Dynamic Spectrum Access for Battery-Operated IoT Using Multi-Armed Bandit

***Abstract*—The multi-armed bandit (MAB) provides a key theoretic framework for spectrum decision for the cognitive radio networks (CRN), where statistical characteristics of time-varying wireless links are unknown. For battery-operated users, classical approaches fail to deal with the energy budget, where dynamics of the residual energy and lifetime have great influence on channel sensing and access decisions. In this paper, we investigate the dynamic spectrum access (DSA) problem in a cognitive radio IoT where the secondary IoT user (SU) has the limited energy supply. The channel statistical information, including the channel idle probability and channel quality, is unknown to the SU. In each time slot, the SU has to make decisions to choose channels to sense, estimate the channel quality, and learn the channel statistical information with part energy consumption. When channels are sensed idle, the SU adjusts the transmit power and accesses all idle channels for information transmission. To maximize the cumulative data volume within the limited energy, we model the problem within the MAB framework and develop an upper confidence bound based algorithm, which enables the SU to gradually learn the largest throughput-to-energy ratio. Theoretical analysis shows that the data volume loss, which represents the difference from the theoretical maximal cumulative data volume, exhibits a logarithmic relationship with the energy. Furthermore, we validate the algorithm effectiveness under various channel models.**

***Index Terms*—Dynamic spectrum access, Multi-armed bandit, Energy constraint, Internet of things.**

I. INTRODUCTION

THE rapid development of wireless network and its widespread Internet of Things (IoT) have produced a profound impact on industrial paradigm, and put forward various challenges enabling massive wireless connection and high spectrum efficiency. To address these issues, new communication technologies emerge and attract much research attention. In them, cognitive radio, with its capability of improving spectrum efficiency, is a promising technology. Its potential is particularly notable in the context of IoT [1], [2] and low power wide area IoT networks [3].

In DSA within the CR-IoT, there exist two categories of users: primary users (PUs) and secondary IoT users (SUs). PUs hold the license and priority to utilize the spectrum. In contrast, SUs, lacking these privileges, have to employ channel sensing to discover available spectrum. Once identified, SUs can dynamically use these resources while ensuring no harmful interference to PUs. Compared with the conventional CR, the CR-IoT has some distinct characteristics that should be mentioned. A critical aspect of CR-IoT is that the IoT devices are powered by batteries, which provide a limited energy supply. The process of channel sensing and access is energy-sensitive, hence the available energy significantly influences its lifespan and cumulative data volume it can obtain. This makes the energy management and conservation a key consideration when designing the system. Another challenge arises due to the fact that IoT users in low power wide area networks are generally deployed in the congested industrial, scientific, and medical bands, which are invariably plagued by problems associated with spectral congestion, such as increased interference and time-variant channel availability states. In the context, SUs may not possess prior knowledge about the channel statistical information, and they must engage in a learning process alongside the channel sensing and access process to accumulate the information [4]. Overall, balancing the limited energy use with the effective learning and spectrum utilization is a complex challenge that necessitates a careful system design, which is the focus of this paper.

A. Related Works

Extensive research has focused on DSA strategies using machine learning techniques for wireless communication networks where statistical characteristics of time-varying wireless links are unknown. Compared with the machine learning including Reinforcement Learning and Deep Reinforcement Learning, the MAB has the characteristics like optimal solutions, ease of theoretical guarantee, and high computational efficiency [5]. The MAB model is a framework for decision-making under uncertainty, drawing its name from the analogy of a gambler choosing which arm of a multi-armed slot machine to pull to maximize his winnings. In the context of wireless communication, each arm represents a different channel or strategy for accessing the spectrum. The MAB model is particularly suitable for DSA in CRNs because it efficiently balances the exploration of channels (i.e., to gather statistical information) with the exploitation of known channels (i.e., to maximize data transmission). This balance is crucial in environments where the occupancy status of primary channels varies over time and is initially unknown to SUs.

As a research pioneer, Lai et al. in [6] establish the MAB framework for DSA strategy design in CRNs. In the model, the occupancy status of primary channels in each slot follows an independent and identically distributed (i.i.d.) reward model. The challenge for the SU, which has no prior knowledge of the idle probability of primary channels, is to select the best channel for opportunistic access in each time slot. Due to the initial lack of statistical information about channels, the SU must explore multiple primary channels to update their beliefs while simultaneously exploiting the preferred channels based on the information accumulated. To maximize the expected cumulative data volume, an MAB problem has been formulated. This problem equivalently transforms the maximization issue into finding an online learning-based algorithm to minimize the regret, i.e., the difference between the expected transmission traffic in actual and that achieved by the optimal strategy with SU knowing the channel statistic information. To solve the MAB problem, an upper confidence bound (UCB) based DSA strategy has been proposed and proved to be order optimal with regret $O(\ln t)$ with time slot number t. As research deepens, a substantial amount of studies for multiple users based on the MAB framework have been investigated [7]–[9].

When statistics are unknown and users have limited energy supply, the problem becomes much more challenging since users have to learn the statistics online, as well as balancing the energy consumption and transmission performance [10]–[14]. As IoT users are usually battery-operated and the system lifetime depends on the battery reservation, DSA strategies adapting to both varying link condition and energy states at run time are demanded. For future intelligent networks, Sherman et al. in [15] consider the statistical decision problem of channel selection in the visible light frequency band and radio frequency band (including 2.4 GHz, 5.25 GHz, and 38 GHz) where the user has limited energy. Due to the unknown link fading characteristics, authors design two online decision algorithms, namely, energy-aware band selection with upper confidence bound (EABS-UCB) and energy-aware band selection with Thompson sampling (EABS-TS). Moreover, in IoT network applications, [16] considers the scenario where users obtain random rewards and consume random energy during the channel selection process. The paper proposes the c-UCB algorithm, which selects channels by computing the upper confidence bound of the ratio of average reward to average energy consumption for each channel. [17] introduces the Budget-Limited First algorithm, where the user uses εE (ε is the predefined parameter and E is the overall energy) energy to access all channels and calculates the upper confidence bound of the ratio of average reward to average energy consumption for each channel. After the exhaustion of εE energy, the user continues to access the channel with the maximum upper confidence bound using the residual energy.

B. Motivation and Contributions

As can be concluded, the majority of energy constraint-related research lies in the IoT networks and future wireless networks, but does not address the DSA strategies in the context of CR-IoT. In CR-IoT, when the SU senses the channel idle and transmits, it can obtain the data volume but consumes part energy; whereas when the SU senses the channel occupied, it cannot transmit the data but consumes less energy. Therefore, the data volume obtained and energy consumed are highly correlated. However, most current research assumes that the data volume gained and energy consumption are independent variables, and does not fully consider the correlation between the two, thereby failing to assess the efficiency in CR-IoT.

In this paper, we investigate the DSA problem where the channel statistics are unknown and the SU has the energy constraint in the context of CR-IoT. We first model the problem within the MAB framework, where the throughput-to-energy ratio of the channel is regarded as the arm and user as the player. Different from the traditional MAB framework, we additionally consider the energy constraint for the problem, which complicates the problem. Then, we develop an algorithm based on the upper confidence bound for the estimation of the throughput-to-energy ratio and greedy policy, from which the SU can gradually distinguish the optimal channel. To verify the effectiveness, we theoretically analyze the algorithm performance in the form of data volume loss, showing it has the sub-linear relationship with the energy constraint. In the final, we validate the algorithm performance under various channel models and compare with other algorithms.

C. Organizations and Notations

The paper is organized as follows. The system model and problem formulation are depicted in Section II. We design the algorithm and theoretical analysis in Section III. System performance is evaluated in Section IV, followed by conclusions drawn in Section V.

Notations: The operators $\mathbb{E}$ and $\mathbb{E}_x$ denote the expectation with respect to all the randomness and the expectation with respect to x and other randomness, respectively. Define $\mathcal{I}_N = \{1, 2, \dots, N\}$ as a shorthand as the index set, and $\mathbb{I}[\cdot]$ as the indicator function. Equation is abbreviated as Eq. in the following.

II. System Model and Problem Formulation

In this section, we first depict the system model, transmission process for the system model and then formulate the problem.

A. System Model

We consider a wireless network consisting of N channels denoted as $\mathcal{I}_N = \{1, 2, ..., N\}$. The network operates in a time-slotted manner with the fixed slot length. In each slot, channel $i \in \mathcal{I}_N$ is idle (i.e., without PU activities) with probability θ_i and occupied with $1-\theta_i$. Moreover, channels follow the independent quasi-static block fading during each slot. For each channel, the channel state and quality vary independently from a slot to another. And the N channels have independent channel states.

In the model, we take the energy consumption into consideration referred to [18]. The SU transmitter has the limited energy denoted as $\mathcal{E}$. In the process of the transmission process, the energy is consumed and instantaneous residual energy at the beginning of slot t is $\mathcal{E}(t)$. When the residual energy is less than the threshold ε_{thres}, the SU stops the transmission process.

B. Transmission Process

The transmission process in a slot consists of channel selection, channel sensing, estimation, transmission and acknowledgment (ACK) reception, which are depicted in Figure 1 and described as follows.

1) Channel selection, sensing and estimation: At the beginning of each slot, the SU makes the decision to select M $(1 \leq M < N)$ channels and takes the duration of τ to sense their availability states. We assume that the energy consumption of the selection is c^d and channel sensing for channel

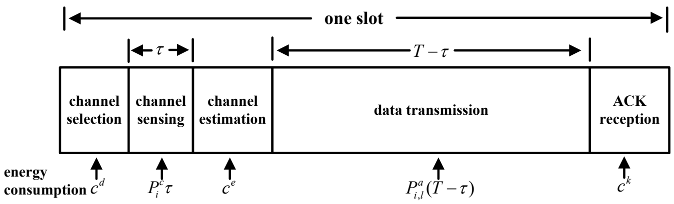

Fig. 1. The relationship between the data volume loss and energy.

i is $P_i^c\tau$, respectively. When the channel is sensed idle, the SU transmitter estimates the instantaneous channel quality. We assume that the energy consumption of the estimation is c^e.

2) Transmission and ACK reception: In this model, we assume that the SU has L power levels denoted as $\mathcal{I}_L = \{1, 2, ..., L\}$ and a transmission rate requirement R bits per second (bps) on a channel. The SU thus has to self-adaptively adjust the transmit power in accordance with the instantaneous channel quality. In specific, for channel i at slot t, the SU estimates the ergodic rate with transmit power $P(t)$ and instantaneous channel quality as

$$R_i(t) = B_i \log_2\left(1 + \frac{|H_i(t)|^2 P(t)}{\sigma_i^2(t)}\right) \tag{1}$$

where B_i is the channel bandwidth, $|H_i(t)|^2$ is the power gain of channel i at slot t and $\sigma_i(t)$ is the channel noise at the SU receiver. Then, the minimum transmit power to meet the transmission rate R is

$$P_i^{\min}(t) \geq \frac{(2^{R/B_i} - 1)\sigma_i^2(t)}{|H_i(t)|^2}. \tag{2}$$

Therefore, the SU transmitter adjusts the transmission power and uses the duration of $T-\tau$ to transmit at a fixed rate of R bits per second (bps) on each sensed-idle channel. The energy consumption of the transmission is thus $P_i^a(t)(T-\tau)$, where $P_i^a(t) \in \{P_{i,l}^a : l \in \mathcal{I}_L\}$ is the transmit power level and $T-\tau$ is the transmission duration.

After the transmission, the SU receiver transmits the Acknowledge (ACK) packet to the transmitter for the confirmation. The energy consumption of the ACK reception is denoted as c^k.

Overall, when the SU senses channel i, the transmission rate at slot t can be

$$R_i(t) = \begin{cases} 0, & i \text{ is occupay} \\ R \text{ (bps)}, & i \text{ is idle} \end{cases}, \tag{3}$$

the expected data volume on channel i can be

$$\mu_i^r = \mathbb{E}\{R_i(t)\} \cdot (T-\tau) = \theta_i R(T-\tau). \tag{4}$$

When SU selects channel i at slot t, the energy consumption can be

$$c_i(t) = \begin{cases} c^d + P_i^c\tau, & i \text{ is occupay} \\ c^d + P_i^c\tau + c^e + P_{i,l}^a(T-\tau) + c^k, & i \text{ is idle} \end{cases}.$$

The expected energy consumption on channel i is thus

$$\mu_i^c = \mathbb{E}\{c_i(t)\} = c^d + P_i^c\tau + \theta_i\left(c^e + \sum_{l=1}^{L} P_{i,l}^a(T-\tau)p_{i,l} + c^k\right),$$

where θ_i represents the availability (idle) probability of channel i, $p_{i,l}$ represents the probability that the SU transmits with power $P_{i,l}$ at channel i.

C. Problem Formulation

In our problem, the statistical information on channel availability states and channel quality is unknown to the SU. The aim for SU is to find an efficient algorithm to learn the unknown channel statistics, make the appropriate decision to select channels and obtain the largest cumulative data volume within the limited energy.

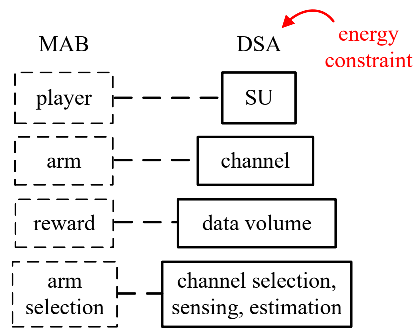

Fig. 2. The relationship between the data volume loss and energy.

To solve the problem, we formulate the problem within the MAB framework. In specific, the SU is regarded as a player, the energy budget is the constraint, the channel is the arm, the energy consumed and traffic volume for one time data transmission over a channel are the cost and reward by puling an arm, respectively. Under this framework, we define an online leaning policy based strategy as $\pi = \{a(t)\}_{t=1}^{\infty}$, which is a sequence of channel selection actions by SU at time slot t until energy budget is depleted. For algorithm π, the expectation of the cumulative data volume can be

$$\begin{aligned} &\mathbb{E}\{U^\pi(\mathcal{E})\} \\ &= \mathbb{E}\left\{\sum_{i=1}^{N}\sum_{t=1}^{\infty} R_i(t) \cdot (T-\tau) \cdot \mathbb{I}[i \in a(t), \varepsilon(t) > \varepsilon_{thres}]\right\} \\ &= \sum_{i=1}^{N}\sum_{t=1}^{\infty} \mu_i^r \cdot \mathbb{E}\{\mathbb{I}[i \in a(t), \varepsilon(t) > \varepsilon_{thres}]\}, \end{aligned}$$

where $\mathbb{I}[\cdot]$ represents the indicator function, $a(t)$ represents the channel set chosen at slot t, $\varepsilon(t) > \varepsilon_{thres}$ represents that the residual energy is larger than the threshold. When SU selects channel i and the residual energy exceeds the threshold at slot t, $\mathbb{I}[i \in a(t), \varepsilon(t) > \varepsilon_{thres}] = 1$.

In all algorithms, there exists an optimal algorithm π^* that the SU can obtain the largest expected cumulative data volume, denoted as $U^*(\mathcal{E})$. Then, the data volume loss, denoted as $R(\pi; \mathcal{E})$, can be given as

$$R(\mathcal{E}; \pi) = \mathbb{E}\{U^*(\mathcal{E})\} - \mathbb{E}\{U^\pi(\mathcal{E})\}. \tag{5}$$

Intuitively, to design the algorithm that maximizes the expected cumulative data volume can be equivalent to minimizing the data volume loss, which can be regarded as the goal of our problem.

D. Performance Benchmark

To analyze the algorithm performance, we first give the upper bound of the cumulative data volume for our problem in Lemma 1.

Lemma 1: For any algorithm π, the upper bound for the expectation on the cumulative data volume with energy constraint $\mathcal{E}$ is $\frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}}\mathcal{E}$, where $\mathcal{C}_{\max} = \underset{\mathcal{C}_m, m\in\mathcal{I}_{\binom{N}{M}}}{\arg\max}\left\{\frac{\mu^r_{\mathcal{C}_m}}{\mu^c_{\mathcal{C}_m}}\right\}$, $\mathcal{C}_m$ represents the m-th channel set, $\mathcal{I}_{\binom{N}{M}} = \{1, 2, ..., \binom{N}{M}\}$ originates from the fact that the SU selects M from N channels in each slot, leading to $\binom{N}{M}$ channel sets. $\mu^r_{\mathcal{C}_m} = \sum_{i\in\mathcal{C}_m}\mu^r_i$ and $\mu^c_{\mathcal{C}_m} = \sum_{i\in\mathcal{C}_m}\mu^c_i$ represent the expected data volume and energy consumption of selecting channel set $\mathcal{C}_m$, respectively.

Proof: The proof is shown in Appendix A. ■

III. Algorithm Design and Performance Analysis

In this section, we separately depict the algorithm design, data volume loss performance and complexity analysis.

A. Algorithm Design

We design an algorithm referred to as π-Energy Constrained (π-EC). The main idea is that the SU can estimate the throughput and energy consumption by observations of channel samplings. Then, the SU in each slot designs indices in the form of throughput-to-energy ratio for all channels with the estimated information and sorts them in a descending order. Channels with the largest indices are regarded better channels for SUs. The overall algorithm is depicted in Algorithm 1 and described as follows.

In initialization process, the SU senses and estimates each channel once, obtains the number of sensing times as

$$T_i(t) = \sum_{l=1}^{t}\mathbb{I}[i\in a(l)], \forall\ i\in\mathcal{I}_N, \tag{6}$$

obtains the estimated mean throughput as

$$\bar{\mu}^r_i(t) = \max\left\{\frac{\sum_{l=1}^{t}R_i(l)\mathbb{I}[i\in a(l)]}{T_i(t)}, 1\right\}, \forall\ i\in\mathcal{I}_N, \tag{7}$$

obtains the estimated mean energy consumption as

$$\bar{\mu}^c_i(t) = \min\left\{\frac{\sum_{l=1}^{t}c_i(l)\mathbb{I}[i\in a(l)]}{T_i(t)}, c_{\min}\right\}, \forall\ i\in\mathcal{I}_N, \tag{8}$$

where $c_{\min} = \min_{i\in\mathcal{I}_N, l\in\mathcal{I}_L}\{c^d + P^a_{i,l}\tau\}$ represents the minimum energy consumption known to SU *a priori*.

In a subsequent slot, i.e., slot t, the SU firstly checks whether the residual energy is larger than the energy threshold. If the residual energy is less than the energy threshold, it stops the process; otherwise, the SU follows the process as follows: it firstly computes index for each channel, depicted as

$$g_i(t) = \frac{1}{\alpha}\cdot\frac{\bar{\mu}^r_i(t)}{\bar{\mu}^c_i(t)} + \sqrt{\frac{\alpha_1\ln t}{T_i(t)}}, \forall\ i\in\mathcal{I}_N, \tag{9}$$

Algorithm 1 $\pi-$EC

Initialization: the SU senses each channel once, obtains the information on throughput and energy consumption, updates the residual energy and sensing number $T_i(t)$, $\bar{\mu}^r_i(t)$, and $\bar{\mu}^c_i(t)$, $\forall\ i\in\mathcal{I}_N$
for slot t **do**
  if residual energy $\varepsilon(t)\geq\varepsilon_{thres}$ **then**
    for $i = 1 : N$ **do**
      Updates $g_i(t) = \frac{1}{\alpha}\cdot\frac{\bar{\mu}^r_i(t)}{\bar{\mu}^c_i(t)} + \sqrt{\frac{\alpha_1\ln t}{T_i(t)}}$
    end for
    Descends channels in accordance with $g_i(t)$, $\forall\ i\in\mathcal{I}_N$, and selects M top channels $\mathcal{C}^*(t)$
    Senses the channel states from $\mathcal{C}^*(t)$ and obtains the idle channels
    Estimates the channel quality of idle channels and adjusts the transmit power
    Accesses all sensed-idle channels from $\mathcal{C}^*(t)$
    Receives the ACK from the receiver
    Updates $\bar{\mu}^r_i(t)$ and $\bar{\mu}^c_i(t)$, $\forall\ i\in\mathcal{C}^*(t)$
    Updates $U^\pi(\mathcal{E})\leftarrow U^\pi(\mathcal{E}) + \sum_{i\in\mathcal{C}^*(t)}R_i(t)$
    Updates the residual energy $\varepsilon(t)\leftarrow\varepsilon(t-1) - \sum_{i\in\mathcal{C}^*(t)}c_i(t)$
  else
    Stops the sensing and access process
  end if
end for

where α and α_1 are pre-defined parameters, enabling $\frac{1}{\alpha}\cdot\frac{\bar{\mu}^r_i(t)}{\bar{\mu}^c_i(t)}\in[0,1]$. $\frac{\bar{\mu}^r_i(t)}{\bar{\mu}^c_i(t)}$ represents the average estimated throughput that can be obtained by sensing channel i at slot t within the average estimated energy consumption.

Subsequently, the SU descends channels in accordance with channel indices and selects top M channels to sense. Then, the SU estimates the channel quality of the idle channels, adjusts the transmit power and transmits for duration of $T-\tau$. After the transmission, the SU transmitter receives the ACK from the receiver, updates the cumulative data volume, residual energy, and estimated information on all channels, i.e., $\bar{\mu}^r_i(t)$ and $\bar{\mu}^c_i(t)$ in accordance with Eq. (7) and (8), $\forall\ i\in\mathcal{C}^*(t)$.

Remark: In our algorithm, we use $\frac{\bar{\mu}^r_i(t)}{\bar{\mu}^c_i(t)}$ to distinguish the optimality of various channels. A higher value of the ratio indicates that the channel may provide a larger throughput under the unit estimated energy consumption. Moreover, the term $\sqrt{\frac{\alpha_1\ln t}{T_i(t)}}$ in Eq. (9) is designed to form the index with the following function: when channel i is sensed relatively few times, the value of $\sqrt{\frac{\alpha_1\ln t}{T_i(t)}}$ will be relatively large. Therefore, channel i is more likely to be chosen. When all channels have been sufficiently sensed, the SU can finally distinguish the difference of various channels.

B. Performance Analysis on Data Volume Loss

In this subsection, we analyze the algorithm performance. Since the SU has the limited energy in our problem, there exists a stopping time that the SU stops the sensing and access process as the energy runs out. The stopping time is a random variable related to the algorithm, channel idle probability, channel quality, and energy budget. Denoting τ as the stopping time, the expectation of the cumulative data volume within energy $\mathcal{E}$ can be

$$\begin{aligned}\mathbb{E}\{U^{\pi}(\mathcal{E})\} &= \sum_{j=1}^{\binom{N}{M}} \sum_{t=1}^{\infty} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\left\{\mathbb{I}[a(t)=\mathcal{C}_j, \varepsilon(t) > \varepsilon_{thres}]\right\} \\ &= \sum_{j=1}^{\binom{N}{M}} \mathbb{E}_{\tau}\left\{\sum_{t=1}^{\tau} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\}, \end{aligned} \quad (10)$$

where $\mathbb{E}\left\{\mathbb{I}[a(t)=\mathcal{C}_j, \varepsilon(t) > \varepsilon_{thres}]\right\} = \mathbb{P}[a(t)=\mathcal{C}_j, \varepsilon(t) > \varepsilon_{thres}]$ represents that the channel sensing set at slot t is $\mathcal{C}_j$ and $\varepsilon(t)$ is larger than the energy threshold. $\mathbb{E}_{\tau}\{.\}$ represents the expectation over the stopping time, $\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\} = \mathbb{P}[a(t)=\mathcal{C}_j]$ represents the probability of sensing the channel set $\mathcal{C}_j$, $j \in \mathcal{I}_{\binom{N}{M}}$.

Then, the data volume loss can be expressed as

$$\begin{aligned} R(\mathcal{E};\pi) &= \mathbb{E}\{U^{\pi^*}(\mathcal{E})\} - \mathbb{E}\{U^{\pi}(\mathcal{E})\} \\ &\le \frac{\mu^{r}_{\mathcal{C}_{\max}}}{\mu^{c}_{\mathcal{C}_{\max}}}\mathcal{E} - \sum_{j=1}^{\binom{N}{M}} \mathbb{E}_{\tau}\left\{\sum_{t=1}^{\tau} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\} && (11) \\ &= \frac{\mu^{r}_{\mathcal{C}_{\max}}}{\mu^{c}_{\mathcal{C}_{\max}}}\mathcal{E} - \mu^{r}_{\mathcal{C}_{\max}}\mathbb{E}_{\tau}\{\tau\} + \mu^{r}_{\mathcal{C}_{\max}}\mathbb{E}_{\tau}\{\tau\} && (12) \\ &\quad - \sum_{j=1}^{\binom{N}{M}} \mathbb{E}_{\tau}\left\{\sum_{t=1}^{\tau} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\} \\ &= \underbrace{\mu^{r}_{\mathcal{C}_{\max}}\left(\frac{\mathcal{E}}{\mu^{c}_{\mathcal{C}_{\max}}} - \mathbb{E}_{\tau}\{\tau\}\right)}_{(A)} && (13) \\ &\quad + \underbrace{\mathbb{E}_{\tau}\left\{\tau\mu^{r}_{\mathcal{C}_{\max}} - \sum_{j=1}^{\binom{N}{M}}\sum_{t=1}^{\tau} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\}}_{(B)}. \end{aligned}$$

From Eq. (13), we can see that we should analyze the upper bound for terms (A) and (B), which are presented in Lemma 2 and Lemma 3, respectively.

Lemma 2: The upper bound of term (A) from Eq. (13) is

$$\begin{aligned} &\mu^{r}_{\mathcal{C}_{\max}} \cdot \left(\frac{\mathcal{E}}{\mu^{c}_{\mathcal{C}_{\max}}} - \mathbb{E}_{\tau}\{\tau\}\right) && (14) \\ &\le \mu^{r}_{\mathcal{C}_{\max}} \cdot \frac{\varepsilon_{thres} + \sum\limits_{\mathcal{C}_j \in \mathcal{C}^{\dagger}} \Delta_{\max}\left(\frac{4\alpha_1}{d^2_{\min}} \ln\left(\frac{\mathcal{E}}{M\mu^{c}_{\min}}\right) + 1 + \frac{2\pi^2}{3}\right)}{\mu^{c}_{\mathcal{C}_{\max}}}, \end{aligned}$$

where $\Delta_{\max} = \max\limits_{j \in \mathcal{I}_{\binom{N}{M}}} \left\{\left|\mu^{c}_{\mathcal{C}_j} - \mu^{c}_{\mathcal{C}_{\max}}\right|\right\}$, $\mathcal{C}^{\dagger} = \left\{\mathcal{C}_j : \mu^{c}_{\mathcal{C}_j} > \mu^{c}_{\mathcal{C}_{\max}}, j \in \mathcal{I}_{\binom{N}{M}}\right\}$, $d_{\min} = \min\limits_{\mathcal{C}_j \ne \mathcal{C}_{\max}, j \in \mathcal{I}_{\binom{N}{M}}} \left\{\left|\frac{\mu^{r}_{\mathcal{C}_{\max}}}{\mu^{c}_{\mathcal{C}_{\max}}} - \frac{\mu^{r}_{\mathcal{C}_j}}{\mu^{c}_{\mathcal{C}_j}}\right|\right\}$, and $\mu^{c}_{\min} = \min\limits_{i \in \mathcal{I}_N} \{\mu^{c}_{i}\}$.

Proof: The proof can be referred to Appendix B. ■

Lemma 3: The upper bound of term (B) from Eq. (13) is

$$\begin{aligned} &\mathbb{E}_{\tau}\left\{\tau \cdot \mu^{r}_{\mathcal{C}_{\max}} - \sum_{j=1}^{\binom{N}{M}}\sum_{t=1}^{\tau} \mu^{r}_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\} \\ &\le \Delta_{\max}\binom{N}{M}\left(\frac{4\alpha_1}{d^2_{\min}} \ln\left(\frac{\mathcal{E}}{M\mu^{c}_{\min}}\right) + 1 + \frac{2\pi^2}{3}\right). \end{aligned} \quad (15)$$

Proof: The proof can be referred to Appendix C. ■

In the final, we can conclude the theorem from Lemma 2 and Lemma 3 for the data volume loss as follows.

Theorem 1: The upper bound for the data volume loss has the following relationship

$$\begin{aligned} &R(\mathcal{E};\pi) && (16) \\ &\le \mu^{r}_{\mathcal{C}_{\max}} \frac{\varepsilon_{thres} + \sum\limits_{\mathcal{C}_j \in \mathcal{C}^{\dagger}} \Delta_{\max}\left(\frac{4\alpha_1}{d^2_{\min}} \ln\left(\frac{\mathcal{E}}{M\mu^{c}_{\min}}\right) + 1 + \frac{2\pi^2}{3}\right)}{\mu^{c}_{\mathcal{C}_{\max}}} \\ &\quad + \Delta_{\max}\binom{N}{M}\left(\frac{4\alpha_1}{d^2_{\min}} \ln\left(\frac{\mathcal{E}}{M\mu^{c}_{\min}}\right) + 1 + \frac{2\pi^2}{3}\right). \end{aligned}$$

It can be known from the above conclusions, the upper bound for the data volume loss of the proposed algorithm is $O(\ln \mathcal{E})$. It shows that the SU has to utilize the $O(\ln \mathcal{E})$ energy to sense various channels to obtain the estimated information, and utilize the $\mathcal{E} - O(\ln \mathcal{E})$ energy to access the optimal channels. The latent meaning is that the SU can finally distinguish the optimal channels if the energy is large enough.

C. Complexity Analysis

In this subsection, we analyze the space and computational complexity of the algorithm, respectively.

1) Space Complexity: The space complexity quantifies the amount of memory that an algorithm requires during its execution. In the process of channel sensing and access, since the data length of $g_i(t)$, $T_i(t)$, $\bar{\mu}^{r}_{i}(t)$ and $\bar{\mu}^{c}_{i}(t)$ is N, the SU takes the complexity $O(N)$ to update the above data. Then, the SU takes the complexity $O(1)$ to update $\varepsilon(t)$, record $U^{\pi}(\mathcal{E})$ and time t. Since the largest lifetime is $\frac{\mathcal{E}}{Mc_{\min}}$, the overall space complexity is thus $O\left(\frac{\mathcal{E}}{Mc_{\min}}N\right)$.

2) Computational Complexity: The computational complexity measures the amount of computational time that the algorithm requires. In the process of channel sensing and access, the SU takes $O(1)$ time at slot t to check whether the residual energy exceeds the energy threshold. Since $g_i(t)$, $\forall\ i \in \mathcal{I}_N$ has the data length of N, the SU takes $O(N)$ and $O(N^2)$ time to update and sort the data, respectively. Since $\bar{\mu}^{r}_{i}(t)$ and $\bar{\mu}^{c}_{i}(t)$ have the data length of M, the SU takes $O(M)$ time to calculate the above data. Subsequently, the SU takes $O(1)$ time to update $\mathcal{C}^{*}(t)$, $U^{\pi}(\mathcal{E})$ and the residual energy. Since the largest lifetime is $\frac{\mathcal{E}}{Mc_{\min}}$, the overall computational complexity is thus $O\left(\frac{\mathcal{E}}{Mc_{\min}}N^2\right)$.

IV. Simulation

In this section, we perform simulations on Python 3.8.5 with two metrics, namely, data volume loss and energy efficiency[1].

[1]The codes are open-sourced at https://github.com/davidloveshet/thesis-code.

The data volume loss is defined in subsection II-C and energy efficiency is defined as the ratio of cumulative data volume and energy consumption. The growth of data volume loss slows down, or the energy efficiency converges to the optimal energy efficiency (indicating that the maximum data volume can be obtained by the SU within unit energy consumption) mean that the SU tends to sense and access the optimal channels.

A. The Discrete Energy Consumption Model

Scenario: We refer to [18] to set up the simulation scenario. We assume that different channels experience various unknown interference and the SU has four transmit powers, i.e., $\frac{1}{4}P_{\max}$, $\frac{1}{2}P_{\max}$, $\frac{3}{4}P_{\max}$, and $P_{\max} = 200$ mW. Since various channels have various channel qualities, the SU selects the transmit power on the channel with a certain probability, which is depicted in Table I. For example, $[0.062, 0.35, 0.379, 0.208]$ represents that the SU chooses to transmit with a probability of 0.062 at a transmit power of $\frac{1}{4}P_{\max}$, with a probability of 0.35 at a transmit power of $\frac{1}{2}P_{\max}$, with a probability of 0.379 at a transmit power of $\frac{3}{4}P_{\max}$, and with a probability of 0.208 at a transmit power of $P_{\max}$ on channel 1, respectively. We assume that the transmission rate requirement is 1.2 Mbps and energy ranges from $0 \sim 250$ J. Part simulation parameters refer to [20], where the energy consumption of channel allocation and channels sensing is $2.5e^{-04}$ J, and the transmission duration in a slot is set as 95 ms[2]. Algorithm parameters are set with $\alpha = \frac{1}{20}$, $\alpha_1 = \frac{1}{2}$, and the number of Monte Carlo is 200.

The simulation results are shown in Figures 3 and 4. Figure 3 shows the relationship between the data volume loss and energy. It can be concluded that for various sensing number $M = 1, 2, 3, 4$, the data volume loss has the sub-linear relationship with the energy. This implies that as the energy increases, the data volume loss tends to approach the benchmark. This result indicates that as energy increases, the SU tends to sense and access optimal channels (otherwise, the data volume loss exhibits a linear relationship with the energy). Figure 4 represents the relationship between the energy efficiency and energy, where 'optimal' represents the optimal energy efficiency (corresponding to the ratio of the upper bound of the data volume and energy consumption). As can be seen from the result, when the energy is low, the energy efficiency is also low. This suggests that the SU is unable to sense and access the optimal channel due to unknown channel statistical information. As the energy increases, the energy efficiency gradually approaches the optimal energy efficiency. The underlying reason is that as the SU samples various channels and obtains the estimated statistical information, it tends to sense and access the optimal channels, which validates the algorithm effectiveness.

[2]In reference [20], the maximum transmit power of the SU is 166.62 mW, and the maximum energy consumption during a duration of 1.5 ms is $2.5 * 10^{-4}$ J. Therefore, we assume that the energy consumption for channel selection, estimation, and ACK reception in the simulation is set as $2.5*10^{-4}$ J.

TABLE I
THE TRANSMIT POWER ON CHANNELS.

Channel	$u = 1$	$u = 2$
1	$[0.062, 0.35, 0.379, 0.208]$	$[0.183, 0.173, 0.095, 0.548]$
2	$[0.265, 0.356, 0.199, 0.179]$	$[0.070, 0.178, 0.238, 0.515]$
3	$[0.445, 0.209, 0.028, 0.318]$	$[0.246, 0.039, 0.309, 0.406]$
4	$[0.572, 0.234, 0.118, 0.076]$	$[0.195, 0.331, 0.452, 0.022]$
5	$[0.136, 0.335, 0.076, 0.453]$	$[0.452, 0.379, 0.064, 0.106]$
6	$[0.026, 0.151, 0.483, 0.339]$	$[0.036, 0.323, 0.328, 0.313]$
7	$[0.152, 0.365, 0.481, 0.003]$	$[0.349, 0.158, 0.133, 0.359]$
8	$[0.247, 0.288, 0.257, 0.208]$	$[0.294, 0.196, 0.429, 0.079]$
9	$[0.219, 0.425, 0.121, 0.235]$	$[0.287, 0.036, 0.386, 0.290]$
10	$[0.316, 0.135, 0.298, 0.251]$	$[0.424, 0.199, 0.284, 0.093]$
Channel	$u = 3$	$u = 4$
1	$[0.061, 0.174, 0.190, 0.575]$	$[0.097, 0.389, 0.285, 0.229]$
2	$[0.242, 0.327, 0.270, 0.160]$	$[0.013, 0.282, 0.268, 0.437]$
3	$[0.299, 0.091, 0.255, 0.355]$	$[0.188, 0.069, 0.164, 0.579]$
4	$[0.230, 0.095, 0.329, 0.346]$	$[0.359, 0.155, 0.134, 0.352]$
5	$[0.207, 0.281, 0.077, 0.435]$	$[0.205, 0.245, 0.281, 0.269]$
6	$[0.331, 0.051, 0.289, 0.329]$	$[0.154, 0.288, 0.158, 0.401]$
7	$[0.108, 0.597, 0.243, 0.052]$	$[0.077, 0.168, 0.177, 0.577]$
8	$[0.030, 0.241, 0.352, 0.377]$	$[0.292, 0.199, 0.396, 0.113]$
9	$[0.053, 0.312, 0.283, 0.352]$	$[0.242, 0.083, 0.655, 0.021]$
10	$[0.295, 0.258, 0.235, 0.212]$	$[0.273, 0.291, 0.244, 0.192]$
Channel	$u = 5$	$u = 6$
1	$[0.051, 0.0522, 0.115, 0.782]$	$[0.075, 0.166, 0.344, 0.415]$
2	$[0.103, 0.159, 0.088, 0.649]$	$[0.129, 0.171, 0.150, 0.549]$
3	$[0.105, 0.161, 0.164, 0.570]$	$[0.145, 0.184, 0.120, 0.551]$
4	$[0.190, 0.118, 0.193, 0.499]$	$[0.185, 0.369, 0.097, 0.349]$
5	$[0.055, 0.319, 0.232, 0.392]$	$[0.321, 0.292, 0.109, 0.279]$
6	$[0.811, 0.063, 0.074, 0.052]$	$[0.415, 0.019, 0.271, 0.295]$
7	$[0.365, 0.004, 0.157, 0.474]$	$[0.957, 0.019, 0.002, 0.022]$
8	$[0.372, 0.089, 0.256, 0.283]$	$[0.851, 0.064, 0.018, 0.067]$
9	$[0.270, 0.266, 0.219, 0.244]$	$[0.827, 0.023, 0.149, 0.001]$
10	$[0.261, 0.245, 0.433, 0.061]$	$[0.222, 0.412, 0.335, 0.031]$

B. The Channel Fading Model

In this subsection, we validate the algorithm performance under the Rayleigh and Ricean fading models, respectively. We consider there exists 10 channels, with each channel having a bandwidth of 1 MHz. The transmission rate requirement is 2 Mbps. The transmission distance is set 150 meters. The antenna numbers of transmission and reception are set 4 and 1, respectively. The sensing duration and transmission duration are respectively set as 5ms and 95ms. The energy ranges from 0 to 160 J. We refer to [20] that the the energy consumption for channel selection, estimation, and ACK reception in the simulation is set as $2.5 * 10^{-4}$ J. Other simulation settings are presented in Table II.

Figures 5 and 6 represent the algorithm performances under the Rayleigh and Ricean fading models, respectively. 'Opti-

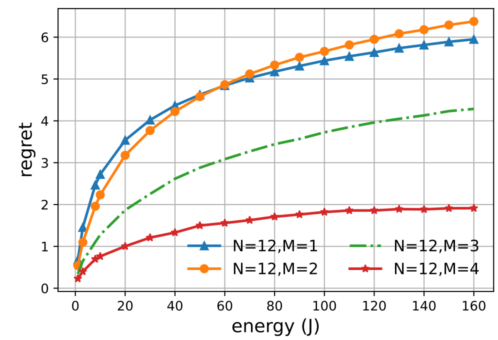

Fig. 3. The relationship between the data volume loss and energy.

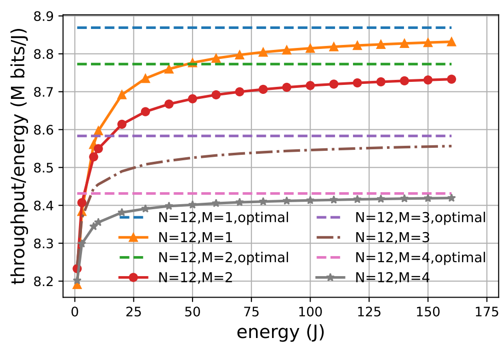

Fig. 4. The relationship between the energy efficiency and energy.

TABLE II
THE SIMULATION SETTINGS UNDER THE RAYLEIGH AND RICEAN CHANNEL MODEL

Parameter	Settings
Fading model	Rayleigh and Ricean (Ricean Factor = 3)
Pathloss model	$\frac{10^{-3.53}}{d^{3.76}}$
Rate requirement	2 Mbps
Bandwidth	1 MHz
Sensing power	110 mW
Transmission duration	90 ms
Channel number	10
Idle probability	$[0.972, 0.824, 0.742, 0.644, 0.528, 0.462, 0.341, 0.225, 0.109, 0.045]$
Monte Carlo number	30
Distance d	150 m
The antenna number of transmission and reception	4,1
Channel noise	$-104 \sim -95$ dBm
The energy	$0 \sim 160$ J
Sensing duration	5 ms
Energy threshold ε_{thres}	0 J
α,α_1	$\frac{1}{100}, \frac{1}{2}$

mal' represents ratio value of the upper bound of data volume and energy. As can be seen from the results, the energy efficiency with various M achieves the optimal energy efficiency as the energy increases. The underlying reason indicates that as the energy increases, the SU can gradually obtain the channel statistical information and tend to sense and access optimal channels, which validates the algorithm effectiveness.

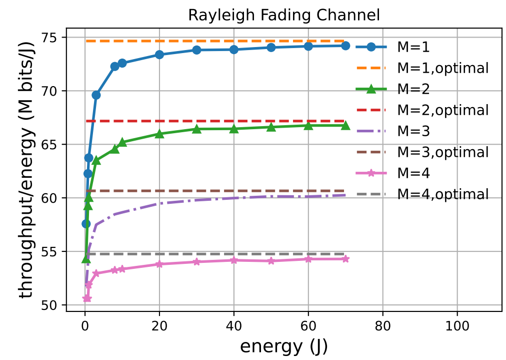

Fig. 5. The relationship between the data loss and energy.

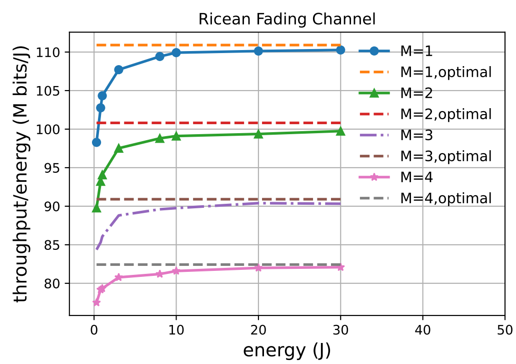

Fig. 6. The relationship between the energy efficiency and energy.

C. Performance Comparison of Various Algorithms

In this subsection, we perform simulations on various algorithms for the comparison on Rayleigh fading model. The channel selection number M is set to 1. The energy ranges from $0 \sim 50$ J. The monte carlo number is set to 5. Other simulation settings are as the same as Table II.

For comparison, we choose algorithm $\pi-$EC, the classic UCB, the $\varepsilon-$greedy algorithm, EABS-UCB, EABS-TS, c-UCB, and Budget-Limited $\varepsilon-$First algorithm.

In specific, in classic UCB, we assume that the SU does not take the energy consumption into consideration, computes the channel indexes by the average estimated throughput, and chooses the channel $\arg\max_{i \in \mathcal{I}_N} \left\{ \bar{\mu}_i^r(t) + \sqrt{\frac{\ln t}{2T_i(t)}} \right\}$ until the energy runs out.

In $\varepsilon-$greedy algorithm, we assume that the SU selects channels using $\pi-$EC with the probability of 0.9 and randomly selects channels with the probability of 0.1.

In EABS-UCB, the SU calculates indexes for all channels in each slot by $\bar{\mu}_i^r(t)+\sqrt{\frac{\ln t}{2T_i(t)}}-\frac{d}{\varepsilon(t)}$ and selects the channel with the largest index, where d represents the distance.

In EABS-TS, the SU calculates indexes for all channels in each slot by $\theta_i(t)-\frac{d}{\varepsilon(t)}$, $\theta_i(t)\sim\mathcal{N}(\bar{\mu}_i^r(t),\frac{1}{T_i(t)+1})$, where $\mathcal{N}(\bar{\mu}_i^r(t),\frac{1}{T_i(t)+1})$ represents the normal distribution with mean value $\bar{\mu}_i^r(t)$ and variance $\frac{1}{T_i(t)+1}$. Then, the SU selects the channel with the largest index.

In c-UCB, the SU calculates indexes for all channels in each slot by $\left(\bar{\mu}_i^r(t)+\sqrt{\frac{\alpha\ln t}{T_i(t)}}\right)/\bar{\mu}_i^c(t)$ with $\alpha=2$, and selects the channel with the largest index.

In Budget-Limited $\varepsilon-$First algorithm, the SU uses $\varepsilon\mathcal{E}$ energy to sense channels in a round-robbin manner. After $\varepsilon\mathcal{E}$ energy runs out, the SU uses $(1-\varepsilon)\mathcal{E}$ energy to continuously sense and access the channel $\arg\max\limits_{i\in\mathcal{I}_N}\left\{\frac{\bar{\mu}_i^r(t)}{\bar{\mu}_i^c(t)}\right\}$.

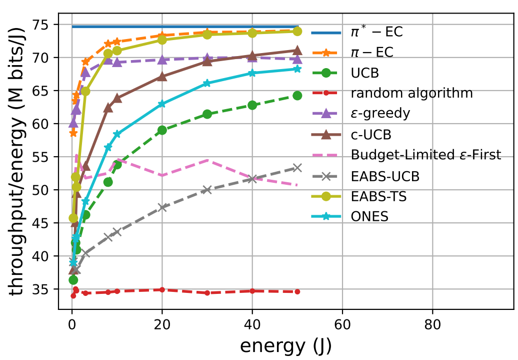

Fig. 7. The performance comparison of various algorithms

The simulation result is shown in Figure 7, where 'optimal' represents the ratio value of the upper bound of data volume and energy. As can be seen from the result, as the energy increases, the energy efficiency of $\pi-$EC and EABS-TS converges to the optimal energy efficiency, representing that the SU tends to sense and access the optimal channels. However, the convergence rate of $\pi-$EC outperforms that of EABS-TS. Compared with c-UCB, UCB and EABS-UCB, the energy efficiency of $\pi-$EC apparently outperforms others, representing that the SU driven by $\pi-$EC can learn the channel statistics, sense and access optimal channels more times. Moreover, we can see that Budget-Limited $\varepsilon-$First and random algorithms cannot converge to the optimal energy efficiency, indicating that the algorithms cannot be effective in our problem.

V. Conclusion

In this paper, we investigate the DSA for CR-IoT where the SU has the limited energy supply. The channel statistical information, including the idle probability and channel quality, is unknown to the SU. To obtain the largest expected cumulative data volume, we propose an algorithm based on the UCB for throughput-to-energy ratio to learn the channel statistics and continuously select the channels with the largest ratios. The theoretical analysis and simulations validate the effectiveness. In the further, multi-user networks will be extended.

Appendix A
Preliminary Conclusion

For the clarity of the subsequent proof, we first give the preliminary conclusion.

Lemma 4: For any time slots n, n_2 and $n_2\le n$, when $\frac{\bar{\mu}_{\mathcal{C}_j}^r(n_2)}{\bar{\mu}_{\mathcal{C}_j}^c(n_2)}\ge\frac{\mu_{\mathcal{C}_j}^r}{\mu_{\mathcal{C}_j}^c}+\sqrt{\frac{\alpha_1\ln n}{n_2}}$, there exist a constant $\alpha_0\ge 0$ enabling at least one of following events holds: $(a):\ \bar{\mu}_{\mathcal{C}_j}^r(n_2)\le\mu_{\mathcal{C}_j}^r-\sqrt{\frac{\alpha_0\ln n}{n_2}}$, $(b):\ \bar{\mu}_{\mathcal{C}_j}^c(n_2)\ge\mu_{\mathcal{C}_j}^c+\sqrt{\frac{\alpha_0\ln n}{n_2}}$.

Proof: We adopt the contradiction method for the prood. Assuming that neither of two events holds, then we have

$$\begin{aligned}&\frac{\mu_{\mathcal{C}_j}^r}{\mu_{\mathcal{C}_j}^c}-\frac{\bar{\mu}_{\mathcal{C}_j}^r(n_2)}{\bar{\mu}_{\mathcal{C}_j}^c(n_2)}=\\&\frac{(\mu_{\mathcal{C}_j}^r-\bar{\mu}_{\mathcal{C}_j}^r(n_2))\cdot\bar{\mu}_{\mathcal{C}_j}^c(n_2)-\bar{\mu}_{\mathcal{C}_j}^r(n_2)\cdot(\mu_{\mathcal{C}_j}^c-\bar{\mu}_{\mathcal{C}_j}^c(n_2))}{\mu_{\mathcal{C}_j}^c\bar{\mu}_{\mathcal{C}_j}^c(n_2)}\\&\ge\frac{\sqrt{\frac{\alpha_0\ln n}{n_2}}}{\mu_{\mathcal{C}_j}^c}+\frac{\sqrt{\frac{\alpha_0\ln n}{n_2}}}{\mu_{\mathcal{C}_j}^c}\cdot\frac{\bar{\mu}_{\mathcal{C}_j}^r(n_2)}{\bar{\mu}_{\mathcal{C}_j}^c(n_2)}\\&\ge\frac{\sqrt{\frac{\alpha_0\ln n}{n_2}}}{\mu_{\mathcal{C}_j}^c}+\frac{\sqrt{\frac{\alpha_0\ln n}{n_2}}}{\mu_{\mathcal{C}_j}^c}\cdot\Upsilon_{\min}>0,\end{aligned}\tag{17}$$

where $\Upsilon_{\min}=\min\limits_{n_2\in\mathcal{I}_n}\left\{\frac{\bar{\mu}_{\mathcal{C}_j}^r(n_2)}{\bar{\mu}_{\mathcal{C}_j}^c(n_2)}\right\}$. Since $\frac{\mu_{\mathcal{C}_j}^r}{\mu_{\mathcal{C}_j}^c}-\frac{\bar{\mu}_{\mathcal{C}_j}^r(n_2)}{\bar{\mu}_{\mathcal{C}_j}^c(n_2)}\le-\sqrt{\frac{\alpha_1\ln n}{n_2}}\le 0$, which contradicts the conclusion from Eq. (17), indicats that at least one of the two events must hold. Thus, the proof completes.

Appendix B
Proof of Lemma 1

In fact, since there exist $\binom{N}{M}$ channel sets, the expected data volume can be expressed as

$$\begin{aligned}&\sum_{m=1}^{\binom{N}{M}}\sum_{t=1}^{\infty}\mu_{\mathcal{C}_m}^r\cdot\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_m]\}\\&=\sum_{m=1}^{\binom{N}{M}}\sum_{t=1}^{\infty}\frac{\mu_{\mathcal{C}_m}^r}{\mu_{\mathcal{C}_m}^c}\mu_{\mathcal{C}_m}^c\cdot\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_m]\}\\&\le\frac{\mu_{\mathcal{C}_{\max}}^r}{\mu_{\mathcal{C}_{\max}}^c}\sum_{m=1}^{\binom{N}{M}}\sum_{t=1}^{\infty}\mu_{\mathcal{C}_m}^c\cdot\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_m]\},\end{aligned}\tag{18}$$

Since the expected cumulative energy consumption is less than the energy constraint $\mathcal{E}$, we thus have

$$\sum_{m=1}^{\binom{N}{M}}\sum_{t=1}^{\infty}\mu_{\mathcal{C}_m}^c\cdot\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_m]\}\le\mathcal{E}.\tag{19}$$

Substituting Eq. (19) into Eq. (18), we can finally include the conclusion that the upper bound for the expectation on the data volume with energy constraint $\mathcal{E}$ is $\frac{\mu_{\mathcal{C}_{\max}}^r}{\mu_{\mathcal{C}_{\max}}^c}\mathcal{E}$.

Appendix C
Proof of Lemma 2

To analyze the upper bound for term (A) in Eq. (13), we first analyze the bound of $\mathbb{E}_\tau\{\tau\}$. In the process of channel sensing and access, the overall expected energy consumption can be expressed as

$$\mathbb{E}_\tau\left\{\sum_{j=1}^{\binom{N}{M}}\sum_{t=1}^{\tau}\mu^c_{\mathcal{C}_j}\cdot\mathbb{E}\{\mathbb{I}[a(t)=\mathcal{C}_j]\}\right\} \tag{20}$$

$$=\mathbb{E}_\tau\left\{\sum_{t=1}^{\tau}\mu^c_{\mathcal{C}_{\max}}\left(1-\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\mathbb{P}\{a(t)=\mathcal{C}_j\}\right)\right.$$
$$\left.+\sum_{t=1}^{\tau}\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\mu^c_{\mathcal{C}_j}\mathbb{P}\{a(t)=\mathcal{C}_j\}\right\} \tag{21}$$

$$=\mathbb{E}_\tau\left\{\sum_{t=1}^{\tau}\mu^c_{\mathcal{C}_{\max}}\right\}+ \tag{22}$$
$$\mathbb{E}_\tau\left\{\sum_{t=1}^{\tau}\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\left(\mu^c_{\mathcal{C}_j}-\mu^c_{\mathcal{C}_{\max}}\right)\mathbb{P}\{a(t)=\mathcal{C}_j\}\right\}$$

$$=\mathbb{E}_\tau\{\tau\}\,\mu^c_{\mathcal{C}_{\max}}+ \tag{23}$$
$$\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\left(\mu^c_{\mathcal{C}_j}-\mu^c_{\mathcal{C}_{\max}}\right)\mathbb{E}_\tau\left\{\sum_{t=1}^{\tau}\mathbb{P}\{a(t)=\mathcal{C}_j\}\right\}$$

$$=\mathbb{E}_\tau\{\tau\}\,\mu^c_{\mathcal{C}_{\max}}+ \tag{24}$$
$$\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\left(\mu^c_{\mathcal{C}_j}-\mu^c_{\mathcal{C}_{\max}}\right)\mathbb{E}_\tau\left\{\mathbb{E}\left\{\sum_{t=1}^{\tau}\mathbb{I}[a(t)=\mathcal{C}_j]\right\}\right\}$$

$$=\mathbb{E}_\tau\{\tau\}\,\mu^c_{\mathcal{C}_{\max}}+\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\left(\mu^c_{\mathcal{C}_j}-\mu^c_{\mathcal{C}_{\max}}\right)\mathbb{E}_\tau\left\{\mathbb{E}\{T_{\mathcal{C}_j}(\tau)\}\right\}, \tag{25}$$

where Eq. (21) comes from the fact that

$$\mathbb{P}\{a(t)=\mathcal{C}_{\max}\}=1-\sum_{\mathcal{C}_j\neq\mathcal{C}_{\max}}\mathbb{P}\{a(t)=\mathcal{C}_j\}. \tag{26}$$

Then, we analyze $\mathbb{E}\left\{T_{\mathcal{C}_j}(\tau)\right\}$ in Eq. (25). Since τ is a random variable, we cannot directly obtain the expectation over $\mathbb{E}\left\{T_{\mathcal{C}_j}(\tau)\right\}$. Therefore, the analysis is divided into two steps: we first set a fixed time τ' with $\tau'\leq\mathbb{E}_\tau\{\tau\}$ and obtain the expectation for $\mathbb{E}\left\{T_{\mathcal{C}_j}(\tau')\right\}$. Then, we determine the bound for τ' with the expectation over τ. The detailed analysis is given as follows.

For any integer l and $j\in\mathcal{I}_{\binom{N}{M}}$, the expectation $\mathbb{E}\left\{T_{\mathcal{C}_j}(\tau')\right\}$ for channel set $\mathcal{C}_j$ during τ' slots can be

$$\mathbb{E}\{T_{\mathcal{C}_j}(\tau')\}=1+\sum_{n=\binom{N}{M}+1}^{\tau'}\mathbb{E}\{\mathbb{I}[a(n)=\mathcal{C}_j]\} \tag{27}$$

$$=1+\sum_{n=\binom{N}{M}+1}^{\tau'}\mathbb{E}\{\mathbb{I}[a(n)=\mathcal{C}_j,T_{\mathcal{C}_j}(n-1)<l]\}+$$
$$\mathbb{E}\{\mathbb{I}[a(n)=\mathcal{C}_j,T_{\mathcal{C}_j}(n-1)\geq l]\} \tag{28}$$

$$\leq l+\sum_{n=\binom{N}{M}+1}^{\tau'}\mathbb{E}\{\mathbb{I}[a(n)=\mathcal{C}_j,T_{\mathcal{C}_j}(n-1)\geq l]\} \tag{29}$$

$$\leq l+\sum_{n=\binom{N}{M}+1}^{\tau'}\mathbb{E}\left\{\mathbb{I}\left[g_{\mathcal{C}_{\max}}(n)\leq g_{\mathcal{C}_j}(n),T_{\mathcal{C}_j}(n-1)\geq l\right]\right\} \tag{30}$$

$$\leq l+\sum_{n=\binom{N}{M}+1}^{\tau'}\mathbb{E}\left\{\mathbb{I}\left[\min_{0<n_1\leq n}\{g_{\mathcal{C}_{\max}}(n_1)\}\right.\right. \tag{31}$$
$$\left.\left.\leq\max_{l\leq n_2\leq n}\left\{g_{\mathcal{C}_j}(n_2)\right\}\right]\right\}$$

$$\leq l+\sum_{n=\binom{N}{M}+1}^{\tau'}\sum_{n_1=1}^{n}\sum_{n_2=l}^{n}\mathbb{E}\{\mathbb{I}\left[g_{\mathcal{C}_{\max}}(n_1)\leq g_{\mathcal{C}_j}(n_2)\right]\} \tag{32}$$

$$\leq l+\sum_{n=1}^{\tau'}\sum_{n_1=1}^{n}\sum_{n_2=l}^{n} \tag{33}$$
$$\mathbb{E}\left\{\underbrace{\mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_{\max}}(n_1)}{\bar{\mu}^c_{\mathcal{C}_{\max}}(n_1)}+\sqrt{\frac{\alpha_1\ln n}{n_1}}\leq\frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)}+\sqrt{\frac{\alpha_1\ln n}{n_2}}\right]}_{(A)}\right\}.$$

where Eq. (28) originates from the fact that the sensing number of channel set $\mathcal{C}_j$ is larger than l and less than l in $n-1$ slots, respectively. Eq. (30) originates from the fact that when the SU chooses channel set $\mathcal{C}_j$ at slot n, we have $g_{\mathcal{C}_j}(n)\geq g_{\mathcal{C}_{\max}}(n)$. The term (A) in Eq. (33) represents that at least one of following events holds: (a) : $\mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)}\geq\frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}}+\sqrt{\frac{\alpha_1\ln n}{n_2}}\right]$, (b) : $\mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_{\max}}(n_1)}{\bar{\mu}^c_{\mathcal{C}_{\max}}(n_1)}+\sqrt{\frac{\alpha_1\ln n}{n_1}}\leq\frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}}\right]$ and (c) : $\mathbb{I}\left[\frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}}<\frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}}+2\sqrt{\frac{\alpha_1\ln n}{n_2}}\right]$.

In accordance with Lemma 4 in Appendix A, event (a) can be depicted as

$$\mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)}\geq\frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}}+\sqrt{\frac{\alpha_1\ln n}{n_2}}\right]$$
$$\leq\mathbb{I}\left[\bar{\mu}^r_{\mathcal{C}_j}(n_2)\leq\mu^r_{\mathcal{C}_j}-\sqrt{\frac{\alpha_0\ln n}{n_2}}\right] \tag{34}$$
$$+\,\mathbb{I}\left[\bar{\mu}^c_{\mathcal{C}_j}(n_2)\geq\mu^c_{\mathcal{C}_j}+\sqrt{\frac{\alpha_0\ln n}{n_2}}\right], \tag{35}$$

In accordance with CH inequality, the expectation over Eq. (35) can be

$$\mathbb{E}\left\{\mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)}\geq\frac{\mu^r_{\mathcal{C}_j}}{\mu^c_j}+\sqrt{\frac{\alpha_1\ln n}{n_2}}\right]\right\}$$
$$\leq\mathbb{P}\left\{\bar{\mu}^r_{\mathcal{C}_j}(n_2)\leq\mu^r_{\mathcal{C}_j}-\sqrt{\frac{\alpha_0\ln n}{n_2}}\right\} \tag{36}$$
$$+\,\mathbb{P}\left\{\bar{\mu}^c_{\mathcal{C}_j}(n_2)\geq\mu^c_{\mathcal{C}_j}+\sqrt{\frac{\alpha_0\ln n}{n_2}}\right\} \tag{37}$$
$$\leq 2n^{-2\alpha_0}\overset{\alpha_0=2}{=}2n^{-4}. \tag{38}$$

The above proof also holds for event (b). Moreover, when $l \geq \left\lceil \frac{4\alpha_1 \ln n}{d_{\min}^2} \right\rceil$ where $d_{\min} = \min_{\mathcal{C}_j \neq \mathcal{C}_{\max}} \left\{ \left| \frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}} - \frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}} \right| \right\}$, the event (c) does not hold.

In conclusion, when $l \geq \left\lceil \frac{4\alpha_1 \ln n}{d_{\min}^2} \right\rceil$, the upper bound for Eq. (33) can be

$$\begin{aligned}
&\mathbb{E}\{T_{\mathcal{C}_j}(\tau')\} \leq l \\
&+ \sum_{n=1}^{\tau'} \sum_{n_1=1}^{n} \sum_{n_2=l}^{n} \mathbb{E}\left\{ \mathbb{I}\left[\frac{\bar{\mu}^r_{\mathcal{C}_{\max}}(n_1)}{\bar{\mu}^c_{\mathcal{C}_{\max}}(n_1)} + \sqrt{\frac{\alpha_1 \ln n}{n_1}} \right.\right. \\
&\qquad\qquad \left.\left. \leq \frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)} + \sqrt{\frac{\alpha_1 \ln n}{n_2}} \right] \right\}
\end{aligned} \tag{39}$$

$$\begin{aligned}
&\leq l + \sum_{n=1}^{\tau'} \sum_{n_1=1}^{n} \sum_{n_2=l}^{n} \mathbb{P}\left\{ \frac{\bar{\mu}^r_{\mathcal{C}_j}(n_2)}{\bar{\mu}^c_{\mathcal{C}_j}(n_2)} \geq \frac{\mu^r_{\mathcal{C}_j}}{\mu^c_j} + \sqrt{\frac{\alpha_1 \ln n}{n_2}} \right\} + \\
&\quad \mathbb{P}\left\{ \frac{\bar{\mu}^r_{\mathcal{C}_{\max}}(n_1)}{\bar{\mu}^c_{\mathcal{C}_{\max}}(n_1)} + \sqrt{\frac{\alpha_1 \ln n}{n_1}} \leq \frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}} \right\} + \\
&\quad \mathbb{P}\left\{ \frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}} < \frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}} + 2\sqrt{\frac{\alpha_1 \ln n}{n_2}} \right\}
\end{aligned} \tag{40}$$

$$\leq l + \sum_{n=1}^{\infty} n^2 \left(4n^{-4} + \mathbb{P}\left\{ \frac{\mu^r_{\mathcal{C}_{\max}}}{\mu^c_{\mathcal{C}_{\max}}} < \frac{\mu^r_{\mathcal{C}_j}}{\mu^c_{\mathcal{C}_j}} + 2\sqrt{\frac{\alpha_1 \ln n}{n_2}} \right\} \right) \tag{41}$$

$$\leq \frac{4\alpha_1 \ln \tau'}{d_{\min}^2} + 1 + \sum_{n=1}^{\infty} 4n^{-2} \leq \frac{4\alpha_1 \ln \tau'}{d_{\min}^2} + 1 + \frac{2\pi^2}{3}. \tag{42}$$

Since the energy constraint is $\mathcal{E}$ and the minimum energy consumption of channel sensing and access is $c_{\min}$, the expectation on stopping time can be

$$\tau' \leq \mathbb{E}_\tau\{\tau\} \leq \frac{\mathcal{E}}{Mc_{\min}}. \tag{43}$$

Substituting Eq. (43) into Eq. (42), we have

$$\mathbb{E}\{T_{\mathcal{C}_j}(\tau')\} \leq \frac{4\alpha_1 \ln \tau'}{d_{\min}^2} + 1 + \frac{2\pi^2}{3} \tag{44}$$

$$\leq \frac{4\alpha_1}{d_{\min}^2} \ln\left(\frac{\mathcal{E}}{M\mu^c_{\min}} \right) + 1 + \frac{2\pi^2}{3}. \tag{45}$$

Substituting Eq. (45) into Eq. (25), we have

$$\begin{aligned}
&\mathbb{E}_\tau \left\{ \sum_{j=1}^{\binom{N}{M}} \sum_{t=1}^{\tau} \mu^c_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \right\} \\
&= \mathbb{E}_\tau\left\{\tau\right\} \mu^c_{\mathcal{C}_{\max}} + \sum_{\mathcal{C}_j \neq \mathcal{C}_{\max}} \left(\mu^c_{\mathcal{C}_j} - \mu^c_{\mathcal{C}_{\max}} \right) \mathbb{E}_\tau\left\{ \mathbb{E}\{T_{\mathcal{C}_j}(\tau)\} \right\} \\
&\leq \mathbb{E}_\tau\left\{\tau\right\} \mu^c_{\mathcal{C}_{\max}} + \sum_{\mathcal{C}_j \in \mathcal{C}^\dagger} \Delta_{\max} \mathbb{E}_\tau\left\{ \mathbb{E}\{T_{\mathcal{C}_j}(\tau)\} \right\} \\
&\leq \mathbb{E}_\tau\left\{\tau\right\} \mu^c_{\mathcal{C}_{\max}} \\
&\quad + \sum_{\mathcal{C}_j \in \mathcal{C}^\dagger} \Delta_{\max} \left(\frac{4\alpha_1}{d_{\min}^2} \ln\left(\frac{\mathcal{E}}{M\mu^c_{\min}} \right) + 1 + \frac{2\pi^2}{3} \right),
\end{aligned} \tag{46}$$

where

$$\Delta_{\max} = \max_{j \in \mathcal{I}_{\binom{N}{M}}} \left\{ \left| \mu^c_{\mathcal{C}_j} - \mu^c_{\mathcal{C}_{\max}} \right| \right\}, \tag{47}$$

$$\mathcal{C}^\dagger = \left\{ \mathcal{C}_j : \mu^c_{\mathcal{C}_j} > \mu^c_{\mathcal{C}_{\max}}, j \in \mathcal{I}_{\binom{N}{M}} \right\}. \tag{48}$$

Noting that the residual energy is at most ε_{thres}, we have

$$\mathcal{E} - \varepsilon_{thres} \leq \mathbb{E}_\tau \left\{ \sum_{j=1}^{\binom{N}{M}} \sum_{t=1}^{\tau} \mu^c_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \right\}. \tag{49}$$

Combining Eq. (46) with Eq. (49), we have

$$\begin{aligned}
&\mathbb{E}_\tau\left\{\tau\right\} \geq \\
&\frac{\mathcal{E} - \varepsilon_{thres} - \sum_{\mathcal{C}_j \in \mathcal{C}^\dagger} \Delta_{\max} \left(\frac{4\alpha_1}{d_{\min}^2} \ln\left(\frac{\mathcal{E}}{M\mu^c_{\min}} \right) + 1 + \frac{2\pi^2}{3} \right)}{\mu^c_{\mathcal{C}_{\max}}}.
\end{aligned} \tag{50}$$

Then, substituting Eq. (50) into term (A) in Eq. (13), we can finally include the conclusion. ■

Appendix D
Proof of Lemma 3

In fact, the term (B) in Eq. (13) can be expressed as

$$\mathbb{E}_\tau \left\{ \tau \cdot \mu^r_{\mathcal{C}_{\max}} - \sum_{j=1}^{\binom{N}{M}} \sum_{t=1}^{\tau} \mu^r_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \right\} \tag{51}$$

$$= \mathbb{E}_\tau \left\{ \sum_{t=1}^{\tau} \left(\mu^r_{\mathcal{C}_{\max}} - \sum_{j=1}^{\binom{N}{M}} \mu^r_{\mathcal{C}_j} \cdot \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \right) \right\} \tag{52}$$

$$\begin{aligned}
= \mathbb{E}_\tau \Bigg\{ \sum_{t=1}^{\tau} \Bigg(&\sum_{j=1}^{\binom{N}{M}} \mu^r_{\mathcal{C}_{\max}} \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} - \\
&\sum_{j=1}^{\binom{N}{M}} \mu^r_{\mathcal{C}_j} \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \Bigg) \Bigg\}
\end{aligned} \tag{53}$$

$$= \mathbb{E}_\tau \left\{ \sum_{t=1}^{\tau} \sum_{j=1}^{\binom{N}{M}} (\mu^r_{\mathcal{C}_{\max}} - \mu^r_{\mathcal{C}_j}) \mathbb{E}\{\mathbb{I}[a(t) = \mathcal{C}_j]\} \right\} \tag{54}$$

$$= \mathbb{E}_\tau \left\{ \sum_{j=1}^{\binom{N}{M}} (\mu^r_{\mathcal{C}_{\max}} - \mu^r_{\mathcal{C}_j}) \mathbb{E}\left\{ \sum_{t=1}^{\tau} \mathbb{I}[a(t) = \mathcal{C}_j] \right\} \right\} \tag{55}$$

$$\leq \Delta_{\max} \binom{N}{M} \left(\frac{4\alpha_1}{d_{\min}^2} \ln\left(\frac{\mathcal{E}}{M\mu^c_{\min}} \right) + 1 + \frac{2\pi^2}{3} \right), \tag{56}$$

where Eq. (56) originates from Eqs. (27)-(45). We finally include the conclusion. ■

References

[1] N. Javaid, A. Sher, H. Nasir, and N. Guizani, "Intelligence in iot-based 5g networks: Opportunities and challenges," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 94–100, 2018.

[2] D. Mishra, N. R. Zema, and E. Natalizio, "A high-end iot devices framework to foster beyond-connectivity capabilities in 5g/b5g architecture," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 55–61, 2021.

[3] A. J. Onumanyi, A. M. Abu-Mahfouz, and G. P. Hancke, "Cognitive radio in low power wide area network for iot applications: Recent approaches, benefits and challenges," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7489–7498, 2020.

[4] F. Li, K.-Y. Lam, X. Li, Z. Sheng, J. Hua, and L. Wang, “Advances and emerging challenges in cognitive internet-of-things,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5489–5496, 2020.

[5] P. Zhou, J. Xu, W. Wang, C. Jiang, K. Wang, and J. Hu, “Human-behavior and qoe-aware dynamic channel allocation for 5g networks: A latent contextual bandit learning approach,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 436–451, 2020.

[6] L. Lai, H. E. Gamal, H. Jiang, and P. Vincent, “Cognitive medium access: Exploration, exploitation, and competition,” *IEEE Transactions on Mobile Computing*, vol. 10, no. 2, pp. 239–253, 2011.

[7] Z. Xu, Z. Zhang, S. Wang, A. Jolfaei, A. K. Bashir, Y. Yan, and S. Mumtaz, “Decentralized opportunistic channel access in crns using big-data driven learning algorithm,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 1, pp. 57–69, 2021.

[8] W. Wang, A. Leshem, D. Niyato, and Z. Han, “Decentralized learning for channel allocation in iot networks over unlicensed bandwidth as a contextual multi-player multi-armed bandit game,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3162–3178, 2022.

[9] A. Magesh and V. V. Veeravalli, “Decentralized heterogeneous multi-player multi-armed bandits with non-zero rewards on collisions,” *IEEE Transactions on Information Theory*, vol. 68, no. 4, pp. 2622–2634, 2022.

[10] C. Fan, L. Li, M.-M. Zhao, A. Liu, and M.-J. Zhao, “Communication and energy-constrained neighbor selection for distributed cooperative localization,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 4158–4172, 2023.

[11] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, “Ai models for green communications towards 6g,” *IEEE Communications Surveys* & *Tutorials*, vol. 24, no. 1, pp. 210–247, 2022.

[12] Z. Xu, Z. Zhang, S. Wang, Y. Yan, and Q. Cheng, “Resource-constraint network selection for iot under the unknown and dynamic heterogeneous wireless environment,” *IEEE Internet of Things Journal*, vol. 10, no. 14, pp. 12 322–12 337, 2023.

[13] X. Mu, Y. Liu, L. Guo, J. Lin, and Z. Ding, “Energy-constrained uav data collection systems: Noma and oma,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6898–6912, 2021.

[14] M. Patnaik, V. Kamakoti, V. Matyáš, and V. Řchák, “Prolemus: A proactive learning-based mac protocol against puea and ssdf attacks in energy constrained cognitive radio networks,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 2, pp. 400–412, 2019.

[15] M. M. Fouda, S. Hashima, S. Sakib, Z. M. Fadlullah, K. Hatano, and X. Shen, “Optimal channel selection in hybrid rf/vlc networks: A multi-armed bandit approach,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6853–6858, 2022.

[16] Y. Xia, T. Qin, W. Ding, H. Li, X. Zhang, N. Yu, and T. Liu, “Finite budget analysis of multi-armed bandit problems,” *Neurocomputing*, vol. 258, pp. 13–29, 2017.

[17] L. Tran-Thanh, A. C. Chapman, E. M. de Cote, A. Rogers, and N. R. Jennings, “Epsilon-first policies for budget-limited multi-armed bandits,” in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010, pp. 1211–1216.

[18] Y. Chen, Q. Zhao, and A. Swami, “Distributed spectrum sensing and access in cognitive radio networks with energy constraint,” *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 783–797, 2009.

[19] “https://github.com/davidloveshet/thesis-code.”

[20] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, “Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, 2011.