in\mathcal{M}}v_{m,l}+\left(\sum_{m\in\mathcal{M}}v_{m,l}\right)^2\right), \quad (20)$$

$$Q_l\left(u_{k,l},v_{m,l}\right) = \sum_{m\in\mathcal{M}}\sum_{k\in\mathcal{K}}\frac{1}{2}\left(u_{k,l}^2+v_{m,l}^2\right)R\left(\alpha_{k,m,l}\left[s\right]\right)-$$

$$\mu_1\left(\sum_{k\in\mathcal{K}}(u_{k,l})^2+\left(\sum_{k\in\mathcal{K}}u_{k,l}\right)^2\right)-$$

$$\mu_2\left(\sum_{m\in\mathcal{M}}(v_{m,l})^2+\left(\sum_{m\in\mathcal{M}}v_{m,l}\right)^2\right). \quad (21)$$

Since the strong Lagrangian duality holds for Eq. (19) with appropriately chosen $0 < \mu_0 < +\infty$ [17], the optimization problem in Eq. (18) is equivalent to the following convex optimization problem for $\mu_1, \mu_2 \geq \mu_0$:

$$\min_{u_{k,l},v_{m,l}} L\left(u_{k,l},v_{m,l},\mu_1,\mu_2\right), \text{ s.t. } \widetilde{C}_5. \quad (22)$$

Although both terms $P\left(u_{k,l},v_{m,l}\right)$ and $Q\left(u_{k,l},v_{m,l}\right)$ in $L\left(u_{k,l},v_{m,l},\mu_1,\mu_2\right)$ are convex over $u_{k,l}$ and $v_{m,l}$, the difference of two convex functions is not necessarily convex. In order to approximate optimization objective $L\left(u_{k,l},v_{m,l},\mu_1,\mu_2\right)$ to a convex function, we employ DCA algorithm by constructing a convex function using the first order Taylor approximation of minuend $Q\left(u_{k,l},v_{m,l}\right)$. In the s-th iteration, Eq. (18) can be transformed into a DCA function around feasible points $u_{k,l}\left[s\right]$ and $v_{m,l}\left[s\right]$, which is an upper bound of device association problem, expressed as.

$$\max_{u_{k,l},v_{m,l}} P_l\left(u_{k,l},v_{m,l}\right) - Q_l\left(u_{k,l}\left[s\right],v_{m,l}\left[s\right]\right)-$$

$$\nabla_{u_{k,l}}Q_l\left(u_{k,l}\left[s\right],v_{m,l}\left[s\right]\right)\left(u_{k,l}-u_{k,l}\left[s\right]\right)-$$

$$\nabla_{v_{m,l}}Q_l\left(u_{k,l}\left[s\right],v_{m,l}\left[s\right]\right)\left(v_{m,l}-v_{m,l}\left[s\right]\right), \quad (23)$$

$$\text{s.t. } \widetilde{C}_5. \quad (23\text{-}1)$$

Solving the above subproblem, we can obtain

$$\left(u_{k,l}\left[s+1\right],v_{m,l}\left[s+1\right]\right) = \arg\max_{u_{k,l},v_{m,l}} (23). \quad (24)$$

D. DCA-S Algorithm for Access Control Strategy

So far, we have discussed how to optimize coefficient design and device association problems in a distributed way

---

Algorithm 1: DCA-S algorithm to solve access control strategy

Input : Network topology (K, L, M) and channel state $(h_{k,m}, g_{m,l} \text{ and } f_{k,l,a})$

Output: $\alpha_{k,m,l}$, $u_{k,l}$ and $v_{m,l}$

1: Initialize:
   Set $s = 0$, $\lambda_l\left[s\right] = 0$, choose $\alpha_{k,m,l}\left[s\right]$ from the feasible set $\widetilde{C}_3 \cap \widetilde{C}_5$, choose $u_{k,l}\left[s\right]$ and $v_{m,l}\left[s\right]$ from $\widetilde{C}_6$, set μ_1 and μ_2 to a sufficiently large value

2: Compute $R_l\left[s\right]$ according to Eq. (4)

3: **repeat**

4: **for** $l = 1: L$ **do**

5:    The inner loop: obtain backscatter coefficient $\alpha_{k,m,l}\left[s+1\right]$ according to Eq. (14)

6:    The outer loop: update dual variable $\lambda_l\left[s+1\right]$ according to Eq. (16)

7:    Obtain $u_{k,l}\left[s+1\right]$ and $v_{m,l}\left[s+1\right]$ according to Eq. (24)

8:    Compute $gap = R_l\left[s+1\right] - R_l\left[s\right]$

9: **end for**

10: **until** $gap \leq threshold$

11: **return** $\alpha_{k,m,l}\left[s+1\right]$, $u_{k,l}\left[s+1\right]$ and $v_{m,l}\left[s+1\right]$

---

respectively. On this basis, we present an effective algorithm for DCA-based access control strategy, called DCA-S. In the s-th iteration, the access control strategy can be formulated by solving coefficient design subproblem in Eq. (14) and device association subproblem in Eq. (24). **Algorithm 1** summarizes DCA algorithm to solve coefficient design and device association problems described in this section.

The computational complexity of DCA-S algorithm is $O\left(MKF_\alpha F_{u,v}\right)$, where F_α and $F_{u,v}$ are the number of iterations for updating $\alpha_{k,m,l}$ and $(u_{k,l},v_{m,l})$, respectively. From simulation results, we will see that DCA-S algorithm has a relatively fast convergence speed.

IV. ACCESS CONTROL STRATEGY BASED ON ONLINE OPTIMIZATION

In Section III, we have proposed an offline DCA-S algorithm for solving the access control problem with known CSI. However, DCA-S algorithm requires real-time CSI on IoT devices, which in turn increases signaling overhead in a time slot-based backscatter network. Furthermore, due to the dynamics of wireless channels (e.g., channel fading, double-fading effect, etc.), channel information is unknown in advance and it is infeasible for each IoT device to obtain global information for all APs and BDs. As a result, the access control strategy represented by the proposed DCA-S algorithm is an instantaneous decision and cannot fully reflect the impact of network dynamics on decision-making, which may not always be the best policy for time-varying environments.

To optimize network performance in an online manner, we next propose a multi-armed bandit (MAB) framework for designing an access control strategy, while making use of coefficient design using DCA-S algorithm. The main idea of

MAB is to model a scenario in which an agent makes the optimal decision from a set of actions over time in rounds. For the decision made in each round, the agent will obtain a stochastic reward from the environment. Without knowing underlaying reward distribution, the optimal policy of MAB is needed to balance the exploitation of currently optimal policy while conducting an exploration of other policies that may turn out to be superior in the long run. By interacting with the environment, the decision-maker will gradually learn statistical knowledge about the reward distribution and make the optimal decisions to maximize the cumulative reward instead of immediate reward, so as to achieve long-term satisfactory system performance in dynamic environments.

A. CMAB-based Access Control Strategy (CMAB-S)

In MAB framework, each decision-maker sequentially selects an arm to pull from a set of predefined arms, and feeds back a corresponding reward from an unknown distribution. To formulate the MAB-based solution for the access control problem, each IoT device can be regarded as a decision-maker, and each arm corresponds to a pairing of the BDs to APs. Assuming that the decision-makers have no prior knowledge of network state information, once an arm is pulled in each round, the throughput in Eq. (4) can be regarded as a reward.

Different from the standard MAB mode, the arm in our problem is always a combinatorial action with two-dimension space. To construct the combinatorial action, we represent the joint arm selected in the t-th round as $a_l^b(t) = (u_{k,l}(t), v_{m,l}(t))$, where b is the index of the Cartesian product set, e.g., $b \in \mathcal{B} = \mathcal{K} \times \mathcal{M} = \{(k,m) \mid k \in \mathcal{K}, m \in \mathcal{M}\}$. The joint arm $a_l^b(t)$ of all IoT devices can constitute a strategy set $\Pi = \{a_1^b(t), a_2^b(t), \cdots, a_L^b(t)\}$. To find the optimal strategy Π^*, we formulate the problem as a combinatorial MAB (CMAB) with aim of maximizing the cumulative reward in finite T rounds. As a result, the optimal CMAB-S is formulated as

$$\Pi^* = \arg\max_{\Pi} \sum_{t \in T} \sum_{l \in \mathcal{L}} R_l\left(a_l^b(t)\right), \text{s.t. } C_1, \ C_2, \ C_4, \text{ and } C_5,$$
(25)

where $R_l\left(a_l^b(t)\right) = \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} u_{k,l}(t) v_{m,l}(t) [\rho R_{k,m,l}^{\mathrm{d}} + (1-\rho) R_{k,m,l}^{\mathrm{b}}]$.

In the problem (25), we can observe that the reward $R_l\left(a_l^b(t)\right)$ of IoT device l is independent of other IoT devices. As a result, the strategy made by decision-maker l does not affect the behavior of other decision-makers. Therefore, the independent decision-makers can make the optimal strategy in a decentralized way, such that the overall optimization can be accomplished. The optimal CMAB-S Π^* in problem (25) can be obtained in a decentralized way as follows.

$$a_l^{b^*} = \arg\max_{\Pi} = \sum_{t \in T} R_l\left(a_l^b(t)\right)$$
(26)

A straightforward method for solving a CMAB problem of independent decision-makers is to apply upper confidence bound (UCB) policy [19]. In UCB policy, the optimal joint arm $a_l^b(t)$ can be selected for maximizing the average reward

associated with a confident interval. Specifically, in each round, when joint arm $a_l^b(t)$ is pulled, the corresponding average reward and the number of times that the joint arm $a_l^b(t)$ has been pulled so far can be observed and stored. In turn, the optimal strategy can be updated based on the index of UCB policy, which is the average reward plus a confident interval to balance exploitation and exploration.

Since UCB policy is related to historical data, two tables with $K \times M$ dimensions are needed to store the average reward and the number of times for pulling each arm $a_l^b(t)$ [20]. Here we define the average reward $\breve{\varphi}_l(t)_{K \times M}$ and the number of times $n_l^b(t)_{K \times M}$ that the joint arm $a_l^b(t)$ has been pulled during t rounds. First, in the $(t+1)$-th round, if the joint arm $a_l^b(t+1)$ is pulled, the number of times $n_l^b(t+1)$ can be updated by $n_l^b(t+1) = n_l^b(t) + 1$.

Once the joint arm $a_l^b(t+1)$ is pulled in the $(t+1)$-th round, the weighted sum throughput can be observed as an immediate reward. However, selecting both actions k and m simultaneously may violate the QoS constraints. In order to avoid this problem, we define a penalized average reward to punish pulling an arm that violates QoS constraints in each round. In the $(t+1)$-th round, the average reward $\breve{\varphi}_l^b(t+1)$ from joint arm a_l^b can be expressed as

$$\breve{\varphi}_l^b(t+1) = \begin{cases} \frac{n_l^b(t)\breve{\varphi}_l^b(t) + R_l\left(a_l^b(t+1)\right)}{n_l^b(t+1)}, & \text{if } R_{k,m,l}^{\mathrm{d}} \geq R_l^{\mathrm{th}}, \\ \frac{n_l^b(t)\breve{\varphi}_l^b(t)}{n_l^b(t+1)}, & \text{otherwise.} \end{cases}$$
(27)

By maximizing the index of UCB policy, the optimal joint action is given by

$$a_l^b(t+1) = \arg\max_{a_l^b} \left(\breve{\varphi}_l^b(t) + \theta\sqrt{\frac{3\ln T}{2n_l^b(t)}}\right),$$
(28)

where $\theta = \max_{1 \leq t' \leq t} \breve{\varphi}_l^b(t')$, $\sqrt{\frac{3\ln T}{2n_l^b(t)}}$ is the exploration term to encourage the exploration of less selected arms, and the constant defining confidence interval is $3/2$ [19].

B. Analysis of Expected Regret for CMAB-S

Due to the existence of the exploration component, the strategy selected by decision-makers may not always be the best. To evaluate the quality of the selected action in each round, the conception of expected regret is introduced to measure the performance loss between the expected cumulative reward of the optimal action and that of the selected action, expressed as.

$$\text{Regret} = \varphi_l^* t - \sum_{b \in \mathcal{B}} \mathbb{E}\left[n_l^b(t)\right] \breve{\varphi}_l^b(t),$$
(29)

where $\varphi_l^* = \max_{a_l^b} \breve{\varphi}_l^b(t)$ is the best expected reward, and $\mathbb{E}[n_{k,m}(t)]$ is the expected number of plays for arms k and m in the first t rounds.

Next, the expected regret bound of UCB can be proved based on **Lemma 1** and **Corollary 1**.

Lemma 1: If UCB policy is run on a finite number of arms, the expected regret is uniformly logarithmic over all rounds, for all reward distributions with bounded support in $[0, 1]$.

Proof: cf. [21] for proofs.

Algorithm 2: CMAB-S algorithm for AmBC access control

Input : Network topology (K, L, M) and $\alpha_{k,m,l}$
Output: $u_{k,l}$ and $v_{m,l}$

1: Initialize: set $t = 0$
2: Pull each joint arm $a_l^b(t)$ once, $\forall b \in \mathcal{B}$
3: Set the number of times $n_l^b(t) = 1$ and compute average reward $\check{\varphi}_l^b(t)$ using Eq. (27)
4: **for** $t = 1 : T$ **do**
5:    Select the joint action $a_l^b(t+1)$ according to UCB index in Eq. (28)
6:    Update $n_l^b(t)$ and $\check{\varphi}_l^b(t)$ using Eq. (27)
7: **end for**
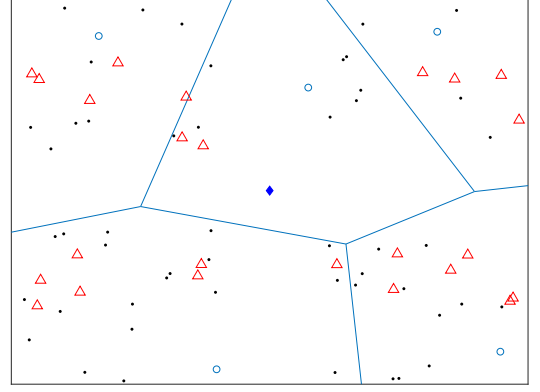8: **return** $a_l^{b*} = \max_{b \in \mathcal{B}} \check{\varphi}_l^b(t+1)$

Fig. 2: A backscatter network with two-tier cell network, where blue diamond is H-AP, light-blue circulars are L-APs, red triangles are BDs, and black dots are IoT devices.

Corollary 1: For the proposed CMAB-S, the expected regret of UCB policy is bounded as: Regret $\leq \sqrt{KMt\left(6\log(t) + \frac{\pi^2}{3} + 1\right)}$.

Proof: To meet the condition of **Lemma 1**, we can rewrite Eq. (28) as

$$a_l^b(t+1) =$$

$$\arg\max_{a_l^b} \left(\frac{\check{\varphi}_l^b(t)}{\theta} + \sqrt{\frac{3\ln T}{2n_l^b(t)}} - \min\left\{0, R_{k,m,l}^d - R_l^{th}\right\}^2 \right).$$

In Eq. (30), θ is a constant value to normalize $\hat{\mu}$ in $[0, 1]$. As the normalized reward is in $[0, 1]$, according to **Fact 1**, Eq. (30) can achieve the logarithmic regret bound over rounds when using UCB policy. Furthermore, since θ is a constant value, Eq. (30) is equivalent to Eq. (28), thereby yielding the same solution. Therefore, the expected regret is bounded by uniformly logarithmic over rounds for the proposed CMAB-S.

Furthermore, the expected regret can be rewritten as

$$\text{Regret} = \varphi_l^* t - \sum_{b \in \mathcal{B}} \mathbb{E}\left[n_l^b(t)\right]\check{\varphi}_l^b(t)$$

$$= \mathbb{E}\left[\sum_{b \in \mathcal{B}} n_l^b(t)\left(\underbrace{\varphi_l^* - \check{\varphi}_l^b(t)}_{\triangle_b}\right)\right]$$

$$= \sum_{b \in \mathcal{B}} \sqrt{\mathbb{E}n_l^b(t)}\sqrt{\mathbb{E}n_l^b(t)} \triangle_b. \qquad (30)$$

Using Cauchy-Schwarz inequality, we have Regret $\leq \sqrt{\sum_{b \in \mathcal{B}} \mathbb{E}n_l^b(t) \sum_{b \in \mathcal{B}} \mathbb{E}n_l^b(t)\triangle_b^2}$.

According to [21] and [22], we can get $\sum_{b \in \mathcal{B}} \mathbb{E}n_l^b(t) = t$ and $\mathbb{E}\left[n_l^b(t)\right] \leq 6\frac{\log(t)}{\triangle_b^2} + \frac{\pi^2}{3} + 1$. Substituting both of them into Regret, we can obtain Regret $\leq \sqrt{KMt\left(6\log(t) + \frac{\pi^2}{3} + 1\right)}$.

Based on above analysis, CMAB-S algorithm for solving AmBC access control problem can be summarized in **Algorithm 2**.

V. PERFORMANCE EVALUATION

In this section, we conduct three sets of experiments to compare the performance of DCA-S and CMAB-S with three baseline schemes as follows.

- Exhaustive searching based algorithm (ESA): The brute-force method is used to find the optimal access solution via exhaustive searching over KML possible solutions, where coefficient design problem can be solved by conventional solvers, e.g., YALMIP optimization tool [23].
- Max-Difference based algorithm (MDA): To the best of our knowledge, there is no explicit solution to address access control strategy with the QoS constraints. Inspired by [9], one solution is to select suitable AP and BD with the largest difference in terms of the channel gain between direct and backscatter links.
- Max-Weighted SINR based algorithm (MWSA): Inspired by [24], a modified scheme is designed for performance evaluation of our proposed DCA-S and CMAB-S based on a conventional Max-SINR scheme. Specifically, under the QoS constraints and given solved coefficient solution, the MWSA scheme is to select the suitable AP and BD with the largest weighted SINR sum of the direct and backscatter links.

A. Simulation Parameter Settings

In this paper, we consider a backscatter network parasitizing a two-tier cellular network. In the two-dimensional Euclidean plane, an H-AP is located in the center of an area of 100 square meters, and the locations of L-APs, BDs and IoT devices are modeled by three homogeneous Poisson Point Processes (PPPs) with average values Φ_K, Φ_M and Φ_L [6]. According to [9] and [11], BDs and IoT devices are located at a distance of 10-100 meters outside from AP, and the transmit power of H-AP and S-AP is set to 40 dBm and 20 dBm respectively. For clarity, we construct the network topology as a poisson-voronoi tessellation graph [25] with $\Phi_K = 5$, $\Phi_M = 5\Phi_K$, and $\Phi_L = 10\Phi_K$, as shown in Fig. 2.

In addition, we set the noise power $\sigma^2 = -101$ dBm, $G_m^b = G_{l,a}^d = G_k^s = 1.5$ dB, $v_1 = 3$, $v_2 = v_3 = 2$, $F = \frac{\lambda^2}{(4\pi)^2} = -40$ dB, and $W = 50$ [9] [11]. The feasible regime of backscatter coefficient is set as $[0.2, 1]$. For simplicity, all the channel information of direct and backscatter links are assumed to follow flat-fading, and can be generated
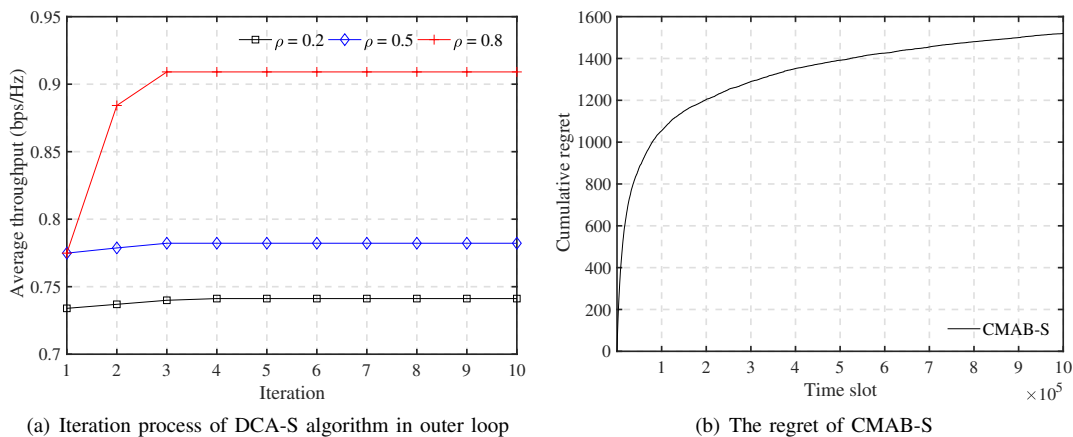
(a) Iteration process of DCA-S algorithm in outer loop

(b) The regret of CMAB-S

Fig. 3: Iteration and Learning Process

based on CSGN., i.e., $f_{k,l,a}$, $g_{m,l,a}$ and $h_{k,m} \sim \mathcal{CN}(0,1)$ [10].

B. Iteration and Learning Process

This experiment shows the iteration process of DCA-S. For DCA-S scheme, the convergence threshold is set to 0.0001 and the initial feasible solution of backscatter coefficient is set to 0.2. In order to obtain DCA-S strategy, the DCA function of coefficient design problem in Eq. (12) and device association problem in Eq. (23) can be solved using optimization solver YALMIP in each iteration. As shown in Fig. 3(a), we can observe that DCA-S scheme can eventually converge a stable value in a fast speed (no more than 5 times). Note that the observed number of iterations is updated by the outer loop, and the iteration process for solving coefficient design and device association depends on the adopted optimization solver. Furthermore, with the increase of weight factor ρ, the convergence speed of DCA-S is accelerated. In particular, as the weight factor approaches 1, the maximum value of the DCA function can be reached closer to the lower bound of the constraints, i.e., the initial feasible solution 0.2. This is because the weight of the primary transmission is larger than that of the backscatter transmission, and the optimal solution can be obtained at the lower bound of the feasible regime.

Next, we illustrate the learning process of CMAB-S with respect to time slot, where the cumulative regret is used to evaluate the cumulative difference between the best expected reward and that by adopting the UCB policy. As shown in Fig. 3(b), we can observe that the cumulative regret increases sharply before 10^5 time slot. During this period, the IoT devices tend to explore the arm that may not be currently optimal, so as to learn the statistic characteristics of the reward distribution in a dynamic scenario. After that, the cumulative reward grows slowly with respect to time slot, which means that IoT devices tend to exploitation (choose the best arm with the highest reward distribution) rather than exploration.

C. Performance Comparisons

In the second experiment, we evaluate the performance of the offline algorithm DCA-S when the complete CSI is known a-priori.

First, we simulate the throughput performance of DCA-S, MWSA and MDA schemes under MRC, MMSE and ZF decoding strategies. For each IoT device, the received signals can be first multiplied by corresponding MRC, MMSE and ZF decoding matrices. Then, primary and backscatter signals are recovered based on SIC-based detectors, as described in [7]. In this experiment, QPSK and BPSK modulations are used to generate primary and backscatter signals. The primary transmission rate R_s is set to 1 kbps and the backscatter throughput is $R_c = \frac{1}{W}R_s$. For simplicity, each time slot includes a primary signal and W backscatter signals. Considering the effect of Rayleigh fading and double-fading, the actual throughput is calculated as the probability of successful decoding multiplied by the transmission rate in a statistical manner. In the following, we vary the direct link SNR[dB] from 0 dB to 20 dB, and backscatter link SNR is -10 dB lower than that of direct link [7].

Fig. 4(a) and Fig. 4(b) show the primary and backscatter throughput under classical three-node model. We can observe that both primary and backscatter throughput of MRC, MMSE and ZF increase with the increase of SNR. Meanwhile, both primary and backscatter throughput of MMSE are better than ZF and MRC, and can achieve suboptimal performance compared with the optimal maximum likelihood detection. Furthermore, Fig. 4(c) and Fig. 4(d) show the primary and backscatter throughput under multi-RF source scenario. We can observe that the primary throughput of the proposed DCA-S is better than MWSA and MDA schemes. However, the backscatter throughput of the DCA-S is not always better than MWSA and MDA schemes, except DCA-S under MMSE. This is because different decoding strategies lead to different access control, and backscatter signal is weaker -10 dB than primary signal, resulting in worse detection errors.

In the following, we consider a particulate time slot and MRC decoding strategy to maximize the weighted sum of primary and backscatter throughput. Fig. 5 illustrates the impact of QoS threshold on the average throughput and the ratio of admissible IoT devices, where the weight factor is set to 0.5. As shown in Fig. 5(a), we can observe that the proposed DCA-S can achieve higher performance gains compared with the benchmark schemes. Meanwhile, the proposed DCA-S
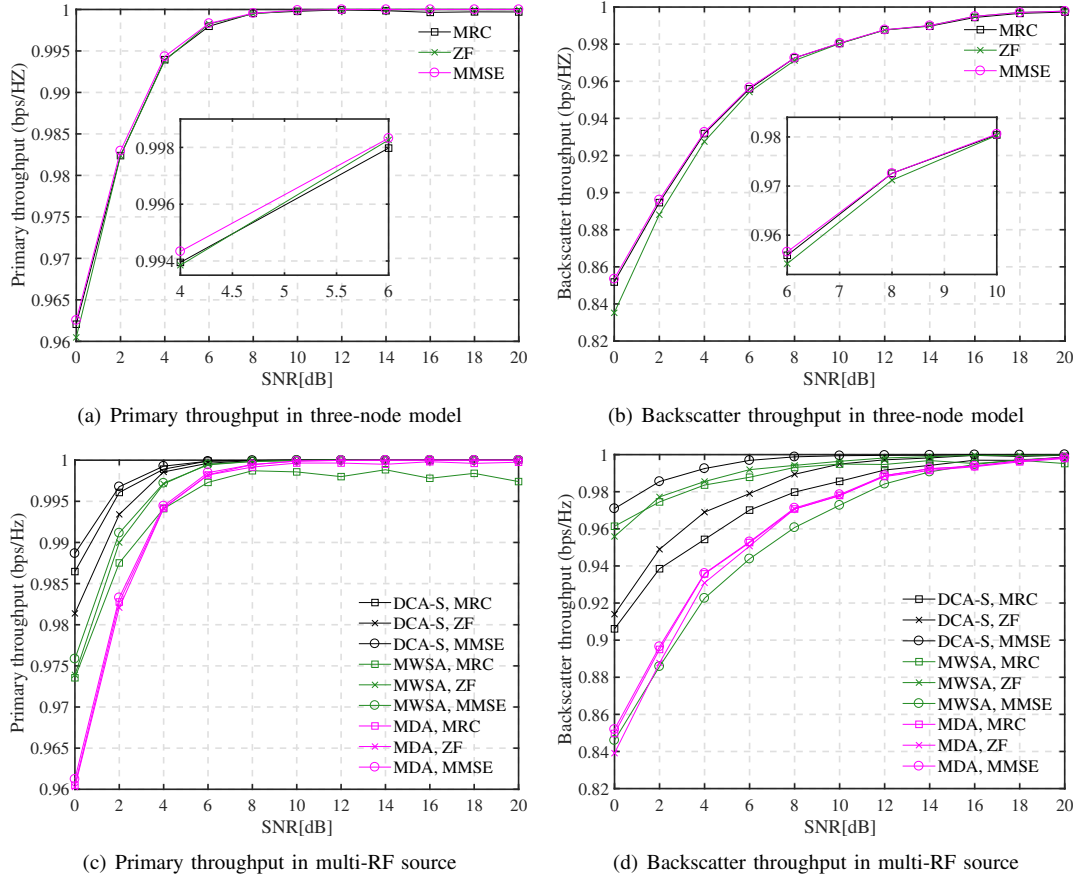
(a) Primary throughput in three-node model

(b) Backscatter throughput in three-node model

(c) Primary throughput in multi-RF source

(d) Backscatter throughput in multi-RF source

Fig. 4: Throughput performance comparisons under MRC, MMSE and ZF.

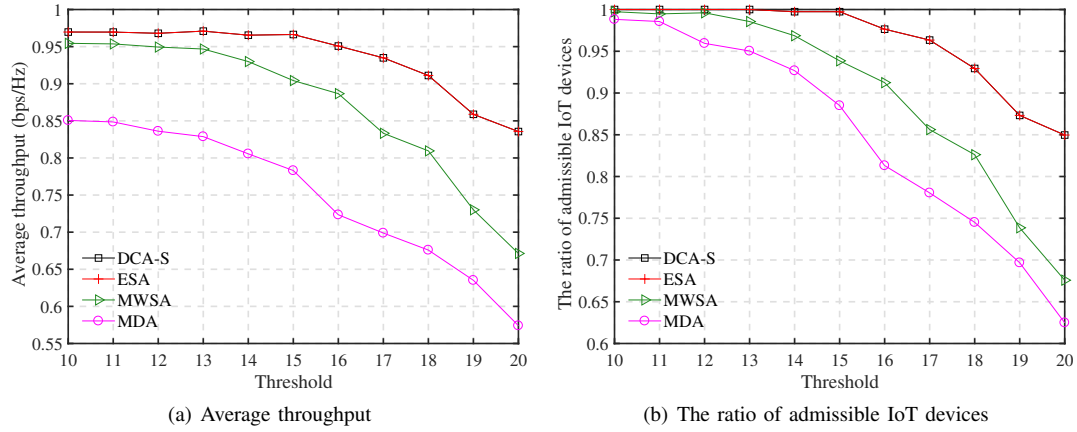(a) Average throughput

(b) The ratio of admissible IoT devices

Fig. 5: Performance comparisons vs. QoS threshold when the channel information is known.

can achieve almost the same performance as ESA. As the QoS threshold increases, the average throughput of the four schemes decrease due to the higher possibility of violating QoS constraints. Meanwhile, Fig. 5(b) shows that the ratio of admissible IoT devices of DCA-S is larger than that of MDA and MWSA, thus resulting in higher performance gain.

Fig. 6 shows the performance of DCA-S with weighted factor varying from 0.1 to 0.9, where the QoS constraint is set to 15. Obviously, the average throughput and the ratio of admissible IoT devices of the proposed DCA-S are always higher than that of other schemes, especially for small

weight factor (smaller than 0.6). Note that MWSA scheme can approach the optimal performance of ESA and DCA-S schemes when the weight factor increases from 0.6 to 0.9. This is because MWSA scheme can be approximately regarded as a Max-SINR scheme with the increase of weight factor, thus resulting in near-optimal solution.

In the third experiment, considering the dynamics of CSI, we evaluate the performance of CMAB-S when the channel information is unknown a-priori. Since each IoT device and BD only communicate within their corresponding time slots, IoT devices can only observe the channel information from
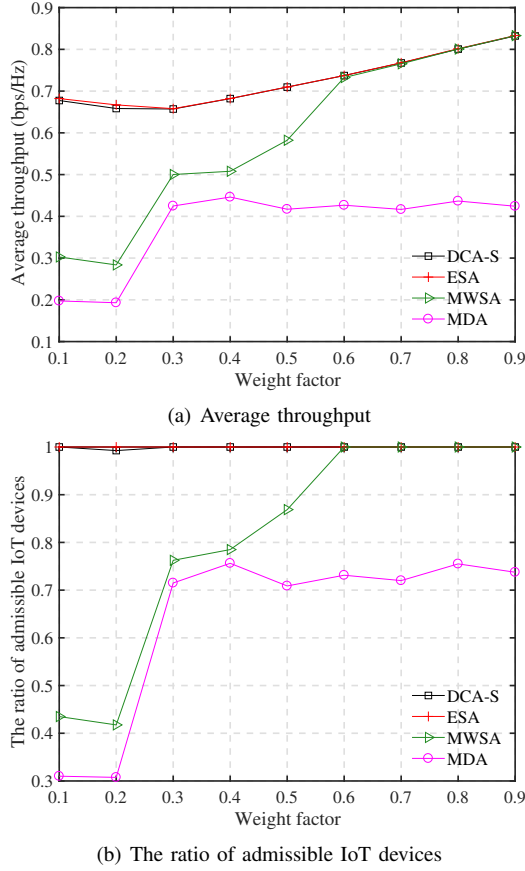
(a) Average throughput

(b) The ratio of admissible IoT devices

Fig. 6: Performance comparisons vs. weight factor when the channel information is known.

(a) Average throughput

(b) The ratio of admissible IoT devices

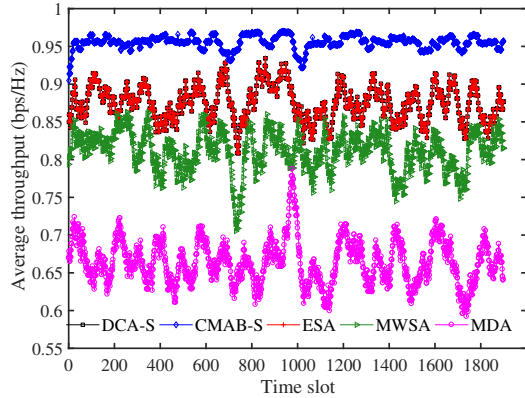Fig. 8: Performance comparisons vs. QoS threshold when the channel information is unknown.

Fig. 7: Comparison of average throughput as a function of time slot for online optimization.

the associated AP and BD and make access control decision instantaneously. Therefore, the online schemes of DCA-S, ESA, MWSA and MDA only have an estimation of incomplete CSI, instead of global channel information for all APs and BDs.

Fig. 7 illustrates the average throughput for online optimization versus time slot, where we set the weight factor to 0.2 and the QoS threshold to 15. We can observe that the average throughput of CMAB-S significantly outperforms other schemes. This is because CMAB-S prefers exploring
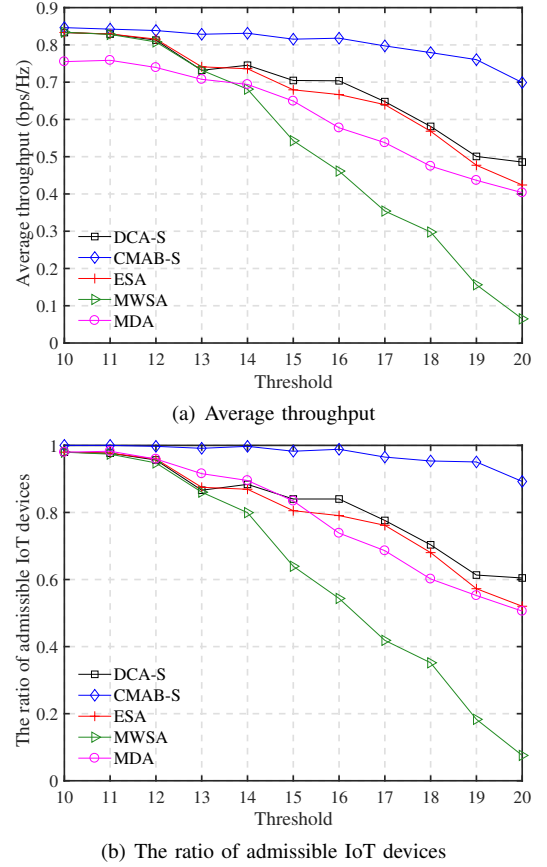
more possibilities rather than exploiting the currently optimal access control strategy in a dynamic environment. With the increase of time slot, CMAB-S can capture the statistic characteristics of uncertain environment and gradually learn the accurate estimation of reward distribution, so that the average throughput of CMAB-S becomes higher than that of other schemes.

Next, we evaluate the impact of QoS threshold on the performance under dynamic CSI scenario, where we set the weight factor to 0.5 and vary the QoS threshold from 10 to 20. As shown in Fig. 8(a), we can observe that CMAB-S scheme outperforms DCA-S, ESA, MWSA and MDA in average throughput. According to Fig. 8(b), the ratio of admissible IoT devices of CMAB-S that satisfies the QoS constraints is always higher than other schemes. Note that in this experiment, ESA may not be able to find the optimal solution due to the inaccurate estimation of channel information. Therefore, the performance of ESA is not always better than that of DCA-S in the dynamic environment. Finally, we examine the performance of the proposed online CMAB-S scheme with varying weight factor. In this experiment, we set QoS constraint to 15 and vary the weight factor from 0.1 to 0.9. As shown in Fig. 9, the performance of CMAB-S is always higher than that of other schemes, which validates the effectiveness of the proposed scheme.

(a) Average throughput

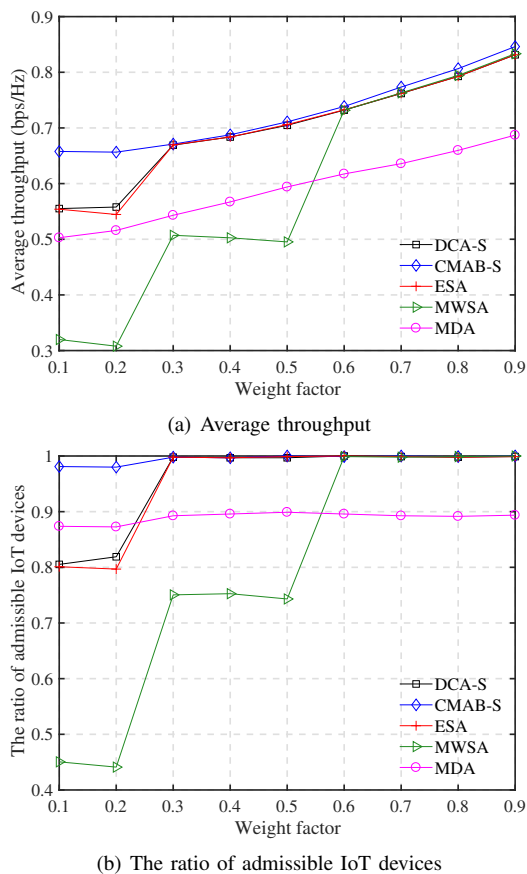(b) The ratio of admissible IoT devices

Fig. 9: Performance comparisons vs. weight factor when the channel information is unknown.

VI. RELATED WORK

A. Coefficient Design for AmBC Networks

In AmBC networks, coefficient design is indeed one of the most important factors affecting backscatter links and thus determines overall system performance. In full-duplex enabled cognitive backscatter network, the authors of [5] consider that a primary AP can transmit and receive signals simultaneously via full-duplex communications. In such scenario, the joint time scheduling, transmit power allocation, and coefficient adjustment are investigated to maximize the throughput of the AmBC system while guaranteeing the minimum rate requirements of the primary system. For monostatic backscatter communication, where carrier emitter and backscatter signal reader are the same entity, the authors of [26] investigate the throughput maximization problem by jointly optimizing precoding vector, beamforming vector and backscatter coefficient. Considering the non-convexity of the joint optimization problem, the authors of [26] design an individually-optimal solution for precoding, beamforming, and backscatter coefficient.

Furthermore, wireless powered backscatter networks [27] allow IoT devices to harvest energy, and then use the energy to transmit data. In this context, the authors of [28] consider a wireless powered backscatter network, and propose to maximize the energy efficiency by jointly optimizing transmission power and backscatter coefficient. In order to ensure fairness

among co-channel backscatter links, the problem is reduced to a max-min fairness optimization to solve. Moreover, in order to increase transmission range and throughput, the authors of [12] propose to integrate a hybrid backscatter communication into wireless-powered communication networks in a heterogeneous network (HetNet) environment. Depending on user location, this work targets at maximizing throughput for Macro-zone and WiFi-zone by optimizing time slots.

B. Device Association for AmBC Networks

From the perspective of network, the access control strategy aims to properly pair IoT devices and BDs in backscatter networks, so as to maximize the transmission capacity of primary and backscatter networks. The authors of [11] consider a symbiotic environment between cellular and IoT networks and focus on the user association problem in symbiotic radio networks. Due to the difficulty of obtaining full real-time channel information, the authors of [11] adopt deep reinforcement learning framework to infer the current information from the historical information, and effectively design centralized and distributed algorithms. In large-scale IoT scenarios, the authors of [29] focus on multi-transmitter multi-receiver communications for a backscatter network. By allowing multiple active communication pairs, the user association problem in such system is formulated as maximizing the number of access pairs, and a priority-based access strategy is proposed.

Considering the diversity gain of multiple tags, the authors of [30] propose a multi-tag selection combining scheme and analyze its outage performance under fully correlated and partially correlated Nakagami-m channels. Moreover, considering BD are capable of sleeping and active states, the authors of [31] investigate the throughput maximization problem through switching between sleep and active states and optimize backscatter coefficient in active state. To solve this optimization problem, a two-stage algorithm is proposed to obtain the optimal control policy. In [32], the authors investigate the throughput and bit error rate for a monostatic multi-tag backscatter communication system and proposed a general order tag selection criterion based on SNR. For the considered multi-tag system, the achievable rate and average bit error ratio are derived under a statistical delay constraint. In [33], the authors consider the multiple-input-multiple-output (MIMO) technology in multi-tag backscatter communications and focused on the optimal transceiver design for throughout fairness maximization. In [6], the authors focus on mathematical modeling and performance analysis for an ambient backscatter underlaying cellular networks. With the aim of maximizing the outage performance, the optimal and near-optimal time switching and power splitting schemes were proposed to enable energy harvesting and ambient backscattering.

Different from the above studies, which mainly focus on the single-RF source scenario and only consider BDs and IoT devices association, we propose to jointly optimize coefficient design and device association problems with the objective of maximizing system throughput in multi-RF source scenario, while satisfying the QoS guarantee of primary network. From the perspective of access control, this is the first study that extends the ideal single-RF source AmBC to a general multi-RF

source scenario, expecting to achieve higher data throughput and connections.

VII. CONCLUSION

Ambient backscatter communication is an emerging technology to achieve spectral, energy and cost efficiency in future wireless networks. In this paper, we have extended the ambient backscatter communication from the ideal single-RF source to a general multi-RF source scenario. In this context, we have investigated the access control strategy in multi-RF backscatter network from the perspective of system throughput maximization. Leveraging the convex optimization and dual decomposition techniques, an offline access control strategy (DCA-S) was proposed when the channel information is available a-priori in static scenario. Inspired by the obtained offline solutions, we have further proposed an online access control strategy (CMAB-S) based on combinatorial multi-armed bandit considering the channel dynamics and incomplete channel information. Numerical results demonstrate that our proposed DCA-S and CMAB-S can achieve significantly performance improvement in terms of average throughput and the number of admissible IoT devices as compared to the traditional schemes in typical scenarios.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] "Internet of things (iot) connected devices installed base worldwide from 2015 to 2025 (in billions)," Nov. 2016. [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide

[3] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient backscatter communications: A contemporary survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2889–2922, 2018.

[4] M. Nemati, M. Soltani, J. Ding, and J. Choi, "Subcarrier-wise backscatter communications over ambient ofdm for low power iot," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13229–13242, 2020.

[5] S. Xiao, H. Guo, and Y.-C. Liang, "Resource allocation for full-duplex-enabled cognitive backscatter networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3222–3235, 2019.

[6] L. Shi, R. Q. Hu, Y. Ye, and H. Zhang, "Modeling and performance analysis for ambient backscattering underlaying cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6563–6577, 2020.

[7] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green internet-of-things," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1116–1130, 2018.

[8] D. L. H. Guo, Q. Zhang and Y. Liang, "Noncoherent multiantenna receivers for cognitive backscatter system with multiple rf sources." [Online]. Available: https://arxiv.org/abs/1808.04316

[9] H. Guo, Q. Zhang, S. Xiao, and Y.-C. Liang, "Exploiting multiple antennas for cognitive ambient backscatter communication," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 765–775, 2019.

[10] J. Qian, F. Gao, G. Wang, S. Jin, and H. Zhu, "Semi-coherent detection and performance analysis for ambient backscatter system," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5266–5279, 2017.

[11] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Intelligent user association for symbiotic radio networks using deep reinforcement learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4535–4548, 2020.

[12] S. H. Kim and D. I. Kim, "Hybrid backscatter communication for wireless-powered heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6557–6570, 2017.

[13] R. Long, Y.-C. Liang, H. Guo, G. Yang, and R. Zhang, "Symbiotic radio: A new communication paradigm for passive internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1350–1363, 2020.

[14] H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local d.c. programming," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 510–515, 2012.

[15] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, 2004, pp. 284–289.

[16] F. Wang, W. Chen, H. Tang, and Q. Wu, "Joint optimization of user association, subchannel allocation, and power allocation in multi-cell multi-association ofdma heterogeneous networks," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2672–2684, 2017.

[17] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5481–5495, 2014.

[18] A. Khalili, S. Zarandi, and M. Rasti, "Joint resource allocation and offloading decision in mobile edge computing," *IEEE Communications Letters*, vol. 23, no. 4, pp. 684–687, 2019.

[19] P. Auer and R. Ortner, "Ucb revisited: Improved regret bounds for the stochastic multi-armed bandit problem," *Periodica Mathematica Hungarica*, vol. 61, no. 1, pp. 55–65, 2010.

[20] Y. Gai, B. Krishnamachari, and M. Liu, "On the combinatorial multi-armed bandit problem with markovian rewards," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011, pp. 1–6.

[21] N. C.-B. P. Auer and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.

[22] R. Munos, "From bandits to monte-carlo tree search: The optimistic principle applied to optimization and planning," *Foundations and Trends in Machine Learning*, 2014.

[23] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in matlab." [Online]. Available: https://yalmip.github.io/reference/

[24] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[25] T. Zhang, J. Zhao, L. An, and D. Liu, "Energy efficiency of base station deployment in ultra dense hetnets: A stochastic geometry analysis," *IEEE Wireless Communications Letters*, vol. 5, no. 2, pp. 184–187, 2016.

[26] D. Mishra and E. G. Larsson, "Sum throughput maximization in multi-tag backscattering to multiantenna reader," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5689–5705, 2019.

[27] B. Lyu, Z. Yang, G. Gui, and Y. Feng, "Wireless powered communication networks assisted by backscatter communication," *IEEE Access*, vol. 5, pp. 7254–7262, 2017.

[28] H. Yang, Y. Ye, and X. Chu, "Max-min energy-efficient resource allocation for wireless powered backscatter networks," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 688–692, 2020.

[29] Q. Chen, X. Zhang, J. Li, and J. Zhou, "Priority-based access strategy for multi-transmitter multi-receiver ambient backscatter communication system," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.

[30] Y. Zhang, F. Gao, L. Fan, S. Jin, and H. Zhu, "Performance analysis for tag selection in backscatter communication systems over nakagami-m fading channels," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–5.

[31] B. Lyu, C. You, Z. Yang, and G. Gui, "The optimal control policy for rf-powered backscatter communication networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2804–2808, 2018.

[32] Y. H. Al-Badarneh, M.-S. Alouini, and C. N. Georghiades, "Performance analysis of monostatic multi-tag backscatter systems with general order tag selection," *IEEE Wireless Communications Letters*, vol. 9, no. 8, pp. 1201–1205, 2020.

[33] D. Mishra and E. G. Larsson, "Multi-tag backscattering to mimo reader: Channel estimation and throughput fairness," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5584–5599, 2019.

Long Zhang received the M.E degree in information and communication engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2019. He currently is pursuing his Ph.D. degree at the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, China. His research areas include next generation mobile networks and Internet of Things.

Bin Cao (M'14) received the Ph.D. (Hons.) degree in communication and information systems from the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Sichuan, China, in 2014. He is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. Before that, he was an Associate Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. From April to December in 2012, he was an international Visitor with the Institute for Infocomm Research, Singapore. From July 2015 and July 2016, he was a Research Fellow with the National University of Singapore, Singapore. His research interests include blockchain system, Internet of Things, and mobile edge computing. He was also the Symposium Co-Chair of the IEEE ICNC 2018, workshop cochair for CyberC 2019 and a TPC Member for numerous conferences.

Gang Feng (M'01–SM'06) received his BEng and MEng degrees in Electronic Engineering from the University of Electronic Science and Technology of China (UESTC), in 1986 and 1989, respectively, and the Ph.D. degrees in Information Engineering from The Chinese University of Hong Kong in 1998. He joined the School of Electric and Electronic Engineering, Nanyang Technological University in December 2000 as an assistant professor and was promoted as an associate professor in October 2005. At present he is a professor with the National Key Laboratory of Science and Technology on Communications, UESTC of China. Dr. Feng has extensive research experience and has published widely in computer networking and wireless networking research. His research interests include resource management in wireless networks, next generation cellular networks, etc. Dr. Feng is a senior member of IEEE.

Shuang Qin received the B.E. degree in Electronic Information Science and Technology, and the Ph.D degree in Communication and Information System from University of Electronic Science and Technology of China (UESTC), in 2006 and 2012, respectively. He is currently an associate professor with National Key Laboratory of Science and Technology on Communications in UESTC. His research interests include cooperative communication in wireless networks, data transmission in opportunistic networks and green communication in heterogeneous.

Yao Sun is currently a Lecturer with James Watt School of Engineering, the University of Glasgow, Glasgow, UK. Dr. Sun has extensive research experience in wireless communication area. He has won the IEEE Communication Society of TAOS Best Paper Award in 2019 ICC. He has been the guest editor for special issues of several international journals. He has been served as TPC Chair for UCET 2021, and TPC member for number of international conferences, including ICC 2022, VTC spring 2022, GLOBECOM 2020, WCNC 2019, ICCT 2019. His research interests include intelligent wireless networking, network slicing, blockchain system, internet of things and resource management in mobile networks. Dr. Sun is a senior member of IEEE.