A Joint Optimization of Sensor Activation and Mobile Charging Scheduling in Industrial Wireless Rechargeable Sensor Networks

Jiayuan Chen*, Changyan Yi*, Ran Wang*, Kun Zhu* and Jun Cai†
*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
†Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, H3G 1M8, Canada
Email: {jiayuan.chen, changyan.yi, wangran, zhukun}@nuaa.edu.cn, jun.cai@concordia.ca

***Abstract*—In this paper, a joint optimization of sensor activation and mobile charging scheduling for industrial wireless rechargeable sensor networks (IWRSNs) is studied. In the considered model, an optimal sensor set is selected to collaboratively execute a bundle of heterogeneous tasks of production-line monitoring, meeting the quality-of-monitoring (QoM) of each individual task. There is a mobile charger vehicle (MCV) which is scheduled for recharging sensors before their charging deadlines (i.e., the time instant of running out of their energy). Our goal is to jointly optimize the sensor activation and MCV scheduling for minimizing the energy consumption of the entire IWRSN, subjected to tasks' QoM requirements, sensor charging deadlines and the energy capacity of the MCV. Unfortunately, solving this problem is non-trivial, because it involves solving two tightly coupled NP-hard problems. To address this issue, we design an efficient algorithm integrating deep reinforcement learning and marginal product based approximation algorithm. Simulations are conducted to evaluate the performance of the proposed solution and demonstrate its superiority over counterparts.**

I. Introduction

With the development of intelligent manufacturing, industrial wireless sensor networks (IWSNs) have been widely used for the automatic control of industrial production process and the monitoring of various parameters. Nevertheless, wireless sensor nodes are severely energy-limited, which hinders the wide application of IWSNs. To tackle such sensor energy provisioning problem, researchers studied how to reduce the energy consumption by optimizing wake-up and sleeping scheduling, data gathering and routing strategies, etc. to prolong the lifetime of IWSNs. However, these methods cannot fundamentally address the shortage of total energy capacities of sensors. Therefore, recent advances of wireless energy transfer technology have inspired the emergence of industrial wireless rechargeable sensor networks (IWRSNs) [1], in which mobile charger vehicles (MCVs) are employed to travel around and replenish energy for sensors without interconnecting wires.

Although IWRSNs can obviously outperform traditional IWSNs in alleviating the heavy burden of energy consumption, there are still some open problems remaining. In practice, sensing tasks for production-line monitoring may be highly heterogeneous in terms of quality of monitoring (QoM) requirements, locations and types. Besides, industrial sensors may also be heterogeneous in terms of sensing radius, types, etc. Therefore, it is crucial to select the optimal set of sensors to activate for collaboratively and continuously execute all monitoring tasks while meeting the QoM of each task, and such problem becomes more complicated since sensors in IWRSNs are rechargeable.

Furthermore, industrial sensors must keep up high-intensity work for long periods and continuously feed data back to controllers or actuators. For example, while a cutting machine is working, industrial camera sensors must collaboratively monitor the position of cutters in real-time and send out the data in a timely manner. Any unpredictable sensor failure may cause serious consequences, e.g., unexpected damages and casualties. Hence, in order to guarantee that all activated sensors can work continuously during the monitoring period, the MCV in IWRSNs should be scheduled to recharge sensors before their charging deadlines (i.e., the instant of running out of their energy). However, the energy capacity of MCV is also limited, and thus the scheduling of MCV is not only subjected to the charging deadlines of sensors, but also its own energy capacity constraint.

To address the aforementioned issues, in this paper, we study a joint optimization of sensor activation and mobile charging scheduling for IWRSNs. The goal is to jointly optimize the sensor activation and MCV scheduling for minimizing the energy consumption of the considered IWRSN, subjected to tasks' QoM requirements, sensor charging deadlines and energy capacity of the MCV. In the considered model, the MCV starts from the depot, travels along the scheduled path and returns to the depot at the end of a trip. While traveling on its path, the MCV charges activated sensors before their charging deadlines. To solve such joint sensor activation and mobile charging scheduling problem, we propose an efficient algorithm integrating deep reinforcement learning (DRL) and marginal product based approximation algorithm.

The main contributions of this paper are summarized in the following.

- A joint optimization of sensor activation and mobile charging scheduling for IWRSNs is formulated, where the objective is to minimize the energy consumption of the entire network.
- An efficient algorithm, called joint sensor activation and charging scheduling algorithm (JSACS), is proposed inte-

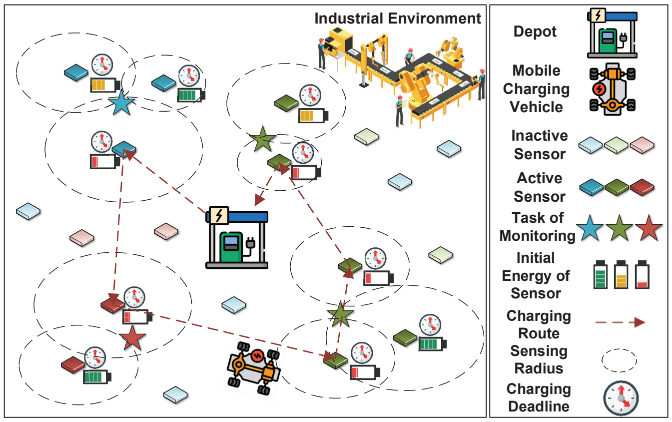

Fig. 1. An illustration of the considered IWRSN.

grating DRL and marginal product based approximation algorithm, which jointly optimizes the sensor activation and the MCV's charging route scheduling.
- Simulations are conducted to show the superiority of the proposed JSACS over counterparts.

The rest of this paper is organized as follows: Section II presents the system model and the problem description. In Section III, an efficient solution for the problem is proposed. Simulation results are provided in Section IV, followed by conclusions in Section V.

II. System Model and Problem Description

A. Network Model

Consider an IWRSN, as illustrated in Fig. 1, consisting of a group of tasks for production-line monitoring, a set of stationary industrial rechargeable sensors $\mathcal{S}$ with cardinality of $|\mathcal{S}| = S$ uniformly distributed in a certain area, and an MCV which starts working from a depot deployed at the center.

At the beginning of a monitoring period, the industrial controller declares its a bundle of monitoring tasks $\mathcal{Z} = \{z_j^m | \forall m \in \{1, 2, ..., M\}, \forall j \in \{1, 2, \ldots, J\}\}$ to the IWRSN, where m and j stand for the index of the monitoring task and its corresponding type, respectively. For meeting the QoM requirements of these tasks, a group of sensors $\mathcal{H} \subseteq \mathcal{S}$ should be activated to collaboratively execute the monitoring tasks.

In practice, sensors' sensing radius are limited, which can be denoted by $R_i, i \in \mathcal{S}$. In addition, different types of sensors can only execute tasks fitting their types, and thus we define $\mathcal{S}_j$ as the set of sensors specialized in task type j. Obviously, each sensor $i \in \mathcal{S}$ can only execute task $z_j^m \in \mathcal{Z}$ that is located within its sensing radius R_i and falls into its targeted type. In each monitoring period, each sensor is able to execute at most one task. In this paper, we adopt the probabilistic sensing coverage (PSC) model [2], [3], and denote p_{i,z_j^m} as the detection probability of z_j^m by sensor i, which can be calculated as

$$p_{i,z_j^m} = \begin{cases} e^{-\alpha_i \cdot dist\left(i, z_j^m\right)}, & \text{if } dist\left(i, z_j^m\right) \leq R_i, i \in \mathcal{S}_j, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where α_i represents the intensity coefficient related to the sensor i's physical characteristics, and $dist\left(i, z_j^m\right)$ indicates the Euclidean distance between sensor i and task z_j^m [2]–[4]. The collaborative coverage probability of sensor set $\mathcal{H}$ to the monitoring task z_j^m is required to be larger or equal to $P_{z_j^m}^{demand}$, i.e.,

$$1 - \prod_{i \in \mathcal{H}} (1 - p_{i,z_j^m}) \geq P_{z_j^m}^{demand}, \tag{2}$$

where $P_{z_j^m}^{demand}$ measures the minimum QoM demanded by each task z_j^m. For sensors that are activated to execute tasks, they should work continuously during the monitoring period due to the application for industrial monitoring. However, the battery capacity of each sensor $E_i^{capacity}$ is limited, and once the battery is completely consumed, the sensor stops working. To this end, the MCV is employed with energy capacity E_{MCV} which travels starting at the depot, charges dying sensors in $\mathcal{H}$ and returns to the depot at the end. Because of the hardware limitation, the MCV can only recharge one sensor at a time. We denote $E_i^{initial}$ as the initial energy of each sensor $i \in \mathcal{S}$ at the beginning of the monitoring period. For simplicity, assume that for each sensor $i \in \mathcal{S}$, $E_i^{initial}$ is sufficiently large to guarantee that $E_i^{initial} \geq E_i^{min}$, where E_i^{min} is the minimum energy for $i \in \mathcal{S}$ to be operational. Here, we characterize the energy consumption rate of each sensor $i \in \mathcal{S}$ by $E_i^{consume}$. Note that it is possible that some sensors may have sufficiently enough energy so that they can work continuously during the monitoring period and are not necessary to be recharged by the MCV. We classify these sensors into the set $\mathcal{H}_0 \subseteq \mathcal{H}$, and categorize the others which have to be recharged by the MCV into set $\mathcal{H}_1 = \mathcal{H} \backslash \mathcal{H}_0$. Obviously, the amount of energy that sensor $i \in \mathcal{H}_1$ required to be recharged can be calculated as

$$E_i^{demand} = T \cdot E_i^{consume} - (E_i^{initial} - E_i^{min}), \forall i \in \mathcal{H}_1, \tag{3}$$

where T is the time duration of each production-line monitoring task period.

For ensuring that all activated sensors can execute tasks continuously, the MCV should charge the sensors in set $\mathcal{H}_1$ before their charging deadline ddl_i, $i \in \mathcal{H}_1$, which can be calculated as

$$ddl_i = \frac{E_i^{initial} - E_i^{min}}{E_i^{consume}}, \forall i \in \mathcal{H}_1. \tag{4}$$

Besides, let us denote the charging route of the MCV by a vector $\mathcal{L}^{\mathcal{H}_1} = \{\pi_0, \pi_1, ..., \pi_g, ..., \pi_{|\mathcal{H}_1|}, \pi_{|\mathcal{H}_1|+1}\}$, where π_g signifies the g_{th} visiting target (i.e., the targeted sensor for recharging). Specifically, $\pi_0 = \pi_{|\mathcal{H}_1|+1} = 0$ indicates that the MCV travels starting from the depot and returns at the end, and $\pi_g \in \mathcal{H}_1$ for $g = 1, \ldots, |\mathcal{H}_1|$. Note that, each sensor $i \in \mathcal{H}_1$ can only be visited once, that is $\pi_g \neq \pi_{g'}$ for $g \neq g'$. Furthermore, we define the arrival time of the MCV at a visiting target π_g as A_{π_g}. Clearly, A_{π_g} depends on the arrival time of the last visited target π_{g-1}, the service time (i.e., battery recharging time) for the target π_{g-1}, and the traveling time of the MCV from π_{g-1} to π_g. Hence, A_{π_g} can be expressed as

$$A_{\pi_g} = A_{\pi_{g-1}} + \frac{E^{demand}_{\pi_{g-1}}}{\varepsilon} + \frac{dist\left(\pi_{g-1}, \pi_g\right)}{v}, \forall \pi_g \in \mathcal{L}^{\mathcal{H}_1}, \quad (5)$$

where ε and v stand for the the charging efficiency and the velocity of the MCV, respectively. Following the definition in (3), $E^{demand}_{\pi_g}$ depicts the amount energy that the target π_g (or sensor π_g) demands for recharging. In particular, $E^{demand}_{\pi_0} = E^{demand}_{\pi_{|\mathcal{H}_1|+1}} = 0$, and $A_{\pi_0} = 0$.

In this paper, we assume that when a sensor $i \in \mathcal{H}$ has been fully recharged, it can work continuously without interruption during the monitoring period, namely $E^{capacity}_i \geq T \cdot E^{consume}_i$.

B. Problem Description

The energy consumption of an IWRSN includes the energy consumption of the MCV and the energy consumption of sensors in $\mathcal{H}$ for executing tasks. Although the energy cost of the MCV further consists of both the traveling energy cost and the recharging energy cost, all recharging energy will be consumed completely by sensors for a higher energy utilization efficiency, and thus such term is implied by the energy cost of sensors in $\mathcal{H}$. Therefore, the total energy consumption of an IWRSN $E_{total}\left(\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}\right)$ can be formulated as

$$E_{total}(\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}) = \sum_{g=0}^{|\mathcal{H}_1|} \gamma \cdot dist\left(\pi_g, \pi_{g+1}\right) + \sum_{i \in \mathcal{H}} T \cdot E^{consume}_i,$$

where γ represents the energy consumption rate from MCV's travelling.

Accordingly, a joint optimization of sensor activation (i.e., the optimal set of sensors to activate $\mathcal{H}$) and mobile charging scheduling (i.e., the optimal charging route $\mathcal{L}^{\mathcal{H}_1}$) for the IWRSN can be formulated as

$$[\mathcal{P}1]: \quad \min_{\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}} E_{total}(\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}) \quad (6)$$

$$s.t., \; 1 - \prod_{i \in \mathcal{H}} (1 - p_{i, z^m_j}) \geq P^{demand}_{z^m_j}, \forall z^m_j \in \mathcal{Z}, \quad (7)$$

$$A_{\pi_g} \leq ddl_{\pi_g}, g=1, \ldots |\mathcal{H}_1|, \quad (8)$$

$$\pi_g \neq \pi_{g'}, g \neq g'; g=1, \ldots |\mathcal{H}_1|, g'=1, \ldots |\mathcal{H}_1|, \quad (9)$$

$$\sum_{g=0}^{|\mathcal{H}_1|} \gamma \cdot dist\left(\pi_g, \pi_{g+1}\right) + \sum_{g=1}^{|\mathcal{H}_1|} E^{demand}_{\pi_g} \leq E_{MCV}, \quad (10)$$

$$\pi_0 = 0, \pi_{|\mathcal{H}_1|+1} = 0, \quad (11)$$

$$\mathcal{H} \subseteq \mathcal{S}, \quad (12)$$

$$\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1, \quad (13)$$

$$\mathcal{L}^{\mathcal{H}_1} = \{\pi_0, \pi_1, ..., \pi_g, ..., \pi_{|\mathcal{H}_1|}, \pi_{|\mathcal{H}_1|+1}\}, \quad (14)$$

where constraint (7) states that each monitoring task's QoM requirement should be met; constraint (8) ensures that the MCV can always be scheduled to arrive before each sensor's charging deadline expires; constraint (9) means that the MCV should not visit the same sensor more than once in the scheduled charging route; constraint (10) indicates that the total energy consumption of the MCV should be less than or equal to its energy capacity E_{MCV}; constraint (11) illustrates that the MCV starts at the depot and returns to the depot at the end. In the following section, we will propose an efficient algorithm to derive the solution of this joint optimization problem.

III. Joint Sensor Activation and Mobile Charging Scheduling

A. Hardness Analysis

From the problem formulation $[\mathcal{P}1]$, we can observe that the joint optimization of sensor activation and mobile charging scheduling actually includes two-layer optimizations. The upper layer optimization mainly addresses the sensor set selection with tasks' QoM constraints, where the objective is to minimize the energy consumption of the activating sensor set $\mathcal{H}$. And the lower layer optimization aims to determine the charging route scheduling for the MCV by taking into account sensors' charging deadlines, where the objective is to minimize the traveling energy consumption of the MCV. Indeed, these two optimization problems are tightly coupled.

Given the charging route $\mathcal{L}^{\mathcal{H}_1}$ of the MCV, we can get the set of candidate sensors $\mathcal{S}^{'} \subseteq \mathcal{S}$, where all sensors in $\mathcal{S}^{'}$ have sufficient energy to execute monitoring tasks continuously during the monitoring period. The upper layer sensor set selection problem turns to be a variant generalized assignment problem, which is NP-hard:

$$[\mathcal{P}2]: \min_{\mathcal{H}} \sum_{i \in \mathcal{H}} T \cdot E^{consume}_i$$

$$s.t., \; (7), \; (13) \text{ and } \mathcal{H} \subseteq \mathcal{S}^{'},$$

While given the set $\mathcal{H}$, the set $\mathcal{H}_1$ can also be obtained and the lower layer mobile charging route scheduling problem can be seen as a reduced traveling salesman with time windows problem, which is NP-hard:

$$[\mathcal{P}3]: \min_{\mathcal{L}^{\mathcal{H}_1}} \sum_{g=0}^{|\mathcal{H}_1|} \gamma \cdot dist\left(\pi_g, \pi_{g+1}\right)$$

$$s.t., \; (8), \; (9), \; (10), \; (11) \text{ and } (14)$$

Based on the above analyses, it is obvious that solving the joint optimization of sensor activation and mobile charging scheduling for the IWRSN directly is very challenging because: i) both the upper layer sensor selection optimization, and the lower layer charging route scheduling problem are NP-hard; ii) the upper and lower layer problems are tightly coupled (i.e., the input of the lower layer problem depends on the output of the upper layer one, while the optimization of the upper problem would impact the lower layer problem). In the following subsection, we first solve the MCV charging route scheduling problem by applying a DRL-based approach. Then, we jointly optimize the sensor set selection and the MCV charging route scheduling by utilizing a marginal product based approximation algorithm.

B. DRL Algorithm for Mobile Charging Route Scheduling

Here, a modified pointer network similar to that in [5] is introduced to model the lower layer problem $[\mathcal{P}3]$, and the Actor-Critic algorithm is utilized for training.

First, we introduce the input structure of the neural network. At each decoding step $g = 0, 1, \ldots, |\mathcal{H}_1| + 1$, let the set of inputs be $X_g = \{x_g^0, x_g^1, \ldots x_g^{|\mathcal{H}_1|}\}$, where $|\mathcal{H}_1|$ indicates the number of targets that need to be recharged. Each x_g^i is represented by a sequence of tuples $\{x_g^i = (s^i, d_g^i)\}$, where s^i and d_g^i stand for the static and dynamic elements of the input, respectively. It is worth noting that dynamic elements of each input are allowed to alter between the decoding steps, while the static elements are invariant. For example, s^i is the attribute of the target i, including target i's location and the charging deadline, which does not change during the charging process. However, the charging requirement of the target i becomes 0 after charging by the MCV. Therefore, x_g^i can be viewed as a vector of features that depicts the state of i at decoding step g. Particularly, x_g^0 represents attributes of the depot, which is set to locate at the center of the area, and its charging deadline is infinite and it has no charging demand.

The output of the model is a permutation of the sensors and the depot, $\mathcal{L}^{\mathcal{H}_1} = \{\pi_0, \pi_1, ..., \pi_{|\mathcal{H}_1|}, ..., \pi_{|\mathcal{H}_1|+1}\}$. At each decoding step $g = 0, 1, \ldots, |\mathcal{H}_1| + 1$, π_g points to a sensor or the depot in X_g, determining the next visiting target. The states of sensors in X_g are updated every time after a target has been visited. When the charging requirements of all sensors are satisfied, the process will be terminated.

To map input X_0 to output $\mathcal{L}^{\mathcal{H}_1}$, the probability chain rule is utilized:

$$P\left(\mathcal{L}^{\mathcal{H}_1} \mid X_0\right) = \prod_{g=1}^{|\mathcal{H}_1|} P\left(\pi_{g+1} \mid \pi_0, \pi_1, \cdots, \pi_g, X_g\right). \quad (15)$$

Firstly the depot is selected as π_0. Eq. (15) provides the probability of selecting the next visiting target according to $\pi_0, \pi_1, \ldots, \pi_g$, i.e., the already visited targets. Then a modified pointer network similar to that in [5] is used to model (15). Its basic structure is the sequence-to-sequence model [6], a powerful model in the machine translation field, which maps one sequence to another. The sequence-to-sequence model consists of two recurrent neural networks (RNNs), namely encoder and decoder.

Encoder encodes the input sequence into a code vector which contains knowledge of the input. Since the attributes of the targets convey no sequential information and the order of targets in the inputs is meaningless, RNN is not necessary to be utilized in the encoder. Therefore, a simple embedding layer is adopted to encode the inputs which decreases the computational compilations without decreasing the efficiency [5]. In this work, we apply a 1-dimensional (1-D) convolution layer to encode the inputs to a *high-dimensional vector* [5] ($d = 128$ in this work). The parameters of the 1-D convolution layer are shared among the inputs.

Different from the encoder, we use RNN to model the decoder network since we need to store the knowledge of previous steps $\pi_0, \pi_1, \cdots, \pi_g$ to assist for obtaining π_{g+1}. The hidden state of RNN decoder d_g can memorize the previously selected visited targets. Then d_g is combined with the encoding of the inputs $\rho_g^0, \rho_g^1, \ldots \rho_g^{|\mathcal{H}_1|}$ to calculate the conditional probability $P\left(\pi_{g+1} \mid \pi_0, \pi_1, \cdots, \pi_g, X_g\right)$.

Algorithm 1: Actor-Critic training algorithm

Output: The optimal model $\mathcal{M}^* = [\theta^*, \phi^*]$.

Initialize: Let the actor network with random weights θ and critic network with random weights ϕ;
for *iteration* $\leftarrow 1, 2, \ldots$ **do**
    generate F problem instances from $\{\Phi_{\mathcal{M}_1}, \Phi_{\mathcal{M}_2}, \ldots, \Phi_{\mathcal{M}_M}\}$;
    for $c \leftarrow 1, \ldots, F$ **do**
        $t \leftarrow 0$;
        while *not terminated* **do**
            select the next target π_{g+1}^c according to $P\left(\pi_{g+1}^c | \pi_1^c, \ldots, \pi_g^c, X_g^c\right)$;
            Update X_g^c to X_{g+1}^c leaving out the visited targets;
        compute the reward R^c ;
    $d\theta \leftarrow \frac{1}{F}\sum_{c=1}^{F}\left(R^c - V\left(X_0^c;\phi\right)\right)\nabla_\theta \log P\left(Y^c \mid X_0^c\right)$;
    $d\phi \leftarrow \frac{1}{F}\sum_{c=1}^{F}\nabla_\phi\left(R^c - V\left(X_0^c;\phi\right)\right)^2$;
    $\theta \leftarrow \theta + \eta d\theta$;
    $\phi \leftarrow \phi + \eta d\phi$;
Determine $\theta^* = \theta, \phi^* = \phi$.

The attention mechanism is utilized to calculate the degree of correlation of each input to the decoding step g. More $attention$ is given to the most relevant one which is more likely to be selected as the next target. The calculation can be expressed as

$$u_g^i = w^T \tanh\left(W_1\rho_g^i + W_2 d_g\right), \quad i \in (0, 1, ..., |\mathcal{H}_1|);$$
$$P\left(\pi_{g+1} \mid \pi_0, \pi_1, \cdots, \pi_g, X_g\right) = \operatorname{softmax}\left(u_g^i\right),$$

where w, W_1, W_2 are $learnable$ parameters. For each target i, its u_g^i is computed by d_g and its encoder hidden state ρ_g^i. The softmax operator is used to normalize $u_g^0, u_g^1, \ldots, u_g^{|\mathcal{H}_1|}$, and probability for selecting each target i at step g can then be obtained. In this paper, the greedy decoder is utilized to select the next target.

We adopt the well-known Actor-Critic method to train the network. The method introduces two networks that require to be trained: i) an actor network, which is the pointer network in this work, is used to calculate the probability distribution for choosing the next target; and ii) a critic network that evaluates the expected reward given a specific problem state. In addition, the critic network uses the same architecture as the pointer network's encoder which maps the encoder hidden state into the critic output. However, during training, the model selects the next target by sampling from the probability distribution instead of choosing the target with the maximum probability.

The training is conducted in an unsupervised way and the training procedure is presented in Algorithm 1. During the training process, we generate instances from distributions $\{\Phi_{\mathcal{M}_1}, \Phi_{\mathcal{M}_2}, \ldots, \Phi_{\mathcal{M}_M}\}$, where $\mathcal{M}$ signifies different input features of the targets, i.e, the targets' locations, charging deadlines, etc. F instances are sampled from $\{\Phi_{\mathcal{M}_1}, \Phi_{\mathcal{M}_2}, \ldots, \Phi_{\mathcal{M}_M}\}$ for training the actor and critic net-

Algorithm 2: Joint Sensor Activation and Charging Scheduling Algorithm (JSACS)

Input: $\mathcal{S}_{z_j^m}^{candidate} = \{i | p_{i,z_j^m} \neq 0, \forall i \in \mathcal{S}\}, \mathcal{S}^{candidate} = \sum_{z_j^m \in \mathcal{Z}} \mathcal{S}_{z_j^m}^{candidate}, \mathcal{Z}^{unsatisfied} = \mathcal{Z}$.

Output: $\mathcal{H}$, $\mathcal{L}^{\mathcal{H}_1}$.

Initialize: Let $\mathcal{H}_0 = \emptyset, \mathcal{H}_1 = \emptyset, \mathcal{H} = \emptyset, E_{MCV}^{travel}(\mathcal{H}_1) = 0$;

while $\mathcal{Z}^{unsatisfied}$ *is nonempty* **do**

  for *each* $i \in \mathcal{S}^{candidate}$ **do**

    if $E_i^{initial} - E_i^{min} \geq T \cdot E_i^{consume}$ **then**

      $E_{MCV}^{travel}(\mathcal{H}_1 \cup \{i\}) = E_{MCV}^{travel}(\mathcal{H}_1)$;

    else

      Call the model $\mathcal{M}^* = [\theta^*, \phi^*]$ in algorithm 1 to get a charging route $\mathcal{L}^{\mathcal{H}_1 \cup \{i\}}$ which meets each sensor's charging deadline (If there is no charging route that meets the sensor's charging deadline or the energy consumption of the MCV exceeds E_{MCV}, delete the sensor i from $\mathcal{S}^{candidate}$.), then compute the energy consumption of the charging route $E_{MCV}^{travel}(\mathcal{H}_1 \cup \{i\})$;

  $i_{selected} =$

$$\arg\max_{i \in \mathcal{S}^{candidate}} \left\{ \frac{(1-\prod_{i' \in \mathcal{H} \cup \{i\}}(1-p_{i',z_j^m})) - (1-\prod_{i' \in \mathcal{H}}(1-p_{i',z_j^m}))}{E_{total}\left(\mathcal{H} \cup \{i\}, \mathcal{L}^{\mathcal{H}_1 \cup h}\right) - E_{total}\left(\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}\right)}, \forall z_j^m \in \mathcal{Z} \right\},$$

  Update $\mathcal{H} = \mathcal{H} \cup \{i_{selected}\}$, $E_{MCV}^{travel}(\mathcal{H}_1) = E_{MCV}^{travel}(\mathcal{H}_1 \cup \{i_{selected}\})$;

  if $E_{i_{selected}}^{initial} - E_{i_{selected}}^{min} \geq T \cdot E_{i_{selected}}^{consume}$ **then**

    Update $\mathcal{H}_0 = \mathcal{H}_0 \cup \{i_{selected}\}$;

  else

    Update $\mathcal{H}_1 = \mathcal{H}_1 \cup \{i_{selected}\}$;

  for *each* $z_j^m \in \mathcal{Z}^{unsatisfied}$ **do**

    if $1 - \prod_{i \in \mathcal{H}}(1 - p_{i,z_j^m}) \geq P_{z_j^m}^{demand}$ **then**

      Update $\mathcal{S}^{candidate} = \mathcal{S}^{candidate} \backslash \{\mathcal{S}_{z_j^m}^{candidate}\}$, $\mathcal{Z}^{unsatisfied} = \mathcal{Z}^{unsatisfied} \backslash \{z_j^m\}$;

  Update $\mathcal{S}^{candidate} = \mathcal{S}^{candidate} \backslash \{i_{selected}\}$;

return $\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}$.

works with parameters θ and ϕ. For each instance, the actor network with current parameters θ produces the permutation of targets, and the corresponding reward can be obtained. Then policy gradient is computed in line 10 to update the actor network. Meanwhile, the critic network is updated in line 11 by reducing the difference between the observed rewards and the approximated rewards.

C. Joint Sensor Activation & Charging Scheduling Algorithm

Based on the MCV's traveling energy consumption calculated by the trained model $\mathcal{M}^*$, the core idea is iteratively selecting a new sensor i which has the largest marginal product [7]. Marginal product is a concept in economics, which refers to the increase in the total output brought about by adding a unit of an input, assuming that the quantities of other inputs are maintained as constant [7]. In this paper, the energy consumption of the IWRSN corresponds to the adding input, and the QoM obtained by all tasks corresponds to the output. Then, in each iteration, a new activating sensor should be selected according to:

$$\arg\max_{i \in \mathcal{S}^{candidate}} \left\{ \frac{(1-\prod_{i' \in \mathcal{H} \cup \{i\}}(1-p_{i',z_j^m})) - (1-\prod_{i' \in \mathcal{H}}(1-p_{i',z_j^m}))}{E_{total}\left(\mathcal{H} \cup \{i\}, \mathcal{L}^{\mathcal{H}_1 \cup h}\right) - E_{total}\left(\mathcal{H}, \mathcal{L}^{\mathcal{H}_1}\right)}, \forall z_j^m \in \mathcal{Z} \right\},$$

where h indicates whether this sensor needed to be recharged or not:

$$h = \begin{cases} \{i\}, & \text{if } E_i^{initial} - E_i^{min} < T \cdot E_i^{consume}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Initially, $\mathcal{H} = \emptyset$, and the details of the proposed JSACS algorithm can be found in Algorithm 2.

TABLE I
MAIN SIMULATION PARAMETERS.

Parameter	Value
Sensor types	[0,1,2,3]
Task types	[0,1,2,3]
Number of sensors	800 (number of each type: 200)
Number of tasks	40 (randomly chosen over [0,1,2,3])
Area dimensions	80 m $\times$ 80 m
Sensing radius R_i	randomly chosen over [10,15,20,25] m
Energy capacity $E_i^{capacity}$	10.8 kJ
Energy consumption rate $E_i^{consume}$	0.5 J/s
Minimum energy E_i^{min}	540 J
Initial energy $E_i^{initial}$	randomly over [1080,3240] J
Intensity coefficient α_i	randomly over [0.1,0.3]
QoM demand $P_{z_j^m}^{demand}$	randomly over [0.5, 0.7]
Charging efficiency ε	15 W
Velocity v	2 m/s
Traveling energy consumption γ	20 J/m
Energy capacity of MCV E_{MCV}	128 kJ
Time duration of monitoring period T	1 hour

IV. SIMULATION RESULTS

In this section, simulations are conducted to numerically evaluate the performance of the proposed JSACS for problem $\mathcal{P}1$. Table I lists the values of main simulation parameters. Similar settings have been employed in the literature [8]. Note that some parameters may vary according to different evaluation scenarios.

For effective and fair comparisons, we introduce the greedy algorithm (GRE) and an existing algorithm named reward-cost ratio algorithm (RC-ratio) [9]. GRE greedily selects sensors into $\mathcal{H}$ that have maximum coverage probability until all tasks' QoM are satisfied and then applies the earliest deadline first policy (EDF) [10] to derive the charging tour of the MCV for $\mathcal{H}_1$. For EDF, MCV always selects a sensor with the earliest charging deadline as its next serving target. Besides, both the charging deadlines of sensors in $\mathcal{H}_1$ and the energy capacity of MCV are taken into account when selecting each sensor. RC-ratio selects sensors into $\mathcal{H}$ according to the marginal product function while the MCV's charging route is determined by EDF.

Fig. 3 demonstrates the superiority of the proposed JSACS in terms of the entire network energy consumption. It is shown that, the energy consumption of the entire network increases monotonically with the number of tasks. This is because with the growth of the number of tasks, more sensors need to be activated, leading to more energy consumption. Meanwhile,

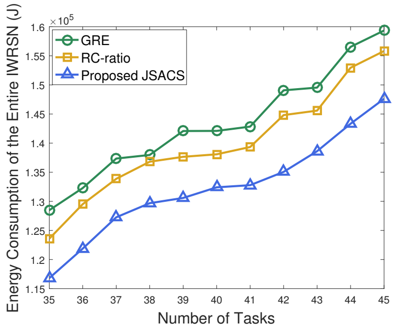

Fig. 3. Comparison of energy consumption of the entire IWRSN w.r.t. number of tasks.

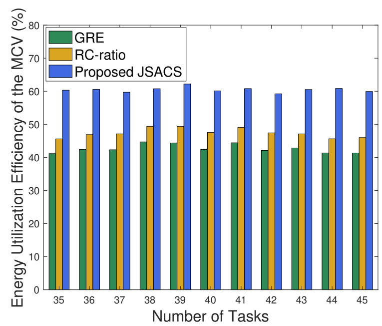

Fig. 4. Comparison of energy utilization efficiency of the MCV w.r.t. number of tasks.

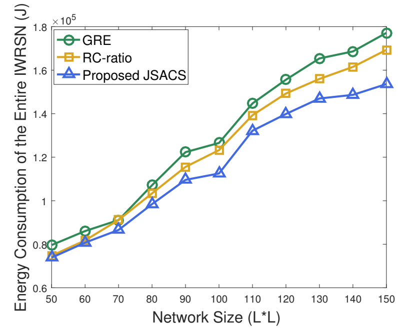

Fig. 5. Comparison of energy consumption of the entire IWRSN w.r.t. network sizes.

with more sensors being activated, a growing number of them need to be recharged within this area, resulting in the increase of the MCV's traveling energy consumption. Additionally, it can be observed that the proposed JSACS outperforms GRE and RC-ratio. The reason is that GRE iteratively selects a sensor with maximum coverage probability while ignores the sensor selection impacts on the total energy consumption. RC-ratio outperforms GRE since RC-ratio selects a sensor with maximum marginal product in each iteration. The proposed JSACS achieves the best performance because it does not only select a sensor with the largest marginal product in each iteration, but also determines the charging route of the MCV by a well DRL model instead of EDF.

Fig. 4 compares the energy utilization efficiency of GRE, RC-ratio and proposed JSACS. The energy utilization efficiency refers to the proportion of the energy for recharging sensors to total MCV energy consumption. It is shown that the proposed JSACS performs better than GRE and RC-ratio. The reason is that the proposed JSACS consider the two-layer optimization simultaneously when selecting a sensor. In addition, the objective of the trained DRL model is to minimize the traveling energy consumption of the MCV while meeting the charging deadlines of sensors. However, the EDF applied in GRE and RC-ratio does not consider the traveling length of the MCV, and it simply recharge sensors in a timely manner. Therefore, the proposed JSACS can prompt the MCV to utilize more energy for task execution to increase the QoM of tasks, rather than wasting energy on traveling.

Fig. 5 shows that the energy consumption of the entire network of these three algorithms increases almost linearly with the network size. The reason is that the larger network size makes the sensor deployment more sparse, leading to more energy consumption on traveling. In addition, a larger network size also makes the distance between the sensor and its monitoring tasks larger, and the detection probabilities of sensors decrease, so that more sensors need to be activated to execute tasks, inducing more energy consumption of sensors. Intuitively, the proposed JSACS outperforms GRE and RC-ratio, benefiting from integrating DRL and marginal product based approximation algorithm to jointly solve the sensor activation and charging scheduling problem.

V. Conclusion

In this paper, the joint optimization of sensor activation and mobile charging scheduling for IWRSNs has been studied. By considering the objective of minimizing the energy consumption of the entire network subjected to tasks' QoM requirements, sensor charging deadlines and the energy capacity of the MCV, an efficient algorithm named JSACS is proposed integrating DRL and marginal product based approximation algorithm. Simulation results show that, compared to counterparts, the proposed algorithm can decrease the energy consumption of the entire IWRSN and improve the energy utilization efficiency of the MCV.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) under Grants 62002164, 62176122, and 62171218.

References

[1] Y. Feng, W. Zhang, G. Han, Y. Kang, and J. Wang, "A newborn particle swarm optimization algorithm for charging-scheduling algorithm in industrial rechargeable sensor networks," *IEEE Sensors J.*, vol. 20, no. 18, pp. 11 014–11 027, 2020.

[2] H. P. Gupta, T. Venkatesh, S. V. Rao, and T. Dutta, "Analysis of coverage under border effects in three-dimensional mobile sensor networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2436–2449, 2017.

[3] C. Yi, J. Cai, K. Zhu, and R. Wang, "A queueing game based management framework for fog computing with strategic computing speed control," *IEEE Trans. Mobile Comput.*, 2022.

[4] C. Yi, J. Cai, T. Zhang, K. Zhu, B. Chen, and Q. Wu, "Workload reallocation for edge computing with server collaboration: A cooperative queueing game approach," *IEEE Trans. Mobile Comput.*, pp. 1–1, 2022.

[5] M. Nazari, A. Oroojlooy, L. V. Snyder, and M. Takáč, "Reinforcement learning for solving the vehicle routing problem," in *Adv. Neural Inf. Process. Syst.*, 2018, pp. 9839–9849.

[6] I. Sutskever and O. Vinyals, "Sequence to sequence learning with neural networks," in *Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[7] A. Brewer, *The making of the classical theory of economic growth*. Routledge, 2010.

[8] T. Liu, B. Wu, S. Zhang, J. Peng, and W. Xu, "An effective multi-node charging scheme for wireless rechargeable sensor networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2020.

[9] T. Wu, P. Yang, H. Dai, C. Xiang, X. Rao, J. Huang, and T. Ma, "Joint sensor selection and energy allocation for tasks-driven mobile charging in wireless rechargeable sensor networks," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11 505–11 523, 2020.

[10] J. A. Stankovic, M. Spuri, K. Ramamritham, and G. C. Buttazzo, *Deadline scheduling for real-time systems: EDF and related algorithms*. Springer Science & Business Media, 2012, vol. 460.