Channel-Occupation-Aware Resource Allocation in LoRa Networks: a DQN-and-Optimization-Aided Approach

Zhen Qin[1], Jinming Li[1] and Bo Gu[1,2]0
[1]School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China
[2]Guangdong Provincial Key Laboratory of Fire Science and
Intelligent Emergency Technology, Guangzhou 510006, China
E-mail: {qinzh23, lijm89}@mail2.sysu.edu.cn, gubo@mail.sysu.edu.cn

2022 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-4266-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/WCNC51071.2022.9771911

***Abstract*—Long Range (LoRa) technology which provides low power consumption and long transmission range is regarded as one of the key technologies for industrial Internet of Things (IIoTs). In this article, we investigate an energy efficiency (EE) maximization problem through the joint allocation of spreading factors (SFs), channels (CHs) and transmit power. Then, the optimal solution is derived by decomposing the problem into two stages: (i) a deep Q-Network (DQN) is trained to generate the allocation of CH/SF based on the channel state information (CSI) collected from end-devices (EDs); (ii) given the CH/SF allocation, the optimal transmission power is then determined by solving a convex optimization problem. Simulation results demonstrate that our algorithm is superior to the random CH/SF and power selecting method, and achieves a near-optimal performance.**

***Index Terms*—Deep Reinforcement Learning (DRL), energy efficiency (EE), Long Range (LoRa), joint resource allocation.**

I. Introduction

WITH the expansion of the industrial Internet-of-Things (IIoTs), networks are facing challenges of limited coverage and insufficient capacity. Fortunately, Low-Power-Wide-Area (LPWA), which has the characteristics of high-capacity and low energy consumption, has the potential to cope with these challenges. Compared with traditional technologies for IIoTs such as Wi-Fi, LPWA can better adapt to the current energy consumption, range, and connectivity requirements. In LPWA, Long Range (LoRa) technology is widely concerned as it works in unlicensed frequency bands and satisfies the requirements of low power consumption and longer transmission range in IIoTs.

In a typical LoRa network, end-devices (EDs) and gateway are connected wirelessly. The LoRa gateway exchanges data with the server through the Internet, forming a star topology. Through chirp spread spectrum (CSS) and assigning orthogonal spreading factors (SFs) to different EDs, channel multiplexing can be realized, which increases the network capacity. It is worth noticing that different SFs have different performance. In particular, when the SF is larger, the transmission range becomes longer, the receiver sensitivity becomes higher, the data rate becomes lower, and vice versa [1].

It is hard to distinguish whether the channel occupancy is caused by noise or real data transmission due to the adopting of CSS. Thus, the Channel Activity Detection (CAD) has been introduced in LoRa networks. It can detect the existence of preamble to determine whether the channel is occupied. Such information is helpful when assigning channels (CHs) and SFs. It has been proved in [2] that energy consumed in CAD process is much less than transmission process. Therefore, adding CAD in the design of allocation scheme will not increase the energy consumption of EDs.

Considering the goal of energy conservation, MAC layer of LoRa adopts ALOHA instead of CSMA. As a random method, ALOHA cannot cope with high-pressure scenario, which is manifested in network congestion and system performance degradation [3]. Many resource allocation methods have been proposed to reduce the collisions and improve channel resource utilization. Specifically, authors in [4] studied the SF and transmit power allocation problems based on the adaptive data rate (ADR). Kumari *et al.* [5] allocated the best time of SF using for each ED by establishing a Nash equilibrium game. In addition, many joint allocation schemes have been proposed.

However, in wireless system, the above-mentioned joint allocation problems are NP-hard, and it is difficult to obtain the optimal solutions. Therefore, traditional methods [6], [7] may simplify the problems, and obtain the suboptimal solution through a large number of iterations, which limits the performance of the network. Fortunately, deep reinforcement learning (DRL) performs well in solving such problems [8]. Gu *et al.* [9] and Chen *et al.* [10] adopted DRL to solve the joint optimization of device-to-device (D2D) communication and mobile edge computing, respectively.

In this article, we study the uplink energy efficient transmission framework in LoRa network, realizing the maximization of energy efficiency (EE). The deep Q-network (DQN) [11] is firstly adopted to solve the joint allocation problem of

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1713800 and in part by the National Science Foundation of China (NSFC) under Grand U20A20175. *(Corresponding author: Bo Gu.)*

SF and CH in uplink LoRa network. Then, we propose an optimization-based power allocation framework, which is based on the CH/SF allocation strategy given by DQN. The main contributions are given as follows.

1) We formulate the joint optimization problem of uplink LoRa network as the maximization problem of EE. The goal is to maximize the system EE through the joint allocation of CH/SF and transmit power.
2) The proposed algorithm contains a DRL module and an optimization module. DRL module makes the algorithm converge quickly and adapts to the dynamic wireless communication environment. The optimization module can ignore the complex communication environment and give the optimal power allocation scheme.
3) We take the channel occupancy based on CAD as one of the basis for CH/SF allocation. The gateway follows a merging process to merge the CAD results from EDs into a channel occupancy matrix.
4) Simulation results show that the proposed algorithm outperforms existing methods in terms of improving system EE.

The remainder of this paper is organized as follows. Section II introduces the system model and problem formulation of uplink LoRa network. Section III presents the DQN and optimization based resources allocation algorithm. Section VI gives the simulation results. Section VII is the conclusion of this article.

II. System Model And Problem Formulation

A. System Model

We focus on the uplink transmission in LoRa networks, in which M EDs connect to one LoRa gateway through N channels. Different EDs can share the same channel simultaneously by using different SFs. We assume that all EDs and the gateway are equipped with a single antenna and support half duplex communication. We denote the channel and EDs sets as $\mathcal{N}=\{1,2,...,N\}$, $\mathcal{M}=\{1,2,...,M\}$, respectively. The bandwidth of channel n is denoted as BW_n.

Multiple EDs can access to one channel by choosing different SFs. The SF allocated to ED m is represented by $k_m \in \{7,...,12\}$. Moreover, we use a binary index $b_{m,n}$ to indicate whether ED m is accessing to channel n, where $b_{m,n}=1$ if ED m is allocated to channel n, $b_{m,n}=0$ otherwise. The transmit power of ED m in channel n is denoted by $p_{m,n}$. We assume that the LoRa gateway can get the perfect channel state information (CSI) contained in the uplink data frames from EDs.

Formally, the received signal-to-interference-plus-noise ratio (SINR) at the LoRa gateway for ED m on channel n can be expressed as :

$$S_{m,n,k_m}=\frac{b_{m,n}p_{m,n}h_{m,n}}{\sum_{i=1,i\neq m}^{M}b_{i,n}p_{i,n}h_{i,n}\omega_{k_i,k_m}+z_n} \tag{1}$$

where $h_{m,n}$ is the channel gain of ED m on channel n, $\omega_{k_i,k_m}=1$ or 0 denotes whether $k_i=k_m$ or not, and z_n is the additive white Gaussian noise (AWGN) which obeys $z_n \in \mathcal{CN}(0,\sigma_m^2)$. The first term in the denominator represents the co-SF interference [7] caused by other EDs using the same CH/SF.

TABLE I: Capture threshold and distance ranges of different SFs.

SF	**Receive Sensitivity** [dBm]	**SINR Threshold** [dB]	**Distance** [km]
7	-123	-7.5	2
8	-126	-10	4
9	-129	-12.5	6
10	-132	-15	8
11	-134.5	-17.5	10
12	-137	-20	12

According to Shannon theorem, the data rate of ED m on channel n is expressed as

$$R_{m,n,k_m}=BW_n\log_2(1+S_{m,n,k_m}) \tag{2}$$

Therefore, EE of a single ED can be denoted as the ratio of the data rate to the power consumption, which can be characterized as

$$\eta_m=\frac{\sum_{n=1}^{N}b_{m,n}R_{m,n,k_m}}{\sum_{n=1}^{N}b_{m,n}p_{m,n}+C} \tag{3}$$

the parameter C denotes the fixed power consumption of standby, CAD, or other operations.

B. Problem Formulation

The problem is formulated to maximize the system EE through allocating the binary index $b_{m,n}$, choosing an appropriate k_m and assigning power $p_{m,n}$. Therefore, the problem can be formulated as

$$\begin{aligned}\max_{\{b_{m,n},k_m,p_{m,n}\}} \quad & \sum_{m=1}^{M}\eta_m\\ \text{s.t.}\quad & (a)\quad p_{\text{min}}\le p_{m,n}\le p_{\text{max}},\\ & (b)\quad b_{m,n}\in\{0,1\},\\ & (c)\quad \sum_{n=1}^{N}b_{m,n}\le 1,\forall m\in\mathcal{M},\\ & (d)\quad \sum_{m=1}^{M}b_{m,n}\le \chi_{max},\forall n\in\mathcal{N},\\ & (e)\quad S_{m,n,k_m}\ge\theta_{SF}^{(k_m)},\forall m\in\mathcal{M},\forall n\in\mathcal{N}.\end{aligned} \tag{4}$$

where constraint (a) limits the range of transmit power. Constraint (b) indicates whether CH n is allocated to ED m or not. Constraint (c) implies that each ED can access at most one channel. Although we can achieve channel multiplexing by using different SFs, the number of EDs is upper bounded by χ_{max}, which is constrained in (d). Constraint (e) is the SF selecting constraint, ensuring that the SINR of the ED is above the SINR threshold in table I.

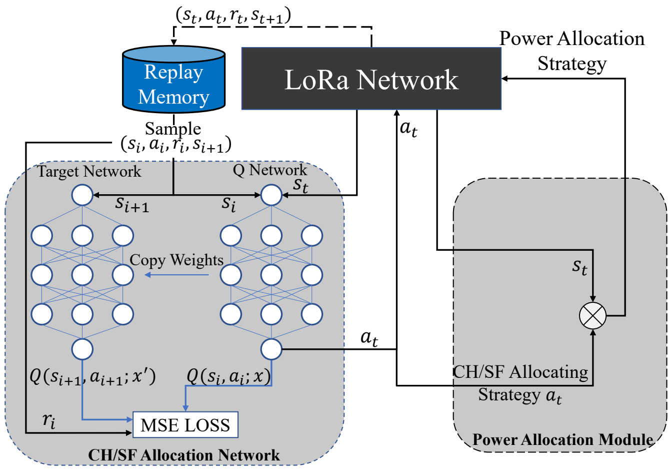


Fig. 1: Framework of JADO

Noting that the formulated problem is NP-hard due to the nonconvex optimization objective and linear inequalities in the constraints. Although we can calculate the EE of all possible allocation strategies to get the global optimal solution, it is unrealistic to use this exhaustive search in real scenes due to the high consumption and low efficiency. With the above considerations, we propose a joint resource allocation framework based on DRL and optimization (JADO). As shown in Fig. 1, we decompose the whole optimization process into two stages. In the first stage, a DQN is trained to generate the optimal allocation of CH/SF based on the CSI collected from EDs; in the second stage, given the CH/SF allocation, the optimal transmission power is then determined by solving a convex optimization problem.

III. DRL-Based CH/SF Allocation Framework

We assume that the LoRa gateway is a learning agent that interacts with its operating environment to learn the optimal CH/SF allocation policy. The gateway obtains CSI from the uplink frames of EDs and takes actions based on CSI. Through continuous interaction with the LoRa network, the gateway will update the allocation strategy.

A. Overview of Reinforcement Learning

In this part, we introduce the basic technology of JADO, which is known as reinforcement learning (RL). The goal of RL is to maximize the long-term cumulative reward which can be expressed as

$$G_t = R_{t+1} + \gamma R_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{5}$$

where γ is the discount rate. Specifically, in RL, agent selects an action a_t based on the observed state s_t at time slot t. The environment returns a feedback to the agent after the action is taken and then move to the next state s_{t+1}. By repeating the above process, agent stores tuples like $\{s_i, a_i, r_i, s_{i+1}\}$ as experience samples, which can be used to update the policy $\pi(a_t|s_t)$.

In Q-learning algorithm, Q is the action-utility function and is used to evaluate quality of an action, which is defined as

$$Q_\pi(s, a) = E_\pi[R_{t+1} + \gamma Q_\pi(s_{t+1}, a_{t+1})|s_t = s, a_t = a] \tag{6}$$

when γ is close to 1, the agent will focus on the value of the future states and it is opposite when γ closes to 0. The idea of Q-learning is to obtain the optimal action-utility function $Q^*(s, a)$, which is defined as

$$Q^*(s, a) = E[R_{t+1} + \gamma \max_{a'} Q^*(s', a')|s_t = s, a_t = a] \tag{7}$$

where s' is the following state of s_t after taking the action a_t. The above equation is derived from the Bellman equation. In Q-learning, the agent maintains a Q-table which maps the relationship between action, state and Q value. The value in this table can be updated by

$$\begin{aligned} &Q(s_t, a_t) \\ &\quad \leftarrow Q(s_t, a_t) + \alpha[R_{t+1} + \gamma \max_{a'} Q(s', a') - Q(s_t, a_t)] \end{aligned} \tag{8}$$

where α is the learning rate and $\alpha \in [0, 1]$. However, when the state space is too large, the estimation of Q-table will take a lot of time. To solve this problem, DQN adopts convolutional neural networks to approximate the Q value. The detail of DQN will be introduced in Section V.

To avoid getting the local optimal solution, we adopt ϵ-greedy algorithm. In each state, the agent will explore with probability ϵ, which is to select a random action, or select the action which has the largest Q value with probability $1 - \epsilon$. The process of ϵ-greedy algorithm can be described as

$$a_t = \begin{cases} \text{random} & \text{with probability } \epsilon \\ \arg\max_{\mathrm{a}} \ \mathrm{Q(s_t, a)} & \text{with probability } 1 - \epsilon \end{cases} \tag{9}$$

B. DRL-Based CH/SF Allocating Framework

In our proposed framework, EDs move to a random location in the beginning of each time slot and remain fixed in this time slot. As shown in Fig. 1, given the environment information, the learning agent gives a CH/SF allocation strategy and send it to the optimization module for calculating the optimal transmit power. Given the CH/SF/P allocation strategy, the agent broadcasts it to all EDs. EDs select the CH/SF/P based on the most recent received strategy. Specifically, the setting of RL is as follows.

States : To ensure the agent making reasonable decisions, the state should contain information about the communication environment. Thus, the state space is defined as

1) $\boldsymbol{g} = \{g_{m,n}\}_{m \in \mathcal{M}, n \in \mathcal{N}}$*:* The set of channel gains, where $g_{m,n}$ represents the channel gain between ED m and the gateway on channel n.

2) $\boldsymbol{d} = \{d_m\}_{m \in \mathcal{M}}$*:* The distances from EDs to the gateway.

3) $\boldsymbol{\phi} = \{\phi_{n,k}\}_{n\in\mathcal{N},k\in\{7,...,12\}}$*:* The set of global historical occupancy rate of each CH/SF pair. In particular, ED m stores CAD result of each CH/SF pair in a local channel occupancy matrix $\Psi^{(m)} \in R^{N\times 6}$, the elements in $\Psi^{(m)}$ are calculated by

$$\psi_{m,n,k} = \frac{CAD^{\text{occupied}}_{m,n,k}}{CAD^{\text{total}}_{m,n,k}} \tag{10}$$

where the denominator is the number of CAD that performed on channel n and SF k; the numerator is the occupied number of CAD on channel n and SF k. The matrix will be transmitted to the gateway in uplink frames. Then the gateway merges the received matrices to obtain the global occupancy matrix $\Phi \in R^{N\times 6}$, elements in the global occupancy matrix is calculated by

$$\phi_{m,n,k} = \frac{1}{M}\sum_{m=1}^{M}\psi_{m,n,k} \tag{11}$$

The joint state of the system can be denoted as: $\boldsymbol{s} = \{[\boldsymbol{g}],[\boldsymbol{d}],[\boldsymbol{\phi}]\}$.

Action : After receiving the state information, agent generates the CH/SF assignment for EDs. The joint allocation of CH/SF could reduce collision, and ensure that the SINR is above the threshold in table I. The action space include

1) $\boldsymbol{b} = \{b_{m,n}\}_{m\in\mathcal{M},n\in\mathcal{N}}$*:* Whether the ED m is assigned to channel n.

2) $\boldsymbol{k} = \{k_m\}_{m\in\mathcal{M}}$*:* The SF allocation decision and $k_m \in \{7, 8, ..., 12\}$.

The joint action of the system can be denoted as: $\boldsymbol{a} = \{[\boldsymbol{b}],[\boldsymbol{k}]\}$.

Reward :As our goal is to maximize the EE of the system, making EE be the reward can intuitively reflect the current performance of the system and the quality of the selected actions. From Eq.(3) Eq.(4), we can express the reward as

$$r = \sum_{m=1}^{M}\eta_m, \quad S_{m,n} \geq \theta^{(k)}_{SF},\ \forall m \in \mathcal{M} \tag{12}$$

IV. Optimization-Based Power Allocation Framework

With the CH/SF allocation strategy from the DRL module, we can further allocate transmission power for EDs. Specifically, given CH/SF allocation, Eq.(4) can be expressed as

$$\begin{aligned} \max_{p_m} \quad & \sum_{m=1}^{M}\eta_m \\ \text{s.t.} \quad & p_{\min} \leq p_m \leq p_{\max}, \end{aligned} \tag{13}$$

Noting that with a certain channel selection, $p_{m,n}$ can be simplified to p_m. Since using the same CH/SF will cause serious co-SF interference, SINR cannot meet the reception threshold in Table I. Thus, DQN has to avoid assigning the same CH/SF to different EDs. With the above consideration, we think the interference term of denominator in Eq.(1) can be omitted, and Eq.(13) can be further rewritten as

$$\begin{aligned} \max_{p_m} \quad & \sum_{m=1}^{M}\frac{BW_n\log_2(1+\frac{h_{m,n}}{z_n}p_m)}{p_m + C} \\ \text{s.t.} \quad & p_{\min} \leq p_m \leq p_{\max}. \end{aligned} \tag{14}$$

Algorithm 1 Training Process of DQN-and-Optimization-Based Joint Resource Allocation

Initialize:
- Initialize $Q(s, a; \boldsymbol{\theta})$ with weight $\boldsymbol{\theta}$.
- Initialize the target network with weight $\boldsymbol{\theta}' = \boldsymbol{\theta}$.
- Initialize replay buffer $\mathcal{D}$.
- Initialize the proposed model.
- Start environment simulator.

for episode $t = 1, 2, \cdots, W$ **do**
- Receive initial observation state.
- $m \leftarrow 1, idle$.
- **while** $m \leq M$ **do**
  - After observing the state $o^{(t)}_m$, agent selects action $a^{(t)}_m$ based on the probability ϵ according to Eq.(9).
  - Pass the action $a^{(t)}_m$ into the optimization module, calculate an optimal power ${p^{(t)}_m}^*$, and then pass it to the ED for allocation.
  - Environment feeds back reward $r^{(t)}_m$ and updates next state s_{m+1}.
  - Agent observes reward $r^{(t)}_m$ and next state $s^{(t+1)}_m$.
  - Store transition $(s^{(t)}_m, a^{(t)}_m, r^{(t)}_m, s^{(t+1)}_m)$ in $\mathcal{D}$.
  - Sample a random mini batch of transitions (s_i, a_i, r_i, s_{i+1}) from $\mathcal{D}$.
  - Update DQN using the stochastic gradient descent.
  - Update the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ using $\boldsymbol{\theta} = \boldsymbol{\theta} + \beta\nabla\mathcal{L}(\boldsymbol{\theta})$.
- **end while**

end for

---

where the cumulative term is only related to its corresponding p_m. Therefore, to obtain the max EE, we only need to maximize each cumulative term. The derivative of term η_m with respect to p_m can be expressed as:

$$\frac{\partial\eta_m}{\partial p_m} = BW_n\frac{\frac{p_m+C}{\log_2(1+\frac{h_{m,n}}{z_n}p_m)} - \log_2(1+\frac{h_{m,n}}{z_n}p_m)}{(p_m + C)^2} \tag{15}$$

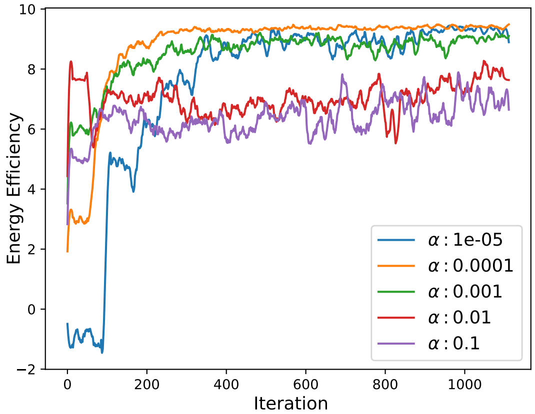
Considering that $(p_m + C)^2{>}0$, we only need to discuss the zero point of the numerator, which can be denoted as

$$y_m = \frac{p_m + C}{\log_2(1+\frac{h_{m,n}}{z_n}p_m)} - \log_2(1+\frac{h_{m,n}}{z_n}p_m) \tag{16}$$

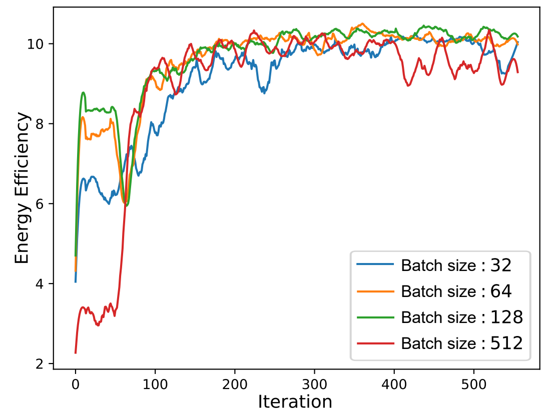
Therefore, the derivative of y_m with respect to p_m can be expressed as:

$$\frac{\partial y_m}{\partial p_m} = -\frac{1}{\ln 2}\frac{\frac{h_{m,n}}{z_n}(p_m + C)}{(1+\frac{h_{m,n}}{z_n}p_m)^2} \tag{17}$$

where $\frac{\partial y_m}{\partial p_m}{<}0$, thus y_m is a monotonously decreasing function with $p_m \in [0, +\infty)$. When $p_m = 0$, $y_m = \frac{C}{\ln 2}{>}0$; when $p_m \to \infty$, $y_m{<}0$. Thus, there exists point p_{peak} makes $y_m = 0$, which is the maximum point of η_m.



(a) different learning rates.



(b) different batch sizes.

Fig. 2: Convergence of JADO under different learning rates and batch sizes

Since the transmission power is limited between $p_{\min}$ and $p_{\max}$, the optimal power for ED m is

$$p_m^* = \begin{cases} p_{\min}, & p_{peak} \in [0, p_{\min}] \\ p_{peak}, & p_{peak} \in (p_{\min}, p_{\max}) \\ p_{\max}, & p_{peak} \in [p_{\max}, +\infty) \end{cases} \tag{18}$$

V. Joint Allocation Framework Based on DRL and Optimization

Algorithm 1 shows the training process of JADO. After observing the state s_t, LoRa gateway selects action a_t based on the probability ϵ and passes the action a_t into the optimization module. Then it calculates an optimal power p^*, and then passes them to the EDs for transmission. Environment feeds back reward r_t and updates next state s_{t+1}. Afterwards, the experience replay buffer stores transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$ and samples a random mini-batch of transitions (s_i, a_i, r_i, s_{i+1}) from $\mathcal{D}$. The loss function is defined as the mean square deviation between the current Q-value and the target Q-value.

$$loss(\boldsymbol{\theta}) = \mathbb{E}\left[\left(r_t + \gamma \max_{a_t'} Q(s_t, a_t'; \boldsymbol{\theta}) - Q(s_t, a_t; \boldsymbol{\theta})\right)^2\right] \tag{19}$$

where $\boldsymbol{\theta}$ is the parameter set of the Q-network, and the terms in the square represent the temporal-difference error. Parameter $\boldsymbol{\theta}$ is updated through

$$\boldsymbol{\theta} = \boldsymbol{\theta} + \alpha \nabla \mathcal{L}(\boldsymbol{\theta}) \tag{20}$$

where $\nabla \mathcal{L}(\boldsymbol{\theta})$ represents the first-order partial derivative with respect to $\boldsymbol{\theta}$.

VI. Simulation Results

A. Simulation Setup

We consider a cell with the radius $R = 12$ km. A LoRa gateway is located at the center of the cell. The minimal transmit power (i.e., $p_{\min}$) is 10 mW and the maximal is $p_{\max} = 100$ mW. The duty cycle is set to 1%. The LoRa gateway operates in sub-GHz band and can listen to up to 3 channels. The channel bandwidth is set to 125 kHz. Each channel follows the Rayleigh fading channel model. Path loss exponent is set to 3. Furthermore, the channel gain is randomly generated with Rayleigh distribution at the beginning of each time slot. The noise power (i.e., z_n) is -124 dBm. Without loss of generality, we set $C = 1$ mW.

B. Convergence of JADO

Supposing that there are 6 LoRa EDs sending data simultaneously. As shown in Fig. 2(a), we evaluate the impact of learning rate on system EE and clearly show that JADO converges under different settings. In particular, the learning rate controls the adjustment speed of the weight in DQN based on loss gradient. We can observe that the fastest convergence speed and the highest EE are achieved when $\alpha = 0.0001$. If the learning rate is too large or too small, the convergence speed will be too slow and unstable. Thus, we use this value in the subsequent experiments. In Fig. 2(b), the value of EE gradually increases and tends to stabilize with the increasing of batch size. We can see that the fastest convergence speed is achieved when the batch size equals to 128. According to the above experimental results, we set $\alpha = 0.0001$ and batch size $= 128$.

C. EE Performances of JADO

As Fig. 3 shown, we consider the performance comparison of EE generated by using JADO to allocate for different numbers of EDs.

Since the number of CHs and SFs is fixed, the collision probability will rise with the increase of the number of EDs. This puts forward higher requirements for the feasibility and success of JADO. In Fig. 3, with the increase of the number of EDs to be allocated, the system EE also increases in equal proportion, which is in line with the actual situation.In addition, the more EDs, the more intense the EE fluctuation, which is reasonable. Therefore, these prove that JADO is in line with the actual situation and has very good performance.

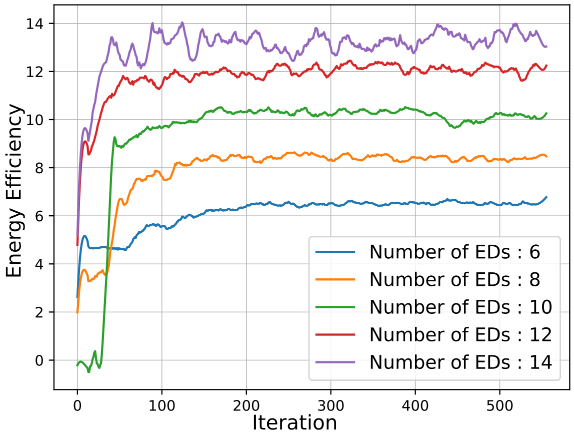


Fig. 3: Energy efficiency of different numbers of EDs

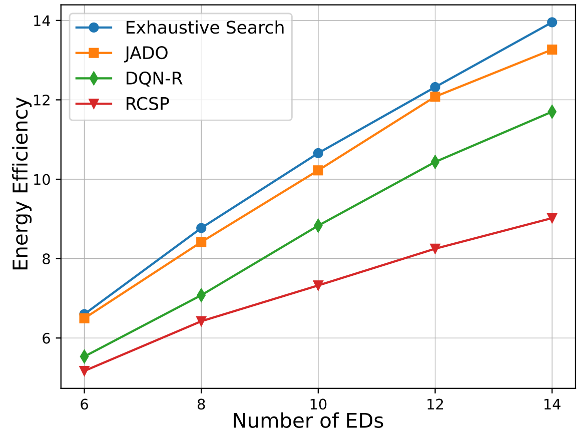


Fig. 4: Comparison of different algorithms

D. Comparison With Existing Algorithms

To further demonstrate the advantages of JADO, we compare it with several existing methods.

1) *Random CH/SF and Power Selecting (RCSP):* EDs randomly select the transmit parameters, ignoring the communication environment and the channel occupation. RCSP is adopted in current LoRa networks.
2) *Exhaustive Search:* The optimal allocation strategy is achieved through exhaustive search. Although exhaustive search is inapplicable for real scenes, it serves as a benchmark for evaluating the performance of JADO.
3) *DQN-Based CH/SF Allocation and Random Power Selecting (DQN-R):* The CH/SF allocation strategy of each ED is obtained via the DQN framework proposed in Section III, while the power is randomly selected.

Fig. 4 illustrates the system EE versus the number of LoRa EDs. The results of RCSP, DQN-R and the exhaustive search approach are provided for comparison. Obviously, the system EE increases linearly with the increase of the number of EDs. It is worth noting that the performance of the RCSP is not good due to its randomness, and the performance of JADO is close to that of exhaustive search. Otherwise, the performance of DQN-R is between JADO and RCSP, which proves that our proposed CH/SF allocation framework and power optimization scheme are both effective. In addition, as the number of active EDs increases, the gap between JADO and RCSP becomes larger. This is because when the number of active EDs increases, the resources allocation scheme proposed by JADO can well avoid the co-SF interference caused by EDs sharing the same CH/SF, while RCSP cannot suppress this interference due to the randomness of transmission parameters selecting.

VII. Conclusion

In this paper, we proposed a DQN and optimization based resource allocation algorithm for LoRa networks. The DQN obtains CSI from uplink frames of EDs without introducing additional transmission overhead. The learning agent running on the LoRa gateway interacts with the operating environment and updates the CH/SF allocation policy according to the feedback from the environment. Given the policy, an optimization module is then constructed to derive the optimal power allocation strategy to maximize the system EE. Simulation results have shown that our algorithm converges under different simulation settings. Furthermore, the proposed algorithm is superior to the RCSP and achieves near-optimal system EE.

References

[1] D. Magrin, M. Capuzzo, A. Zanella, L. Vangelista, and M. Zorzi, "Performance analysis of lorawan in industrial scenarios," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6241–6250, 2021.

[2] A. Gamage, J. C. Liando, C. Gu, R. Tan, and M. Li, "Lmac: Efficient carrier-sense multiple access for lora," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '20. New York, NY, USA: Association for Computing Machinery, 2020.

[3] L. Beltramelli, A. Mahmood, P. Österberg, and M. Gidlund, "Lora beyond aloha: An investigation of alternative random access protocols," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3544–3554, 2021.

[4] R. Marini, W. Cerroni, and C. Buratti, "A novel collision-aware adaptive data rate algorithm for lorawan networks," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2670–2680, 2021.

[5] P. Kumari, H. P. Gupta, and T. Dutta, "Estimation of time duration for using the allocated lora spreading factor: A game-theory approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 090–11 098, 2020.

[6] L. Amichi, M. Kaneko, E. H. Fukuda, N. El Rachkidy, and A. Guitton, "Joint allocation strategies of power and spreading factors with imperfect orthogonality in lora networks," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3750–3765, 2020.

[7] B. Su, Z. Qin, and Q. Ni, "Energy efficient uplink transmissions in lora networks," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4960–4972, 2020.

[8] X. Wang, Y. Zhang, R. Shen, Y. Xu, and F.-C. Zheng, "Drl-based energy-efficient resource allocation frameworks for uplink noma systems," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7279–7294, 2020.

[9] B. Gu, X. Zhang, Z. Lin, and M. Alazab, "Deep multiagent reinforcement-learning-based resource allocation for internet of controllable things," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3066–3074, 2021.

[10] X. Chen and G. Liu, "Energy-efficient task offloading and resource allocation via deep reinforcement learning for augmented reality in mobile edge networks," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 843–10 856, 2021.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness *et al.*, "Human-level control through deep reinforcement learning," *Nat.*, vol. 518, no. 7540, pp. 529–533, 2015.