ORIGINAL RESEARCH Open access

Multi-Agent Reinforcement Learning for Efficient UAV Charger Deployment in IoT Energy Networks

Karolina Nowak[1*], Tomasz Wrobel[2], Julia Mazur[1]

Abstract

Employing unmanned aerial vehicles (UAVs) for wireless power transfer (WPT) opens up new possibilities for extending the lifespan of wireless sensor networks (WSNs), a foundational element of the Internet of Things (IoT). Nonetheless, prior studies on charging deployment frequently neglect the network's heterogeneous energy needs, leading to suboptimal charging of high-consumption nodes. This study formulates the multiple UAVs optimal cooperative charging deployment problem (MUAVs-OCCDP) and introduces a phased optimization methodology. The network topology is first built, and the energy demands of individual nodes are cataloged. Drawing upon the strength advantage relationship (SDR), an enhanced NSGA-II algorithm is crafted to formulate an initial deployment layout. Following this, a two-stage reinforcement learning architecture is constructed: the first stage seeks to shrink the UAV fleet size by refining the number of covered nodes and mean charging efficiency; the second stage encourages cooperative behavior via multi-agent experience sharing and a composite reward function, aiming for an equitable distribution of charging energy.

Keywords Reinforcement learning, IoT, WSNs, Energy requirement, NSGA-II, UAVs deployments

*Correspondence:
Karolina Nowak
karolina.nowak@pw.edu.pl
[1] Department of Intelligent Energy Engineering, Faculty of Electrical Engineering, Warsaw University of Technology, Warsaw, Poland
[2] Department of AI and Smart Infrastructure, Faculty of Engineering, AGH University of Science and Technology, Krakow, Poland

Introduction

In the current era, the maturation of Unmanned Aerial Vehicle (UAV) technology, together with the growing proliferation of low-altitude platforms, has cemented UAVs as a pivotal technology underpinning the Internet of Things (IoT) [1]. UAV-aided IoT systems offer heightened adaptability and scalability, making them integral to present-day communication infrastructures [2]. Wireless Sensor Networks (WSNs), a core IoT component, are broadly implemented across complex domains—from military reconnaissance and smart urban management to disaster early warning—thanks to their self-configuring nature and straightforward deployment [3]. Yet, the limited energy reserves of sensors mean that even marginal performance degradation in a few nodes can trigger substantial cost burdens and degrade overall network efficiency [4]. In contexts such as covert operations in hostile areas or installations in inaccessible regions, the impracticality of routine battery replacement makes energy provisioning a fundamental bottleneck for sustained network stability [5].

Energy scavenging technologies [6-9], which supplement node power by capturing ambient environmental sources like wind and solar radiation [8], offer a potential pathway [9]. However, their performance hinges on environmental variability and is highly unpredictable, rendering them typically uncontrollable. By contrast, UAV-enabled Wireless Power Transfer (WPT) [10] has emerged as a superior alternative due to its maneuverability and low observability. The integration of WPT

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

with WSNs has spawned a new research area: Wireless Rechargeable Sensor Networks (WRSNs) [11]. This is especially salient in stealth-critical military missions; UAVs can maintain a minimal signature through omnidirectional hover-based charging, thus lowering the risk of exposing target coordinates [12]. WPT permits the simultaneous replenishment of numerous nodes within a designated radius, directly addressing the challenges of finite network energy. Deploying a fleet of UAVs to power WRSNs is therefore a viable strategy. Given the trade-offs between charging expenditures and network imperatives, the critical challenge lies in optimizing energy dispatch while keeping deployment costs at a minimum.

Existing charging deployment approaches predominantly target cost reduction [4, 13-17] and efficiency gains [13-16]. These studies frequently presuppose uniform energy drain across all nodes or ignore nodal consumption entirely. Consequently, deployments derived from such models can suffer from skewed energy distribution [18]. This leads to energy wastage through overcharging of low-demand nodes, while failing to meet the requirements of high-demand nodes relative to their consumption rate. The resulting imbalance forces vital backbone nodes into a dormant state, compromising network performance and generating significant cost inefficiencies. In addition, most current research relies on heuristic or clustering techniques—such as NSGA-II [13], greedy methods [19], K-means [14], and firefly algorithms [20, 21]—to solve the deployment problem. These methods are inherently limited by the spatial distribution of nodes, which can trap the optimization process in local minima and promote superfluous, redundant deployments. Accordingly, when planning UAV-based charging deployments, adequately accommodating the energy needs of nodes to extend the network's functional lifetime is paramount. Moreover, charging cost constitutes a crucial factor, encapsulating various elements [16] such as hardware unit price and electricity tariffs. Assuming reasonable transmission power levels, the total count of deployed chargers directly dictates the overall charging expenditure. Hence, reducing this count is an essential first step toward a comprehensive solution.

Figure 1 depicts the specific contexts under consideration—wartime, post-catastrophe, and isolated terrains, including urban zones, forests, oceans, and wilderness—where sensor nodes are routinely relied upon for sensing or communication relay. This paper explores the deployment of UAVs acting as mobile chargers to fortify the energy supply of WRSNs, presenting a WRSNs architecture built on multi-UAV cooperation (MUAVs-WRSNs). The challenge is formally defined as the multi-UAV optimal collaborative charging deployment problem (MUAVs-OCCDP), whose objective is to satisfy network requirements at minimal charging cost. The MUAVs-OCCDP is decomposed into three distinct objective functions, which are addressed sequentially. We first exploit the self-organizing property of the sensor network to map the communication topology and precisely gauge each node's energy consumption using a detailed energy dissipation model. Following this, we propose an improved NSGA-II algorithm to establish preliminary UAV coordinates. Lastly, we introduce a two-phase reinforcement learning adaptive refinement algorithm, incorporating both single-agent and multi-agent paradigms (a portion of these findings has been accepted for the IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA 2025). This enables the UAVs to dynamically calibrate their positions from the initial layout, thereby cutting the total number of deployed units and optimizing the energy top-up for nodes with high consumption rates. Within this process, a mechanism for multi-agent collaboration and experiential reinforcement is implemented to enhance the learning pace and flexibility of the agents. The deployed network is evaluated experimentally through normalized objective functions, showing marked improvements over other state-of-the-art techniques. The principal contributions and findings of this research are consolidated below.

1. An enhanced and innovative NSGA-II variant is developed, incorporating a novel Strength Dominance Relationship (SDR) [22]. By tuning the weights attached to the objective functions, the algorithm normalizes multiple goals and applies traditional Pareto dominance to rank non-dominated solutions. This steers convergence toward configurations that minimize the number of deployed UAVs. Further, strategies like sampling and tournament selection are woven in to boost the algorithm's diversity and convergence strength, particularly effective for scaling to larger problem instances.

2. The fusion of NSGA-II and reinforcement learning yields significant gains in solution quality, environmental adaptability, and global search proficiency within dynamic and complex landscapes. Compared to standard greedy heuristics and clustering routines, this hybrid approach better addresses typical challenges involving multiple prioritized objectives, dynamic shifts, and black-box optimization scenarios.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

3. A two-phase reinforcement learning optimization scheme, anchored in both single-agent and multi-agent setups, is established. Here, the global search capability of reinforcement learning is harnessed to fine-tune the initial positions, framed conceptually as the agent's relocation process. The optimization outcome secures the smallest feasible deployment count under the constraint of complete network coverage. Simultaneously, a multi-agent cooperation framework driven by local and global reward signals is constructed to adjust UAV placements further, thereby attaining a judicious distribution of network charging energy.

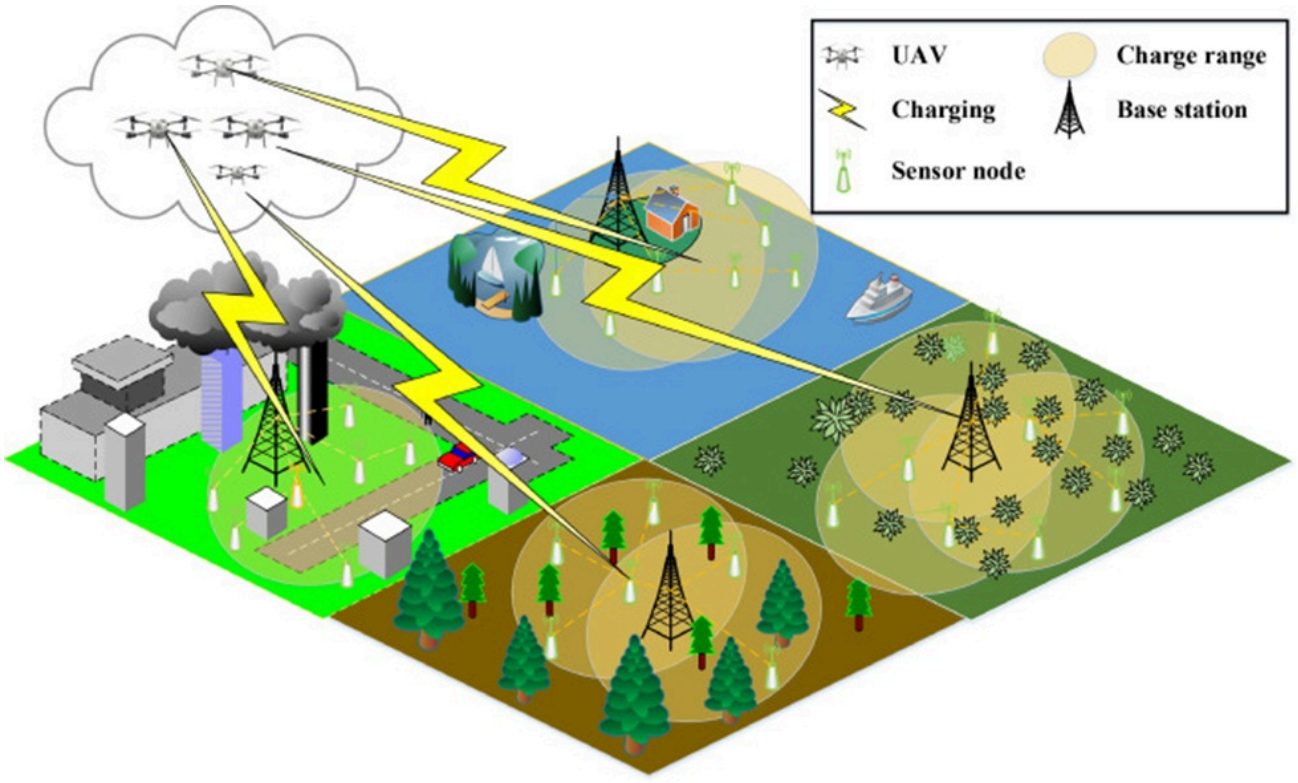


Figure 1. Multiple application scenarios in which UAVs supply energy to WRSNs.

Related Work

This section offers a thorough examination and comparison of current scholarship on charger positioning within WRSNs, highlighting the distinctiveness and importance of our investigative work. Based on their defining features, the body of published material on charger placement in WRSNs can be sorted into two separate branches: static charger positioning and dynamic charger positioning. **Table 1** collates relevant studies from the preceding five-year period, clarifying the differences between those investigations and our own methodological stance.

Table 1. The characteristics of different research.

Paper	Omnidirectional charging	Free deployment location	Energy distribution	Multi-attribute optimization based	Deployment location adjustment
Our	Yes	Yes	Yes	Yes	Yes
[4]	No	No	Yes	Yes	No

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

[13]	No	Yes	No	Yes	No
[14]	Yes	Yes	No	Yes	Yes
[15]	Yes	Yes	Yes	No	No
[16]	Yes	No	No	No	No
[17]	Yes	Yes	No	No	No
[20]	Yes	Yes	No	Yes	No
[23]	Yes	Yes	No	No	No
[24]	Yes	No	No	Yes	No
[25]	No	Yes	Yes	Yes	No
[26]	Yes	Yes	Yes	Yes	No
[27]	Yes	Yes	No	No	No

Fixed charger deployment

Arranging immobile chargers to energize WRSNs calls for refining the positioning of each charger within the network to meet a range of performance benchmarks. Static chargers chiefly depend on WPT technology for one-to-many energy delivery, a category that can be subdivided into omnidirectional [13-21, 23-25, 28] and directional [15, 29-31] modes according to charger attributes. Within the sphere of directional charging layout, scholars have mapped the charging coverage zone through assorted regular geometric forms, applying mathematical formulation tactics like sector-shaped models. Prominently, Ding *et al*. [15, 16] cover both omnidirectional and directional charging schemes. The technique draws on a fusion of candidate direction set pruning and a greedy algorithm to address the problem, practically converting an unbounded search space into a manageable, finite set of candidate directions and thereby reducing the difficulty inherent to NP-hard formulations. Dai *et al*. [29], Liao *et al*. [30], and Dai *et al*. [31] centered on positioning a set number of chargers where device orientation varies unpredictably, to boost the total expected charging benefit. Yet, directional charging might not be broadly suitable to every network arrangement. Turning to omnidirectional charging placement, investigators ordinarily view the effective area of a charger as a circular disk centered on the device [16]. The task is framed as a set covering challenge, striving to ensure no node remains uncovered. The central obstacle is to keep the charger tally as low as achievable. Seeking to trim deployment expenditure, Lian and Yao [17] pioneered the examination of joint sensor and charger placement within WRSNs. That study split the challenge into a pair of NP-hard stages: sensor positioning (P1) and charger positioning (P2), and originally put forward harnessing the sensor “aggregation effect” to shrink the required charger count. You *et al*. [28] explored the task of situating wireless chargers where obstacles exist (WAIT), marking the first comprehensive consideration of obstacle composition, scale, and placement effects on charging behavior. Much of the research turns to metaheuristic techniques to handle these obstacles, including, among others, greedy methods [15, 16, 19], Genetic Algorithm [13], and firefly algorithms [21]. Crucially, these algorithms factor in both charging efficiency and charger quantity side by side, recasting the matter as a multi-objective optimization undertaking. Clustering approaches such as K-means [14], DBSCAN [32], and Mean Shift [24] also see wide use owing to modest computational demands and strong flexibility. They are frequently called upon to determine prime locations for charger installation. Cho *et al*. [23] narrow the solution landscape by outlining the Region of Interest (FoI) and deploying a meta-heuristic algorithm to uncover a cost-efficient layout plan for stationary chargers. The fundamental aim is to maximize the economic efficiency of the fixed arrangement. That said, these algorithms often rely on the starting population and how nodes are distributed, a dependency that can steer the optimization toward local minima and consequently inflate charging costs. For example, Lee *et al*. [25] used an angle-driven clustering multi-way algorithm to identify candidate

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

charging coordinates. Unfortunately, the performance of clustering algorithms is regularly bounded by nodal distribution patterns, possibly yielding imprecise results. For this reason, some researchers have begun merging clustering algorithms with intelligent and machine learning methodologies to polish the outputs derived from preliminary clustering. For instance, Deng *et al*. [24] fed the products of mean shift clustering into the initial population of the NSGA-II algorithm for fixed charger siting. Subsequently, they ran the algorithm to identify superior solution sets—similarly, the strategy of Wang *et al*. [14] blended clustering algorithms with reinforcement learning paradigms. Initially, the K-means algorithm executed a preliminary segmentation of the network, grouping nodes into discrete charging clusters, with their centroids designated as the starting charger locations. Thereafter, reinforcement learning iteratively tuned the deployment coordinates, reducing the charger count while maintaining satisfactory charging efficiency across nodes. Much of the existing published work stresses reducing the charger count to drive down costs while sustaining strong network performance. Departing from these contributions, this paper emphasizes attaining optimal energy distribution at the lowest possible charging cost, tailored to the energy demands of the network's constituent nodes.

Mobile chargers deployment

In the context of WRSNs served by mobile chargers, the quantity of deployed mobile units scales with the size of the network. Stationary and mobile chargers employ WPT technology, accommodating two separate energy delivery patterns: one-to-one [4, 24, 33-35] and one-to-many [20, 36-39]. WRSNs commonly construct mobile charging itineraries for the chargers using accurate time and energy forecasting to satisfy varied network imperatives, such as elevating charging efficiency [20, 26], curbing operational expenditure [36, 27, 33, 38], and stretching network durability [4, 34, 35, 37, 39]. To lift charging efficiency, charger coordination necessitates optimization. Lee *et al*. [34] and Wang *et al.* [35] utilize clustering algorithms to divide the network zone, devoting particular attention to raising charging efficiency within high-demand sectors. Further, Chen *et al*. [38] have advanced a QoS-based on-demand charging scheduling model (QOCS) to confront the tension between Quality of Service (QoS) constraints and charging scheduling overheads in WSNs. The model incorporates a mobile charging vehicle powered by swappable batteries and fitted with multiple chargers, together with a hybrid energy supply scheme to improve energy allocation efficiency while balancing QoS requirements related to data acquisition completeness and charging response delays. A sizeable fraction of the research focuses on reducing operational costs. These inquiries often formulate the charging policy for mobile chargers (MCs) as a charging circuit, applying spatial or temporal partitioning methods to reshape these problems into submodular maximization tasks under assorted network bounds. Liu *et al.* [27] leveraged unmanned aerial vehicles (UAVs) as mobile chargers along with a joint scheduling and trajectory optimization problem (JSTOP) to jointly refine UAV loitering points, travel distances, and node overlap coverage ratios.

Regarding the non-time-out stay probability requirement of mobile nodes at task points (TP), a novel joint deployment challenge was examined [33], targeting the deployment of a minimal number of chargers and receivers to serve those mobile nodes. Liu *et al*. [37] introduced a partial charging scheme and developed a multi-node spatiotemporal partial charging algorithm (MTSPC) to co-optimize dead node count and energy usage efficiency. Across Jia *et al*. [26], Xu *et al*. [40], and Wang *et al*. [41], charger capacity acts as a constraining resource, and the selection of each charging pause is strategically governed by node energy demand and response delay. Wang *et al*. [41] pursue the maximization of nodes covered during an individual stop, simultaneously optimizing adjacent charging beam placements and cutting aggregate mobile energy consumption through a theorem-led dynamic positional search when fixing charging stop locations. Jia *et al*. [26] break free from being limited to deployment strictly at sensor positions or circumscribed areas, and likewise impose no restrictions on node spread or the distribution of charging demand among nodes. This methodology enables worldwide optimization of MC charging coordinates, thereby satisfying network requirements while reducing the energy drained by MC travel, ultimately fostering the self-sustaining functioning of the network.

Alternative lines of inquiry have examined the matter through differing lenses [4, 24, 39, 42, 43]. As a case in point, targeting the harmonization of nodal energy depletion, Dong *et al.* [4] implement a cyclical replenishment protocol for the effective dispatching of chargers, thereby achieving a pronounced drop in the occurrence rate of charging events. The contributions documented in Deng *et al.* [24] and Mallikarjuna and Amgoth [39] each merge stationary and ambulatory

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

charging assets to service the power needs of both routine and high-priority nodes alike. Blending these two charger categories reduces the population of nodes forced to depend on mobile units, as well as the energy burden associated with them. Chen *et al*. [42] unveil an original synthesis of three separate classes of charging hardware, streamlining coordinated operation through a territorial segmentation scheme. Per this scheme, the core ring is provisioned directly by autonomous UAVs, the middle ring draws sustenance from Wireless Charging Vehicles (WCVs), and the peripheral ring falls under the purview of WCVs tasked with extended-range replenishment missions. An indispensable thread running through these inquiries is the detection of prime charger siting coordinates—a puzzle that dovetails with the central thrust of the current manuscript. After all, if static chargers were positioned optimally inside motionless WRSNs, the network's energy requirements could be met in full, obviating any reliance on mobile chargers and the supplementary logistics they demand.

Parallels can be drawn between our own study and the investigation documented by Wang *et al*. [14], since both bodies of work grapple with the charger placement optimization dilemma within WRSNs, each pursuing a reduction in rollout expenditure concurrent with an amplification of the network's power delivery potential. A principal divergence, nevertheless, manifests in the weight we accord to the phenomenon of lopsided energy attrition among the network's constituent nodes. Uniquely, this work positions every node in harmony with the communication topology, undertaking to secure ideal energetic allotment via shrewd placement employing the fewest chargers attainable. In addition, while confronting the minimal-charger-deployment hurdle, deliberate steps are taken to sidestep the hazards of convergence upon local minima; it follows, then, that the concluding charger tally may fall short of representing the strict mathematical minimum. To this end, an unconventional synthesis is forged between the diversity-preserving traits of the NSGA-II algorithm and the global scouting faculties of reinforcement learning. Such a hybridization introduces novel challenges and amplifies the complexity of the solution procedure when set alongside the contribution in Wang *et al*. [14]. Another consideration is that a subset of network nodes fulfills relay functions and shoulders sizeable traffic burdens, potentially leaving their energy demands unmet. Accordingly, attempting to elevate aggregate charging efficiency while neglecting how energy is parcelled out among nodes proves entirely fruitless.

System Model

Network model

Depicted in **Figure 2** is a set of MUAVs-WRSNs within a volumetric (3D) setting. The system is built from a base station (BS), a squadron of UAVs labelled U = {u1, u2, u3, …, um}, and a collection of immobile sensing nodes marked as N = {i1, i2, i3, …, in}. The sensing nodes and the base station are assumed to reside on a shared planar surface, whilst the entire UAV fleet maintains a constant altitude throughout charging activities. The UAVs behave as makeshift fixed replenishment points, drawing upon their energy stores only when actively recharging the sensor nodes. The onboard power reservoirs of the UAVs are assumed ample enough to render energy scarcity concerns moot. The sensor nodes are dispersed without pattern across a two-dimensional expanse, the base station constituting the pivotal aggregation sink for data within said expanse. Every sensor node is obliged to dispatch data, which may propagate toward the base station either by a lone hop or through a chain of several hops. The UAVs can replenish numerous sensor nodes at the same time within their designated charging radius, and, reciprocally, any sensor node can harvest energy from several distinct UAVs. The principal notational conventions adopted throughout this paper are collated within **Table 2.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

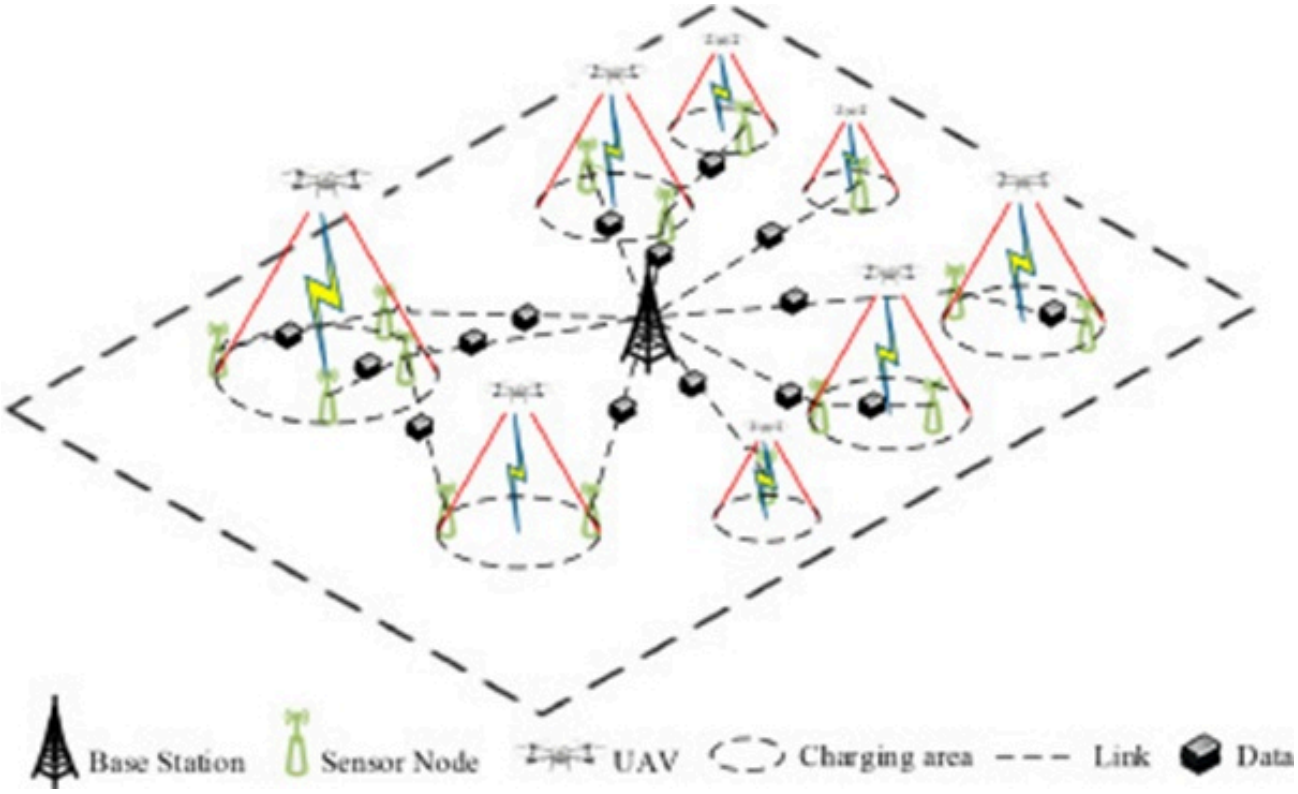


Figure 2. Network model.

Table 2. Description of key parameters.

Symbol	**Description**	***Symbol***	**Description**
N	Set of sensor nodes	U	Set of UAVs
n	Number of nodes	m	Number of UAVs deployed
i_n	n-th node location	u_m	m-th UAV center coordinates
R	UAVs Charging radius	h	UAVs charging suspension height
σ	Charging efficiency factor	P_t	Transmit power of UAVs
Con_i	Energy consumption rate of i_n	d_0	Transmission channel split distance
d_{max}	Maximum transmission distance	$du\ i$	Distance from i_n to UAV u_m

Charging model

When projected onto a planar coordinate system, the operative charging region of a UAV is represented as a disk whose geometric center coincides with the UAV's deployment site. In consequence, the coupling between a sensor node's incoming power and the outgoing power of the emitter can be captured through a Wireless Identification and Sensing Platform (WISP) [44]. Following this formulation, the coefficient of charging efficiency σ is definable via:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\sigma = \frac{G_t G_r \kappa}{L_p}\left[\frac{\omega}{4\pi\left(d_u^i + \beta\right)}\right]^2 \quad (1)$$

Within this construct, the symbols Gt, Gr, κ, Lp and ω correspond respectively to the antenna transmit gain, antenna receive gain, rectification efficiency, polarization loss, and the wavelength of the radio frequency carrier. The variable d_u^i captures the transmission-reception gap between UAV u and sensor node i_u. Incorporating the influence of the UAVs' hovering height h, $d_u^i = \sqrt{{d_{iu}}^2 + h^2}$, where d_{iu} denotes the horizontal offset from node i to UAV u. The factor β serves as an adjustment term adapting the Friis relation for near-field power conveyance. Consequently, except d_u^i , all remaining quantities are governed by circumstantial conditions and can be collapsed into constants. Thus, the charging efficiency operative between nodes and UAVs can be condensed to:

$$\sigma = \frac{\chi}{\left(d_u^i + \beta\right)^2} \quad (2)$$

Each sensor node retains the ability to scavenge energy from an array of UAVs. With the charger's transmission power Pt given, the charging efficiency Ei tied to sensor node in is obtainable through (3):

$$E_i = \begin{cases} \sum_{u=1}^{U} \sigma \cdot P_t & d_{iu} \le R \\ 0 & d_{iu} > R \end{cases} \quad (3)$$

Per **Eq. 3**, at the juncture where the transmission-reception spacing equals R, σ = 0. Hence, R signifies the outermost charging boundary, a value dictated by the hardware profile of the charger alongside its transmission power.

Energy consumption model of sensor nodes

Inside WRSNs, sensor nodes are tasked with continuously surveilling their allotted sensing territories and subsequently conveying gathered readings to a base station. A predominant fraction of energy outlay is incurred during the data-forwarding phase. The roles assigned to particular nodes depend on their physical distribution; nodes lying proximate to the BS might also be charged with relay duties, which entails accepting data streams originating from other nodes and thereafter passing them onward, bundled with their own harvested measurements, toward the base station. As a result, the energy draw of nodes is heterogeneous. To gauge the energy appetite of these nodes, we employ a widely adopted first-order radio model [4]. In accordance with this model, the energy toll for data transmission hinges upon the span traversed. In particular, for transmission gaps of considerable magnitude, the channel obeys a multipath fading paradigm; for more modest spans, it hews to a free-space channel paradigm. Hence, the energy toll Es entailed by dispatching a single bit of data is computable as:

$$E_s = \begin{cases} E_e + \varepsilon_{fs} \cdot d_{ij}^2 & d_{ij} < d_0 \\ E_e + \varepsilon_{amp} \cdot d_{ij}^4 & d_{ij} \ge d_0 \end{cases} \quad (4)$$

Within this notation, d0 designates the crossover threshold dividing the two channel regimes, roughly valued at d0≈87.7 m. The term dij specifies the separation from source point i to destination point j. εfs corresponds to the quantum of energy drawn by the transmitter amplifier when radiating one bit of data through free space, sized at εfs=10pj/bit/m2. In

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

parallel, εamp reflects the energy drain of the transmitter amplifier when sending a single bit under the multipath fading channel paradigm, quoted as εamp=0.0013pj/bit/m4. Moreover, Ee represents the energetic cost borne by the transmission circuitry in emitting one bit, whereas Er denotes the energy expenditure associated with receiving one bit; the two are taken as identical, i.e., Ee=Er=50pj/bit. From this, the holistic energy outlay Esr sustained by a relay node in receiving and subsequently retransmitting one bit of data is derivable as:

$$E_{sr} = \begin{cases} E_e + \varepsilon_{fs} \cdot d_{ij}^2 + E_r & d_{ij} < d_0 \\ E_e + \varepsilon_{amp} \cdot d_{ij}^4 + E_r & d_{ij} \geq d_0 \end{cases} \quad (5)$$

Problem Definition and Analysis

Within this segment, the MUAVs-OCCDP under examination is formally structured and dissected.

Problem definition

The system model sketches a network marked by the stochastic scattering of nodes, accompanied by persistent data traffic shuttling between these nodes and the base station. The charging efficiency at any particular node is a function of both the number of UAVs participating in the replenishment process and their geometric layout. Charging for every node is non-negotiable. Charging is effective whenever a node's charging efficiency is at or above its rate of energy usage. The twin ambitions of this inquiry are to pare down charging-related expenditures while elevating the degree to which network requisites are met. The following formal descriptions are advanced:

Definition 1: Every node exhibits a unique energy consumption rhythm governed by its spatial disposition, which subsequently sets the node's charging requirement.

Definition 2: Postulate that every node is required to forward its sensed data toward the base station, with the size of the sensed payload staying invariant across a standardized time slice, considered tantamount to a single data packet.

Definition 3: Nodes that fall under the umbrella of effective charging are catalogued as satisfied nodes.

In response to the obstacles inherent in deploying UAVs for nodal coverage, this manuscript introduces MUAVs-OCCDP. The overarching intent of MUAVs-OCCDP is to ascertain the sparsest UAV contingent able to deliver exhaustive nodal coverage, amplify the satisfaction quotient of WRSNs, and heighten the mean charging efficiency experienced by the nodes. The MUAVs-OCCDP edifice is erected atop a triad of hierarchically ordered aims: 1. Shrinking the UAV deployment count while safeguarding total nodal coverage; 2. Apportioning surplus energy to unsatisfied nodes to maximize the holistic network satisfaction percentage; 3. Augmenting the network-wide charging efficiency. The latter two aims are contingent upon the successful realization of the foremost aim.

To serve the primary aim, trimming the number of UAVs active within the network becomes imperative, formalized through a set wherein m captures the aggregate count of UAVs in operation. The inaugural objective function, denoted f1, is articulated as:

$$f_1 = m \quad (6)$$

In pursuit of a sharper energetic equilibrium coupled with an uptick in the charging efficiency directed at nodes still lacking satisfaction, the second objective function, f2, is crafted to fuse the tally of satisfied nodes with the mean charging efficiency for nodes that remain unsatisfied, expressible as:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$f_2 = n_{satisfy} + \frac{100E_{total}^{unsatisfy}}{k} \quad (7)$$

In the foregoing, $n_{satisfy}$ corresponds to the population of satisfied nodes, $E_{total}^{unsatisfy}$ captures the cumulative quantum of energy funnelled to unsatisfied nodes, and k conveys the count of unsatisfied nodes. Observably, once the full set of nodes attains satisfaction, $E_{total}^{unsatisfy}$=0. These twin constituents are intrinsically reciprocal in character.

To further augment nodal charging efficiency, the UAVs' coordinates must be drawn nearer to the nodes or made to encompass a more extensive node set. The tertiary objective function, referred to as f3, tallies the network's total charging efficiency and is prescribed by:

$$f_3 = \frac{100}{n}\sum_{i=1}^{n} E_i \quad (8)$$

Consistent with the optimization targets delineated, individual weights are apportioned to each objective function for normalization purposes, reflecting the graded importance of each goal. The aggregate evaluative metric f for benchmarking the efficacy of MUAVs-WRSNs is constructible as:

$$f = f_2 + \frac{f_3}{n} - 2f_1 \quad (9)$$

Worth underlining is that, to reinstate the energy levels of unsatisfied nodes, a fraction of the charging efficiency earmarked for satisfied nodes must be relinquished, an act that precipitates a decline in the network's overarching charging efficiency. A measure of friction thus arises between the second and third objectives. The ambition of the MUAVs-OCCDP is to accommodate the energetic consumption demands of the network by judiciously attenuating the weight attached to the third objective during the normalization step, drawing upon the mean charging efficiency of the nodes. Furthermore, since reducing deployment overhead aligns more closely with the aspirations of MUAVs-OCCDP, the negated form of the first objective is used in normalization and assigned a more substantial weight. The boundary conditions requisite for elevating the performance of MUAVs-OCCDP can be codified as:

$$Max(f). \quad (10)$$

$$s.t. \parallel u_m - i_n \parallel \leq R, \exists i_n \in N, u_m \in U, \quad (10a)$$

$$x_{\min} \leq u_x \leq x_{\max}, \quad (10b)$$

$$y_{\min} \leq u_y \leq y_{\max}. \quad (10c)$$

Wherein **Eq. 10** conveys the foundational purpose of this inquiry. Constraint **Eq. 10a** assures that all nodes are situated within the charging ambit. On account of the UAVs' altitude remaining invariant, strictly the two-dimensional projected coordinates of the UAVs warrant attention. Constraints **Eq. 10b** and **Eq. 10c** furnish additional validation concerning the siting of the UAVs.

Problem analysis

Drawing upon the multi-objective prioritization scheme and mapping connections to recognized NP-hard formulations, the NP-hard nature of MUAVs-OCCDP has been established beyond dispute.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

The earlier discussion flagged that the second and third objectives hinge upon the satisfactory completion of the first. A foundational tension among these aims spawns an irreducible Pareto trade-off embedded in the solution space, a phenomenon substantially fueled by the contest for finite spatial resources. Packing UAVs densely together amplifies localized charging performance, yet does so at the cost of widening the gap in energy provision across nodes. As was also observed, the trio of objectives follows a strict hierarchical ordering, with the third objective occupying the lowest rung of importance. Owing to this structure, MUAVs-OCCDP is bisected into two sequential stages: the inaugural stage accords primacy to slashing the deployment footprint, relegating network-wide charging efficiency to a supporting role; the later stage rests on the outputs of the preceding stage and turns its attention to raising the network satisfaction metric and closing the energy deficit of nodes still falling short. A tiered progressive resolution paradigm is deployed to disentangle the rivalry among the objectives as the solving unfolds, thereby furnishing an effective treatment of the multi-objective optimization predicament marked by mutually opposing aims.

Theorem 1. MUAVs-OCCDP is demonstrably NP-hard.

Proof: A canonical specimen of an NP-hard problem is the Minimum Circle Cover problem (MCCP). Given the MCCP formulation, one is confronted with a collection of planar points, designated $P = \{p_1, p_2, \ldots, p_n\} \subset \mathbb{R}^2$, a uniform covering radius r, and a coverage fraction δ. The MCCP task is to expend the fewest circles necessary to envelop the entire point set, subject to the condition that $\delta = 1$. In our setting, the charging disc of any UAV maps onto a planar circle of invariant radius, predicated on the assumption that all UAVs are equipped with identical charging hardware. The challenge becomes one of stationing the minimal UAV fleet such that the energetic demands of the nodes are met to the highest achievable degree. By fashioning an appropriate reduction, MUAVs-OCCDP can be re-expressed as a multi-objective extension of the MCCP. Since the MCCP has been conclusively classified as NP-hard, this transformation confirms that MUAVs-OCCDP inherits the same NP-hard complexity class. Hence, the problem's difficulty floor sits no lower than NP-hard.

Proposed Algorithm

This segment delineates the designed algorithmic solution. To begin, the communication fabric linking the network nodes is established, and each node's replenishment demand is logged. In alignment with the problem statement, the overall challenge is carved into two separate stages. The opening stage reframes objectives 1 and 3 from the problem description into a joint multi-objective optimization form, all the while guaranteeing that no node falls outside a charging footprint. A tailored NSGA-II variant is called upon to supply a starting configuration of deployment positions. Thereafter, a deployment-site tuning mechanism, built atop a two-phase reinforcement learning scaffold, is introduced to prune the UAV count still further and to maximally honour the energy requirements of the network nodes by steering the deployment positions through deliberate, learned adjustments. The macroscopic sequence of steps is visualized in **Figure 3.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

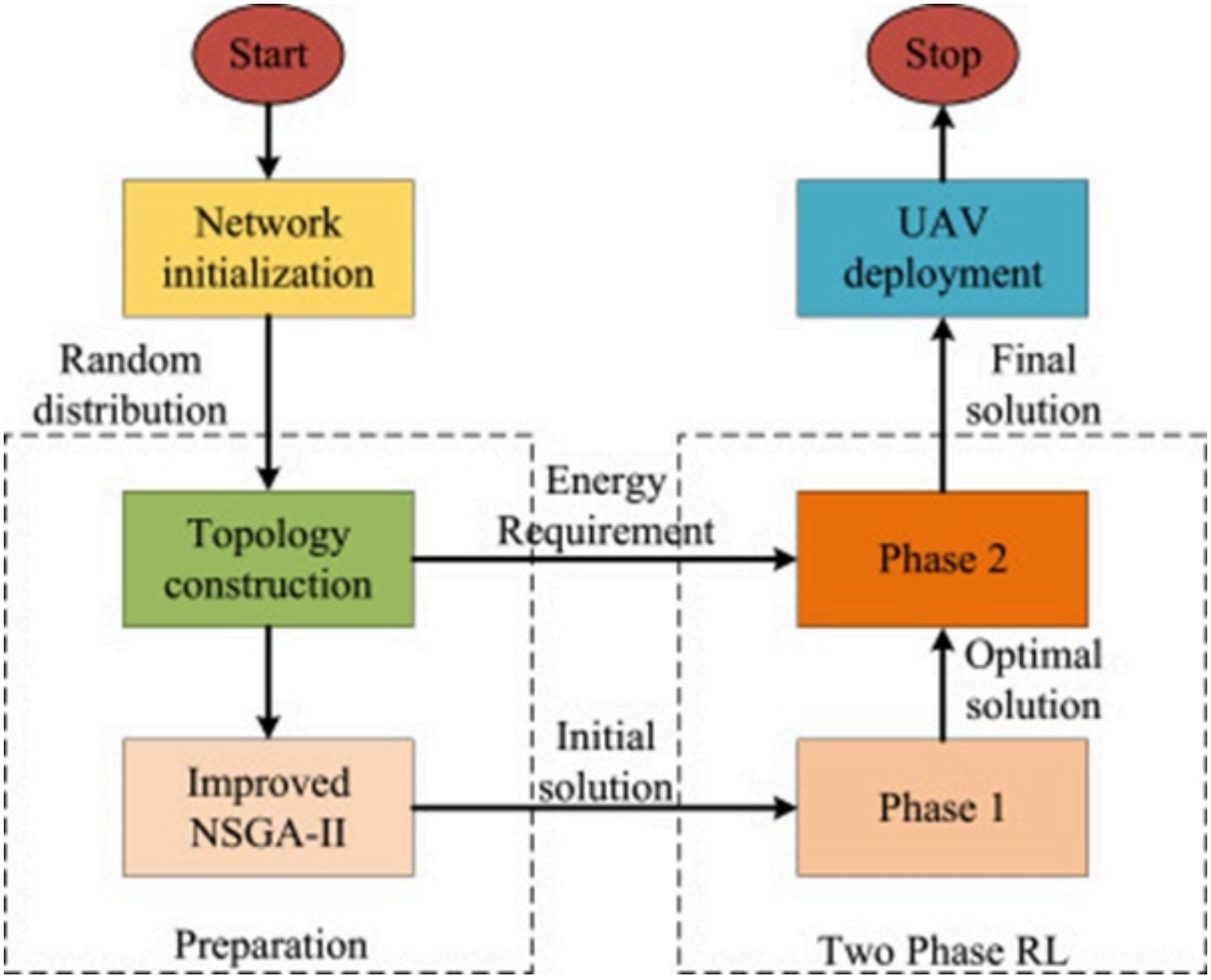


Figure 3. The workflow of the algorithm is proposed.

Communication topology construction

Assembly of the network's communication fabric capitalizes on the innate self-organizing dynamics of WRSNs. To bring the outcome into closer congruence with the behavioural profile of sensor network topologies, the minimization of communication-related energy drain was elevated as the chief priority to the greatest attainable extent; several boundary conditions are enforced: 1. A threshold on tolerable transmission range, written as dmax, is introduced; 2. A restriction on the upper count of concurrently connected neighbours, termed Cmax, comes into effect, signifying that a node cannot be the recipient of data streams originating from more than Cmax fellow nodes; 3. The measurements harvested by any node are obligated to progress along the most direct possible route toward the base station. Accordingly, nodes whose Euclidean separation from the base station satisfies di-BS ≤ dmax may establish a direct link, whereas those for whom di-BS > dmax must route their data traffic via a forwarding node, engaging in a multi-hop conveyance mode. To deliver network-wide nodal interconnection at the leanest energy budget, a modified variant of the Dijkstra algorithm, augmented with a load-levelling heuristic, is used to chart the data forwarding conduits, thereby preserving the shortest-path property throughout the construction of the multi-hop topology. The detailed mechanics of the topology-building procedure are captured in **Algorithm 1**.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Input: N= $\{i_1, i_2, i_3, \ldots, i_n\}$, D_{max}, C_{max}, d_0, E_e, E_r, ds
Output: Consumption [], Network topology, Total_consumption

BS= Network geometric center
Consumption [], Path []
For i in S do:
$d_{i\text{-}BS} = \sqrt{(x_{BS} - x_i)^2 + (y_{BS} - y_i)^2}$
If $distance_{i\text{-}BS} \leq D_{max}$ then:
Node i is connected to the BS, Record path to Path [].
Else:
Best_Path=Dijkstra (i, BS), recorded in Path [].
For each i in Path [] do:
$Downstream(i)$=BFS(i)
$Load(i)$=$|Downstream(i)|$+1
Calculate Consumption[i] using Eq. (10)
Total_consumption+=Consumption[i]
Return Consumption [], Network topology, Total_consumption

Algorithm 1. Topology construction.

Each node pushes one packet toward the base station per standardized time block, with every packet carrying a uniform payload volume of ds bits. A relay node shoulders not merely the transmission of its locally sourced data packets toward the base station but also the retransmission of packets arriving from its downstream dependents. Labelling the count of downstream dependents tethered to a given relay node as Downstream = n, the aggregate traffic load is expressible as Load = n + 1. For a leaf source node absent any relay responsibilities, the downstream count is zero. According to the energy dissipation model, the energetic toll for node i, represented by coni, is obtainable through **Eq. 11** as:

$$Con_i = ((n+1)E_s + nE_r)ds \quad (11)$$

Improved NSGA-II

The canonical NSGA-II framework is a Pareto-dominance-guided search heuristic that has repeatedly proven its merit when applied to multi-objective optimization problems. As problem scale and intricacy mount, however, the discriminating power of the Pareto dominance relation weakens, most conspicuously in scenarios characterized by prioritized objectives and elevated dimensionality. To more productively handle multi-objective settings in which objectives carry explicit precedence rankings, and to reconcile the competing pulls of convergence pressure and solution spread, Tian *et al*. [22] devised the strength dominance relation (SDR). This construct embeds an adaptive niche-preservation scheme informed by the angular separation among candidate solutions. Confined within each niche, the sole solutions granted non-dominated standing are those exhibiting premier convergence quality. A prerequisite step, then, is condensing the multiple objective scores into a unified strength metric within the multi-objective optimization landscape.

Under the scheme advanced in this paper, objective normalization is first performed by attaching differentiated weighting coefficients to each goal. If a candidate solution leaves any network node beyond the charging envelope, the strength penalty affixed to that candidate is driven to infinity. Let the coefficient associated with the UAV count be denoted a, and the coefficient linked to charging efficiency be denoted b. Over the course of the evolutionary loop, if two individuals

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

happen to register equal strength metrics, the standard Pareto dominance check is invoked as a tiebreaker. Once objective priorities are locked in, reasonably calibrated weights should virtually guarantee that strength ties do not materialize. The illustration in **Figure 4b**, corresponding to a and b values of 1000 and 1, respectively, exhibits the strength outcomes following non-dominated sorting. Consequently, individuals with diminished strength metrics have a higher probability of propagation, with the strength metric computed according to **Eq. 12**.

$$Strength = \begin{cases} am - bE & C_{rate} = 1 \\ \infty & C_{rate} < 1 \end{cases} \quad (12)$$

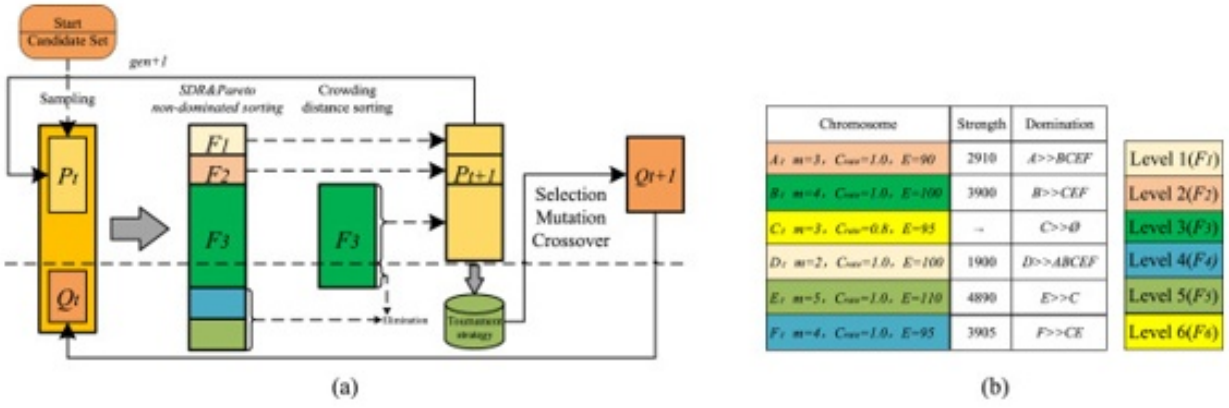


Figure 4. (a) The program flowchart of the improved NSGA-II, and (b) an example of SDR-based non-dominated sorting.

Additionally, the standard elitist preservation operator is set aside to pursue the highest-quality solution across successive generations. Owing to its design, the SDR confers comparatively loftier strength values upon high-calibre individuals, thereby operating as a latent preservation force. This design feature ensures that improved individuals ascend to a more prominent rank through the combined influence of non-dominated sorting and crowding distance [13], smoothing the passage of elite individuals into ensuing generations. The crowding distance computation follows the prescription of **Eq. 13.**

$$CrowdDis = \sum_{o=1}^{K} \frac{G_O(X_{i+1}) - G_O(X_{i-1})}{G_O^{Max}(X) - G_O^{Min}(X)}, i = 2, 3, \ldots, W - 1. \quad (13)$$

In this expression, K denotes the dimensionality of the objective vector, i.e., the count of fitness functions. O captures the value realized by a given function on a chromosome X, i indexes the entire chromosomal population, and W refers to the concluding chromosome of that gene sequence.

To guard against premature stagnation, a tournament-based selection mechanism is activated throughout the reproductive stage. This mechanism skews selection pressure toward parents distinguished by higher fitness scores and broader dispersion across the objective landscape. It acts as a supplementary safeguard. Within each non-dominated tier, individuals occupying comparatively barren regions of the objective space are preferentially preserved, meaning that those burdened with smaller crowding distances are more readily discarded when the population budget must be respected.

Because the nodes composing the network follow a random spatial distribution, the initial population must be enriched to quicken the algorithm's convergence trajectory. Each node is treated as a focal point, with its energy reception footprint and the UAV charging footprint modelled as circles of equivalent radius. The geometric centroid of any intersection zone formed by these circles is appended to the starting population as a plausible deployment coordinate. At the same time,

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

spatially isolated nodes are injected without further processing. It should be underscored that for large-scale network instances, the cardinality of the starting population can explode with the insertion of additional nodes, posing a threat to the algorithm's convergence speed. To neutralize this threat, a down-sampling mechanism is enforced when collecting intersection centroids, combined with a hard cap on the maximum admissible number of candidate sites.

Additionally, the mean shift algorithm is used to perform adaptive pre-clustering of the network, with its bandwidth parameter synchronized with the charging radius. The resulting cluster centroids are subsequently incorporated into the starting population to bolster the legitimacy of the contained individuals. Candidate deployment locations within the population are represented via a binary encoding scheme, with 0 signalling the absence of deployment and 1 signalling its presence. The core genetic operators of selection, mutation, and crossover native to the NSGA-II architecture are preserved without modification. The full particulars of the enhanced NSGA-II procedure are available in **Algorithm 2** and **Figure 4a.**

Input: Initial population P_0, Crossover rate *Cr*, Mutation rate *Mr*, *Iteration*, N
Output: final *P*

Network initialization;
Cluster center=Mean shift(N);
Intersection center = Sampling (N);
$P_0 \leftarrow$ *Cluster center*$\cup$*Intersection center*;
Normalize the objective [*m*, C_{rate}, *E*] value in P_0;
F [F_1, F_2...F_n]$\leftarrow$Do non-dominated sorting on P_0 by SDR;
CowedDis$\leftarrow$Crowding Distance (*F*);
while *gen* < *Iteration* **do**
$P_1 \leftarrow$ Tournament Strategy (*F*, *CowedDis*) in P_0;
offerspring=Crossover (P_1, *Cr*);
offerspring= Mutation (*offerspring*, *Mr*);
family = *offerspring* + P_0;
F [F_1, F_2...F_n]$\leftarrow$Do non-dominated sorting on P_0 by SDR;
CowedDis$\leftarrow$Crowding Distance (*F*);
k$\leftarrow$Minimum value s.t. $|F_1 \cup F_2 \cup \ldots \cup F_k| \geq P_1$
If $|F_1 \cup F_2 \cup \ldots \cup F_k| > P_1$ then
$|F_1 \cup F_2 \cup \ldots \cup F_k| - P_1$ solutions from F_k with the worst *CrowdDis*
$P \leftarrow |F_1 \cup F_2 \cup \ldots \cup F_k|$
gen+1
end while
return final *P*

Algorithm 2. Improved NSGA-II.

Adaptive deployment location adjustment algorithm based on two-phase reinforcement learning

An enhanced UAV positioning arrangement emerges from the improved NSGA-II procedure. Nevertheless, that algorithm's reliance on its starting population, coupled with its inherently probabilistic nature, yields configurations that

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

represent little more than rough approximations of the genuine optimum. Thus, paring back the UAV contingent while simultaneously satisfying the energetic appetite of the network's nodes necessitates a further stage of location refinement.

The initial tier of the reinforcement learning architecture invokes Q-learning to iteratively finesse the placement coordinates, pursuing a globally-oriented enhancement rooted in the seed solutions furnished by NSGA-II. The driving ambition here is twofold: trimming the charger count and amplifying charging effectiveness. The subsequent tier ushers in a multi-agent learning paradigm. This tier confronts the heterogeneous energy drain across relay nodes—a direct outgrowth of the communication topology—by remedying the deficits of select high-consumption nodes, all while preserving the replenishment state of the broader node population. The two tiers are woven together in a sequential pipeline. The overarching schema by which deployment positions are recalibrated via the two-phase reinforcement learning methodology is rendered in **Figure 5.**

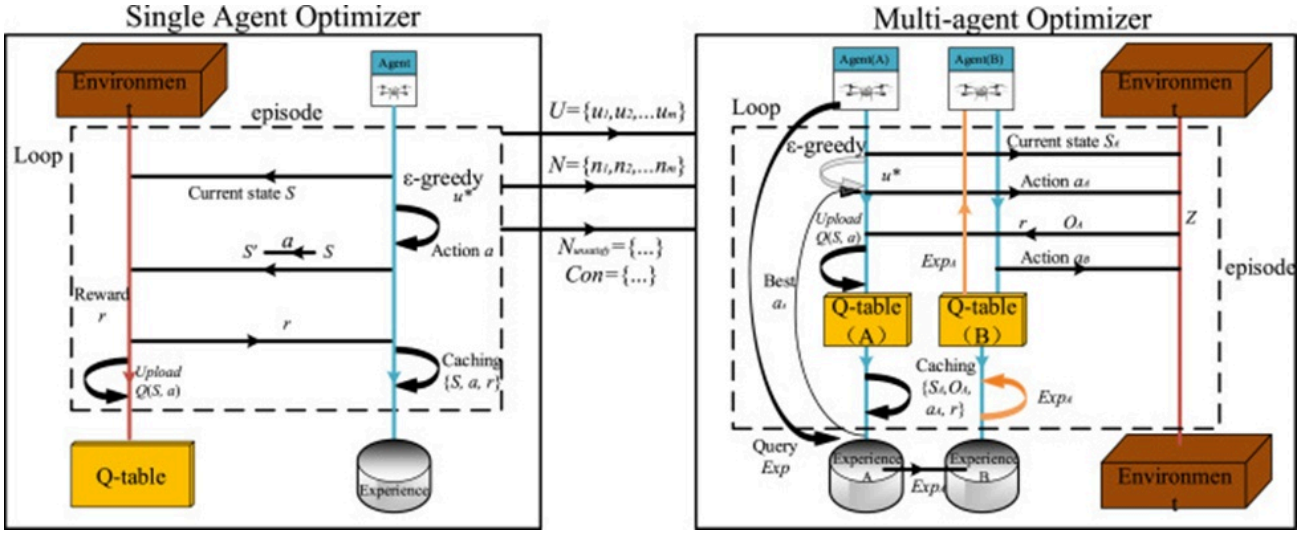


Figure 5. Two-phase reinforcement learning algorithm flow chart.

Inside the dual-phase reinforcement learning construct, the agent governs the UAV coordinates to reduce charging outlay and converge upon a deployment configuration harmonized with the network's energy consumption demands. The planar region is first carved into a dense grid of unit-dimension cells, with every deployment centre and node situated precisely at a cell's centroid. Relocation of a deployment centre is thus conceptualized as a modality of network mobility. Even though the optimization criteria diverge between the two reinforcement learning layers, the agent is presented with an unchanged repertoire of directional moves, each selected with equal odds. The permissible headings for motion across the 2D surface are consolidated into a set of eight discrete choices, as illustrated in **Figure 6a**. Consequently, reinforcement learning within either layer operates under a confined state space alongside a confined action space, permitting its formulation as a textbook Markov Decision Process (MDP). Reinforcement learning has consistently proven its mettle in MDP settings, with its central function being to shepherd the agent toward an ever-improving behavioural policy through environmental reward feedback. In sprawling network deployments, the stock of seed solutions balloons in tandem with network scale, swelling the agent's cumulative learning load; meanwhile, the experience replay buffer remains bounded in size. To avert overwriting hard-won earlier experiences due to repeated forays into identical states, it becomes necessary to continually refresh and preserve high-value experiences as learning progresses within each tier. Doing so fortifies the agent's capacity to endure protracted learning episodes. The deployment site recalibration algorithm, underpinned by the two-phase reinforcement learning framework, capitalizes on these retained experiences to attain a stabilized learning regime, attenuating the disruptive influence of randomness, sharpening the agent's exploratory acuity, and ultimately hastening the pace of learning.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

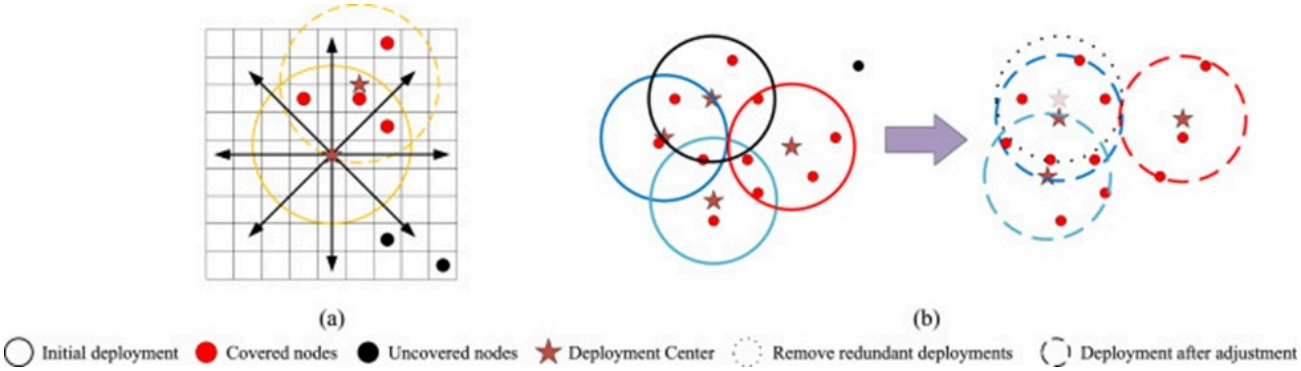


Figure 6. (a) Diagram of deployment location adjustment, and (b) an example of multiple deployment location adjustments.

Phase 1: single-agent reinforcement learning

The opening optimization tier adopts the standard Q-learning protocol, which shares the same end goal as **Algorithm 2** to push the seed solutions further forward. As displacement unfolds, distinct deployment coordinates correspond to distinct states, and the deployment centres elect a heading with uniform probability. Upon executing an action, the environmental state transitions and the agent is furnished with a reward that reflects the nature of that transition. The agent's overarching motivation is to maximize the cumulative reward stream. To broaden the agent's exploratory horizons and reduce the likelihood of entrapment in a local optimum, specific guardrails are placed on the repositioning of deployment sites:

1. Throughout the movement process, the deployment positions must at all times retain those nodes that enjoy only a single layer of coverage, to ensure that no node ever slips outside any charging footprint.

2. The count of nodes under coverage is updated without delay whenever a node exits or enters the charging boundary, guaranteeing that the coverage status of every node stays current.

The historically optimal configuration is bookmarked, and once the step budget is exhausted, the exploration episode ends with a reversion to that historical best. Furthermore, once the optimization cycle completes, any deployments rendered redundant are culled. As depicted in **Figure 6b**, repositioning of deployment sites proceeds under the condition that the affected nodes are confirmed to be enclosed by no fewer than one deployment footprint. If the initial coverage relationships were to be rigidly enforced, the agent's exploration domain would shrink considerably, potentially imprisoning it in a local optimum.

In consequence, deployment sites marked in red can extend their reach to previously unserved nodes, rendering the deployment site marked in black superfluous after adjustment and hence subject to removal. Every deployment site contributes to refreshing the global coverage ledger as repositioning occurs; this ledger is inherently global and tethered to the MDP. The agent's learning trajectory within phase 1 can therefore be formally expressed as a five-tuple {S, A, P, R, γ}, with S signifying the state space, A standing for the action space, P representing the policy, R being the reward function, and γ the discount factor. The reinforcement learning setup for this opening phase is elaborated as follows:

Agent: Guided by a reinforcement learning algorithm, the agent constructs the mobility policy for the deployment centre through continuous interaction with the WRSNs.

Action: This encompasses the set of behaviours the agent may exhibit while interacting with the environment. Given that the initial position under optimization already resides in the vicinity of a strong solution, a large step increment risks

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

overshooting the optimum; consequently, the step increment is limited to one unit. The movement repertoire, defined by directional bearings, is as follows: A = {(0, 1), (0, 1), (1, 1), (1, 1), (1, 1), (1, 1), (1, 0), (1, 0)}.

State: The state encompasses the UAV identifier together with its coordinates, the coverage condition of the nodes, and their energy condition. The coverage condition specifies which deployment sites cover which nodes, while the energy condition captures the charging effectiveness received by the nodes. The state space is definable as *S* = {*UAVs_idx*, (x_u, y_u), *node_status, energy_status*}.

Policy: Denoted as P=μ{x}, this notation signifies the linkage between states and the actions selected. The agent charts its course by adhering to the optimal movement policy and adjusts its coordinates accordingly. An ε-greedy behavioural strategy is implemented, with the exploration fraction tapering dynamically across successive iterations. In the exploration-dominant phase (early), actions are drawn randomly with likelihood ε, whereas in the exploitation-dominant phase (late), the action boasting the highest Q-value is favoured.

Reward: The reward (R) serves as evaluative feedback delivered after the agent's chosen action, measuring the merit of that action. The signal can be either reinforcing or penalizing. To shrink the number of deployed units while upholding complete nodal coverage, the average charging efficiency enters the calculation. The reward function is laid out in **Eq. 14:**

$$R = 50n_{cover} + 500n_{new} + 300\frac{E}{m} \quad (14)$$

Within this expression, n_{cover}, n_{new}, E and m correspond respectively to the count of nodes encompassed by a deployment site, the count of nodes freshly encompassed, the overall charging efficiency, and the total number of deployments. Given the premium placed on maximizing nodal coverage, additional weight is assigned to n_{new}. Furthermore, should a network node fall entirely outside the coverage umbrella, an infinite penalty is imposed, resulting in a reset to the initial position. The agent's mission is to refine the deployment site via policy μ to extract the maximal accumulated reward.

Throughout the learning of the agent's motion policy, a sequential displacement protocol is observed. To counteract the possibility that a deterministic selection order might channel convergence toward a local optimum, the sequence in which deployment sites are chosen for adjustment is randomized. The agents draw upon a shared Q-table [14] that records the valuation linked to each state, termed the Q-value, which informs the choice of action a. To keep memory usage lean, only those states actually encountered are retained. The Q-value is iteratively updated according to **Eq. 15** shown below.

$$Q(s,a) = Q(s,a) + \alpha\left[R_a + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \quad (15)$$

In this equation, α, Ra, and γ designate the learning rate, the instantaneous reward, and the discount factor, respectively. Fixing α at an overly elevated level can introduce volatility into the learning path and weaken the retention of prior knowledge. Conversely, setting it too low can retard algorithmic convergence. A larger γ tilts the agent's preference toward distant future rewards. The optimal state is therefore characterized as the valuation that jointly maximizes immediate gain and the discounted stream of future benefits, as formalized in **Eq. 16.**

$$Q^*(s,a) = R_a + \gamma \max_{a'} Q(s',a') \quad (16)$$

Upon convergence of the Q-table, a milestone signifying arrival at the exploitation regime, the action-selection policy will systematically favour the action linked to the supreme Q-value, which can be expressed as:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$u^{*} = \text{argmax}Q^{*}(s,a) \quad (17)$$

Phase 2: multi-agent reinforcement learning

What follows is a reinforcement learning optimization routine constructed around a collaborative multi-agent design aimed at shifting deployment sites so that the energetic needs of every node receive the best attainable accommodation. Much like the single-agent setup, every deployment site functions as a distinct agent that progressively shapes a movement strategy, with motion parameters held invariant from one agent to the next. The bond uniting the agents rests on cooperation, not rivalry. To sidestep the hazard of mutually blocking competitive dynamics typical of multi-agent optimization landscapes, the Q-table is refreshed in parallel, with each agent preserving its own separate Q-table coupled to its private reservoir of past encounters. Additionally, stored encounters are broadcast across agents via the cooperative pipeline, allowing contributions to flow among them. As sketched in **Figure 7**, a given deployment site keeps track of several neighboring nodes that remain unserved and coordinates its actions with surrounding deployment sites. While shifting sites, pre-existing constraints must stay unviolated, and nodes already labelled as satisfied must retain that status. Moreover, the deployment count inherited from phase 1 is treated as irreducible, obviating any pruning of surplus deployments throughout phase 2.

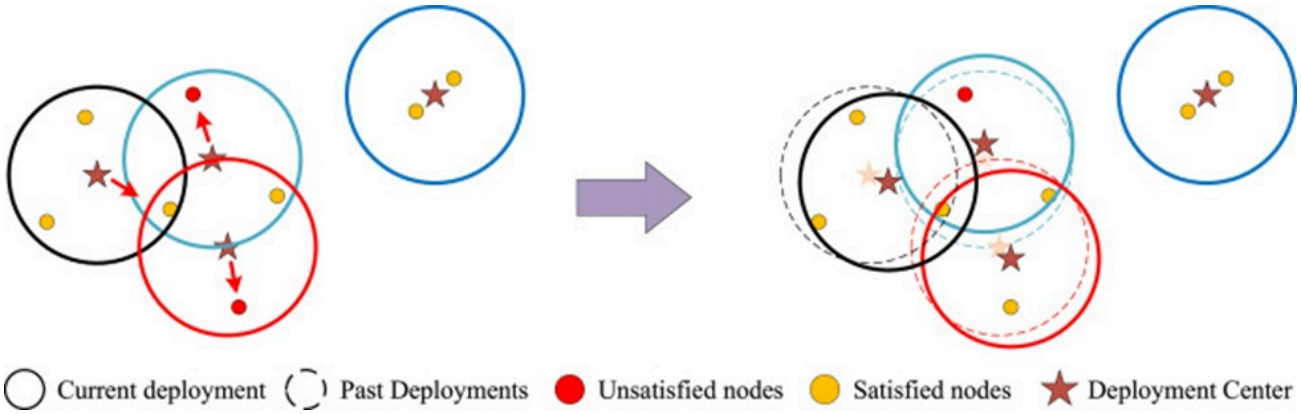


Figure 7. Example of multi-agent cooperative adjustment of position.

At this juncture, the state attached to any single deployment site bundles together data about a limited subset of fellow sites and unserved nodes, thereby qualifying the setup as a POMDP. It lends itself to description through a seven-tuple {S, A, O, P, R, Z, γ}. Set against phase 1, the observation space O together with the observation function Z make their entry, while the remaining elements endure unchanged. Calibrated to the revised target specifications, the multi-agent reinforcement learning formulation is explicitly characterized below.

Agent: The design proposed here rests on the premise that agents ought to swap experiential records among themselves to trim superfluous exploration efforts and accelerate the tempo of learning. The experience store undergoes a bisection, with fifty percent of its capacity reserved for elite experiences and the remaining fifty percent earmarked for freshly harvested experiences.

State: The state space assembles exhaustive intelligence on deployment sites, the coverage linkages connecting nodes to these sites, plus the quantum of energy absorbed by each node. In formal terms, the state space is expressed as $S = \{(x_u, y_u), (i, j, \ldots, k)_u, (E_i, E_j, \ldots, E_k)\}$.

Action: To reconcile the competing imperatives of exploratory breadth and fine-grained search, step increments are partitioned into larger jumps and smaller refinements. Combined with the eight admissible displacement headings, the action space crystallizes as: A = {(0, 5), (0, 5), (5, 0), (5, 0), (5, 5), (5, 5), (5, 5), (5, 5), (0, 10), (0, 10), (10, 0), (10, 0), (1, 0), (1, 0), (0, 1), (0, −1)}.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Observation space: An observation bundles together the positional coordinates of the deployment site (x_{u1}, y_{u1}), descriptive data on proximate unsatisfied nodes {x_i, y_i, E_i, Con_i}, and details on fellow deployment sites (x_{u2}, y_{u2}). The symbol Ei denotes the node's instantaneous charging efficiency, whereas Coni encodes its energy draw rate. In light of the high-dimensional state space—a trait that complicates straightforward Q-table storage—a coarsely binned discretization is imposed. By picking a bin width sensibly matched to the network's magnitude, a tolerable sacrifice in fidelity is accepted. Thus, the observation space is rendered as $O = \{(x_{u1}, y_{u1}), \{x_i, y_i, E_i, Con_i\}, (x_{u2}, y_{u2})\}$.

Policy: The behavioural strategy follows an ε-greedy template, the exploration propensity exhibiting a dynamic decay as iteration cycles accumulate. Beyond this, experiential steering is embedded into the process, conferring upon agents a certain likelihood of internalizing the most effective experiential moves demonstrated by their counterparts.

Observation function: The observation function Z ($O|$ s', a) quantifies the chance of registering observation O upon transitioning to s' following the enactment of action a, and is treated as a deterministic mapping.

Reward: The incentive architecture underpinning the multi-agent system bifurcates into global and local reward currents, each encompassing both reinforcing and penalizing elements. The global reward originates from the satisfied-node headcount $n_{satisfy}$ together with the cumulative energy upswing ΔE_{gain} recorded among those nodes still classified as unsatisfied, as derived in **Eq. 18.**

$$r_{global} = 5000\frac{n_{satisfy}}{n} + 1000\Delta E_{gain} \quad (18)$$

$$\Delta E_{gain} = \sum_{n_{unsatisfy} \in N} \left(E^u_{cur} - E^u_{init}\right) \quad (19)$$

Throughout the optimization trajectory, it remains mandatory that nodes already in the satisfied category stay satisfied and that no node drops out of universal coverage. Any transgression of this condition draws an unbounded penalty in **Eqs. 18 and 19**, nsatisfy stands for the satisfied-node tally, whereas n denotes the total node count. *Eu cur* and *Eu init* respectively capture the baseline and current aggregated charging efficiencies for the set of unsatisfied nodes.

To channel localized refinement of deployment sites, the local reward is sculpted by the incremental benefit each agent delivers to the unsatisfied nodes. The recipe for computing the local reward is supplied in **Eq. 20.**

$$r_{local} = 500\sum_{i=1}^{N}\sum_{u=1}^{U}\frac{E^i_u}{Con_i} \quad (20)$$

Striking a workable equilibrium between localized and system-wide priorities calls for a blended incentive formulation that fuses local and global signals, thereby shepherding the agent along a trajectory aligned with the global optimum. The hybrid reward is obtainable via **Eq. 21.**

$$R = \lambda r_{global} + (1-\lambda) r_{local} \quad (21)$$

Within **Eqs. 20 and 21**, λ serves as the collaboration weighting factor. Ei u quantifies the charging efficiency dispensed by UAV u to node in, and Coni reflects the energy consumption tempo of node in.

Updating of an agent's Q-table proceeds analogously to the single-agent regime, the Q-value being derived per **Eq. 15.** Experiential interchange among agents transpires, meaning action choice is steered not exclusively by the agent's own Q-table but also by the fruitful experiential archives of its peers. The movement policy is prescribed as:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$u^{\bullet} = \begin{cases} argmaxQ^{\bullet}(s,a) & 1-p \\ Exp & p \end{cases} \quad (22)$$

In **Eq. 22,** p denotes the probability of selecting to draw upon the premier experiences stored by fellow agents, and Exp denotes that premier experience.

Experimental Results and Analysis

This portion begins by cataloguing the parameter selections embedded in our experimental pipeline. Subsequent to that, a convergence diagnostic is executed on the enhanced NSGA-II algorithm to home in on the most favourable mutation and crossover settings. We additionally probe a handful of baseline heuristic schemes, namely the K-means algorithm (KM) [14], Greedy algorithm (Greedy) [16], Mean Shift algorithm (MS) [28], and Genetic Algorithm (GA) [13], some of which are regularly coupled with machine learning pipelines to churn out seed solutions. Through a side-by-side comparison of the outputs yielded by the improved NSGA-II and these heuristic baselines, we corroborate the algorithm's performance edge and lay out the reasoning behind selecting the enhanced NSGA-II as the seed-solution engine. One should note that every algorithm is assessed under matching node scatter patterns, and the seed populations fed to the Greedy, modified NSGA-II, and GA algorithms are kept uniform.

The algorithm we propose is brought to bear on the MUAVs-OCCDP. To render its performance more readily comparable, we pit it against a roster of up-to-date sophisticated approaches evaluated under parallel conditions. The roster spans the effectiveness and efficiency of the RL-CCD [14], Greedy algorithm [19], FCDA [28], and IPECS [26] algorithms when tasked with solving MUAVs-OCCDP, plus a variety of assessments of the fluctuation band of the network satisfaction rate throughout our UAV placement process. As a culminating step, we further lift the network satisfaction rate through a topology rebuild and display the rebuilt outcomes. The experimental workflow in its entirety was executed inside PyCharm 2023.2.1, running on a machine powered by a 13th Gen Intel(R) Core (TM) i5–13,400 CPU.

Experimental setup

Herein, the parameter configurations applied within the simulation trials are enumerated. In the standard scenario, n nodes are scattered stochastically throughout a square planar patch measuring 100 m by 100 m, the scattering process itself being steered by random seeds. The radiated power of the UAVs is locked at 3 W, aligning with the output rating of the TX91501 transmitter unit sourced from Powercast, as referenced by [45]. One should acknowledge that a higher radiated power or the adoption of more performant hardware could lift charging efficiency. Additionally, the quantities χ and β, which act upon charging efficiency, are pegged at 36 and 30, respectively, consonant with earlier reports [45, 46]. Wang [14] gauged the functional charging reach of the equipment under matching radiated power and settled on a charging radius of 10 m. In the context of this study, the UAVs are held at a hovering height of 5 m; it is worth remarking that a reduced hovering height would compress the charging separation between UAVs and nodes, thereby boosting charging efficiency. Every node pushes a data packet toward the base station at uniform temporal intervals. Experimental measurement and subsequent calculation revealed that when the packet payload is fixed at 106 bits, the energy draw rate of nodes occupies the same order of magnitude as the charging efficiency figures. With respect to the algorithmic architecture, the seed population size is set to 10, and the iteration budget is capped at 1500. The mutation and crossover rates are finalized in light of the convergence diagnostics that follow. The subsection also records the essential parameters belonging to the two-phase reinforcement learning machinery. A consolidated overview of the parameter assignments can be found in **Table 3.**

Table 3. Simulation parameters.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Symbol	Description of parameter	Value
Z	Network area size	Z = 100 m
n	Total number of nodes	100–300
R	UAV charging coverage radius	10 m
χ	Charging efficiency coefficient	36
β	Adjustment (correction) factor	30
C_r	Crossover probability	1.0
M_r	Mutation probability	0.9
gen	Number of NSGA-II iterations	1500
P_0	Initial population size	10
P_t	Transmission power	3 W
Episodes	Number of reinforcement learning episodes	1000
Maxsteps	Maximum RL steps per episode	200
α	Learning rate	0.1
γ	Discount factor	0.9
ε	Exploration rate (epsilon)	1.0
λ	Synergy coefficient	0.8
d_{max}	Maximum allowable communication distance	100 m
C_{max}	Maximum connection coefficient	5
ds	Data packet size	10^6 bits

Remark 1: Addressing the matter of coordinate precision calls for rescaling the network footprint down to the centimetre range, specifically 1000 cm × 1000 cm. In tandem, when computing charging efficiency alongside node energy draw, the unit of length ought to be switched to metres to preserve the fidelity of the numerical outputs.

Experiments with improved NSGA-II

Convergence test of the improved NSGA-II algorithm

The convergence behaviour of the NSGA-II algorithm is heavily shaped by the crossover rate (Cr) alongside the mutation rate (Mr). Within this investigation, convergence is gauged through the lens of the resulting deployment count under a spectrum of Cr and Mr pairings, with the network node tally locked at 100. The baseline iteration cap and population size are fixed at 1500 and 10, respectively. Owing to the fact that the seed population already comprises viable individuals, setting Mr and Cr at excessively low values can stifle the breadth of the population and hamper the safeguarding of elite individuals. For this reason, only test configurations where Mr ≥ 0.5 and Cr ≥ 0.5 are taken into consideration. Each parameter underwent separate scrutiny, beginning from a value of 1.0 and tapering downward in steps of 0.1. The

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

outcomes, depicted in **Figure 8**, reveal that with Mr pegged at 0.9 and Cr at 1.0, the enhanced NSGA-II variant displays the strongest convergence profile and delivers the sparsest deployment.

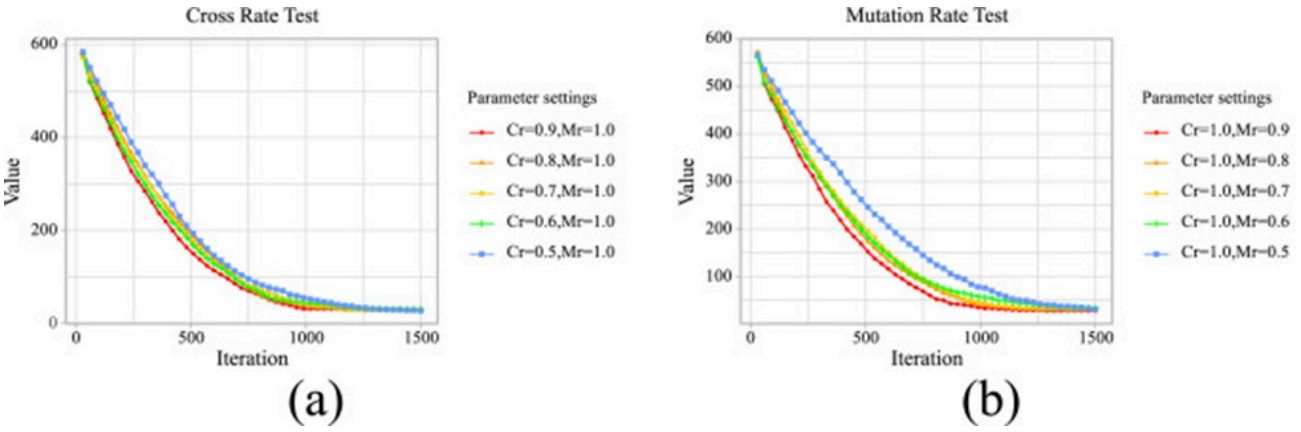


Figure 8. (a) Mutation rate test; (b) cross rate test.

Comparative experiments of the pioneer algorithm

The enhanced NSGA-II procedure was measured against counterpart methods regarding the UAV fleet size, under uniform node arrangements and seed population settings. As portrayed in **Figure 9**, the enhanced NSGA-II method as well as the Greedy method both exhibit pronounced strengths in curtailing deployment numbers. That said, the Greedy approach is widely understood to latch onto the locally best candidate at each step with no capacity to reverse earlier decisions, a trait that renders it susceptible to the "local optimum" pitfall. Beyond this, the configurations produced by the Greedy scheme suffer from homogeneity; when supplied as a starting point for reinforcement learning refinement, they squeeze the exploration envelope, making escape from local optima difficult. By contrast, the enhanced NSGA-II scheme fosters a widely dispersed seed configuration via its diversity-preserving machinery, thus equipping reinforcement learning with a stronger capacity to sidestep convergence upon any single local optimum and to progress more swiftly toward the global optimum. Additionally, superior seed configurations can markedly lower the trial-and-error burden borne by reinforcement learning and shrink the associated computational overhead. It is also noteworthy that the enhanced NSGA-II approach boasts wider applicability and stronger scalability across single-objective, multi-objective, and priority-ranked multi-objective optimization settings alike.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

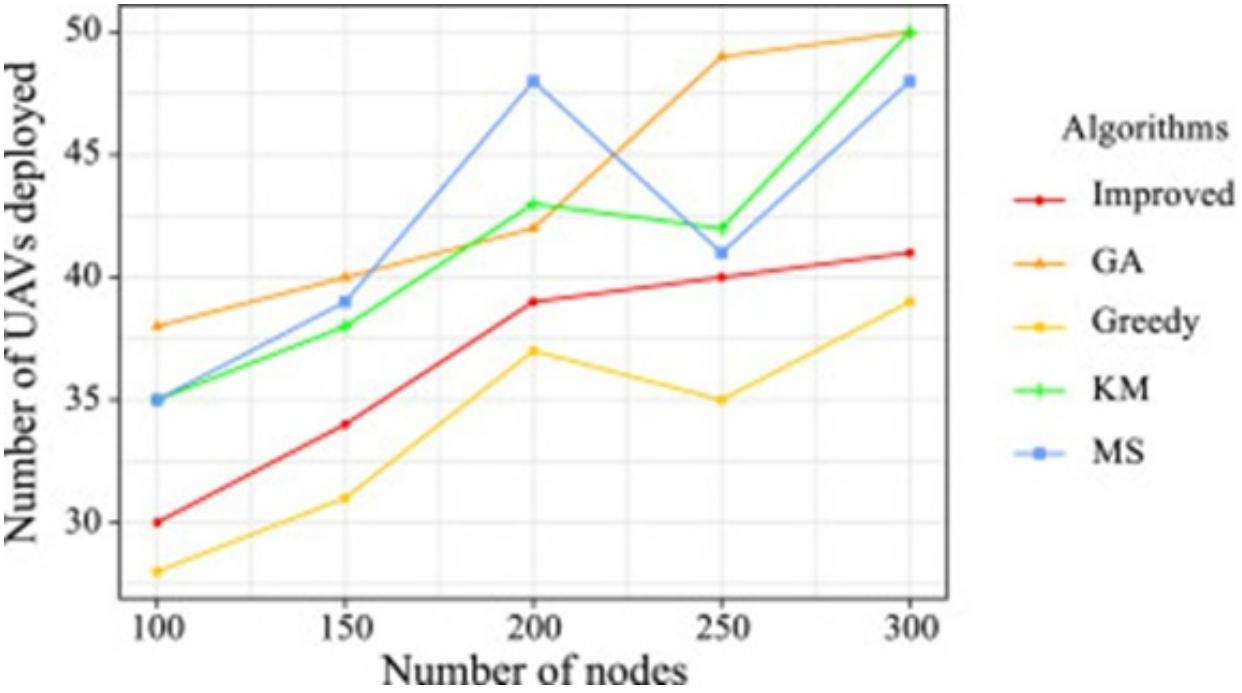


Figure 9. Comparison of the deployment number of the pioneer algorithm under different node distributions.

Under matching conditions, the seed configurations generated by both the Greedy and enhanced NSGA-II routines were fed into phase 1 of the reinforcement learning (RL) procedure for further refinement. As summarized in **Table 4**, even though the Greedy routine yielded the leanest seed configurations, its optimized deployment counts ended up trailing those achieved via the enhanced NSGA-II pipeline. Although the Greedy method can, on occasion, land upon a global optimum, it frequently becomes ensnared in a local one. In juxtaposition, pairing the enhanced NSGA-II method with reinforcement learning demonstrates greater efficiency and raises the odds of breaking free from local optima. The average charging efficiency figures obtained after RL-based refinement underscore the indispensability of employing RL for optimization, irrespective of which algorithm supplied the seed configuration.

Table 4. Data comparison between Improved NSGA-II + RL and Greedy + RL.

Nodes	Average charging efficiency of UAVs (Greedy/RL)	No. UAVs deployed (Improved-NSGA-II/RL)	Average charging efficiency of UAVs (Greedy/ RL)	No. UAVs deployed (Greedy/RL)
100	0.338W/0.381W	30/24	0.348W/0.37W	28/26
150	0.426W/0.48W	34/28	0.454W/0.509W	31/29
200	0.532W/0.603W	39/31	0.55W/0.626W	37/32
250	0.641W/0.764W	40/33	0.684W/0.768W	35/33
300	0.833W/0.913W	41/32	0.787W/0.913W	39/32

Comparison of the proposed algorithm with other state of the advanced algorithms

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

In this part of the study, the algorithm put forward is held up against a selection of sophisticated algorithms originating from other research groups, as depicted in **Figures 10 and 11**. The presented algorithm displays a sizeable edge in terms of both the number of fielded UAVs and the mean charging efficiency delivered by those UAVs, comfortably outstripping the competing advanced methods. The RL-CCD scheme, while likewise aspiring to trim deployment counts and lift charging efficiency, demands that reinforcement learning-based optimization adjust positions in relation to the originally covered nodes. Such a requirement constrains the agent's exploration territory, fostering a tendency to settle into a local optimum. The backbone of the FCDA strategy rests on NSGA-II, an algorithm overly dependent on its seed population and in need of additional polishing. The IPECS method, meanwhile, when handling coverage of fringe nodes, may be forced to trade away coverage opportunities for several intermediate nodes, potentially yielding substandard deployment positions or an inflated tally of such positions. One should remark that phase 2 of the RL-based optimization, geared toward delivering supplementary energy to nodes that remain unserved, could bring about a dip in charging efficiency. Even so, chasing ever-higher charging efficiency without a sensible distribution of energy is unwarranted.

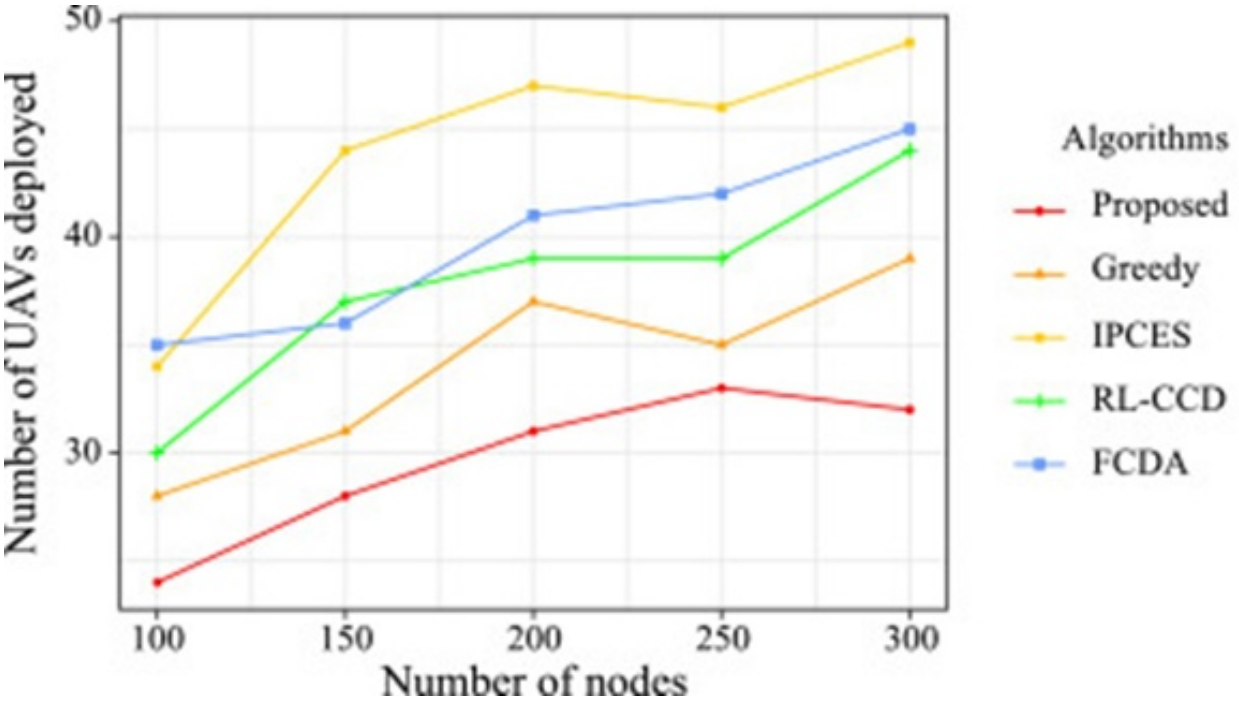


Figure 10. Comparison of each algorithm in the number of UAVs deployed.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

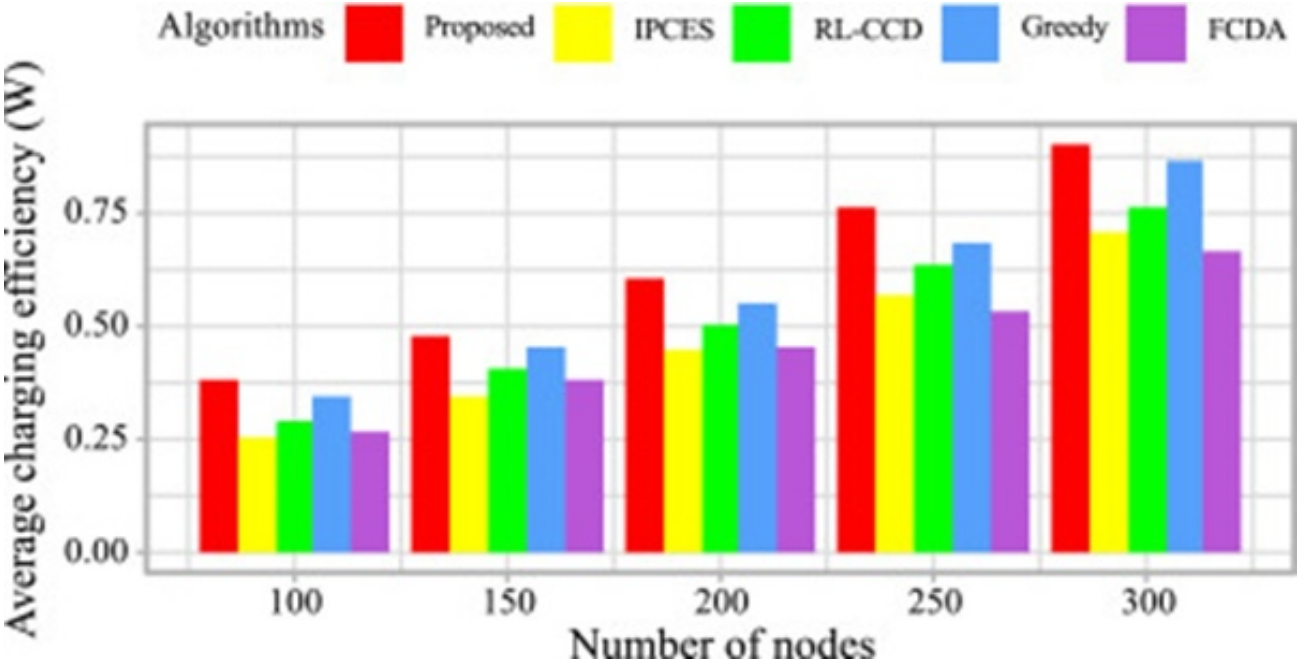
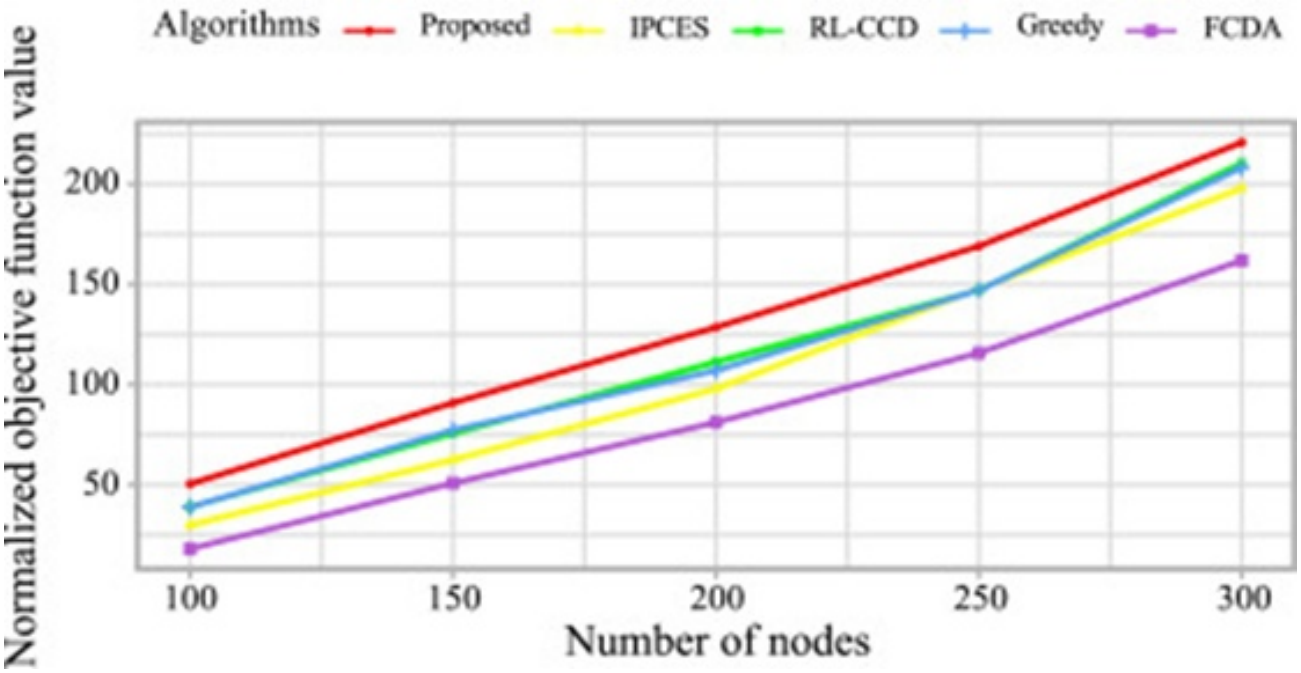


Figure 11. Comparison of each algorithm on the average charging efficiency of UAVs.

Within the problem formulation, three distinct objectives undergo normalization: the UAV fleet size, the network satisfaction percentage, and the aggregate charging efficiency. Normalizing the objective function yields a richer portrayal of the algorithm's capacity to respond to network demands. Against this backdrop, and keeping both node scatter and communication topology unchanged, a comparative evaluation of the objective function achieved by the proposed algorithm versus those attained by other methods is carried out. As rendered in **Figure 12**, the proposed algorithm commands a pronounced advantage over rival state-of-the-art approaches. The construction of the objective function spans three pivotal dimensions, and as corroborated by the comparisons laid out in the prior subsection, our methodology results in the leanest UAV fleet. Hence, the elevated objective function score registered by our approach signals a more judicious deployment arrangement and a closer alignment with network imperatives. The granular performance metrics of each algorithm across the three objective functions are collated in **Table 5.**



© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 12. Comparison of individual algorithms on the normalized objective function.

Table 5. Data presentation of the algorithm in each target.

Nodes	Proposed	IPECS	RL-CCD	Greedy
100	f_1=24, f_2=83+6.2, f_3=914.7	f_1=34, f_2=82+7.2, f_3=867.3	f_1=30, f_2=83+7.1, f_3=878.9	f_1=28, f_2=79+6.1, f_3
150	f_1=28, f_2=129+8.9, f_3=1341.3	f_1=44, f_2=133+7.3, f_3=1509.7	f_1=37, f_2=132+7.2, f_3=1489.5	f_1=31, f_2=123+6.8, f_3
200	f_1=31, f_2=174+7.2, f_3=18.697	f_1=47, f_2=174+7.3, f_3=2116.4	f_1=39, f_2=172+7.4, f_3=1965.7	f_1=37, f_2=164+7, f_3=
250	f_1=33, f_2=218+6.9, f_3=2521.8	f_1=46, f_2=222+7.1, f_3=2631.9	f_1=39, f_2=208+7.1, f_3=2473.8	f_1=35, f_2=200+7, f_3=
300	f_1=32, f_2=268+7.1, f_3=2888.1	f_1=49, f_2=277+7.3, f_3=3470.6	f_1=44, f_2=280+7.2, f_3=3354.3	f_1=39, f_2=268+6.9, f_3

Figure 13 illustrates the deployment layouts produced by various algorithms when the sensor node population stands at 100. The notations RL (1) and RL (2) correspond to the configurations obtained after phase 1 and phase 2, respectively, of the reinforcement learning optimization pipeline. Nodes shaded in red are those still classified as unsatisfied. One can readily observe that the seed solution sourced from the Improved NSGA-II algorithm, once subjected to RL (1) optimization, culminates in the smallest UAV contingent, with the deployment coordinates drawn nearer to the nodes themselves. In contrast, the Greedy algorithm exhibits a bulkier deployment count after RL (1) optimization, underscoring its struggle to evade the local optimum trap during the reinforcement learning refinement stage. Furthermore, the concluding deployment configuration attained post RL (2) optimization markedly strengthens the energy top-up directed at unsatisfied nodes, bringing about a steep reduction in the unsatisfied node count—an outcome that compares favourably against other advanced algorithms.

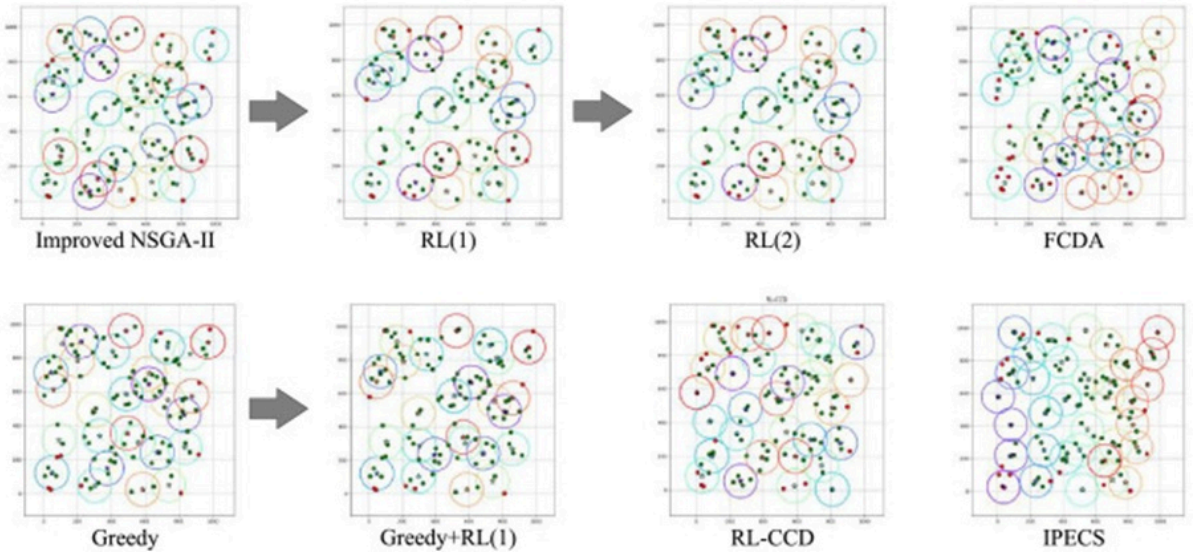


Figure 13. Deployment of each algorithm with the number of nodes being 100.

An inquiry was conducted into how the network satisfaction ratio distributes itself across node populations of 100, 200, and 300, running ten replicate trials for each setting. It is essential to highlight that the node scatter pattern together with the communication topology stayed invariant throughout all trials. As can be gleaned from **Figure 14,** the satisfaction rate yielded by the proposed algorithm persistently remains above 80%, and a discernible uplift in satisfaction is visible across all algorithms as the node count climbs. In particular, upon reaching a node count of 300, the satisfaction rate linked to

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

the proposed algorithm settles around roughly 90%, despite the IPCES and RL-CCD schemes posting even loftier satisfaction figures. Crucially, the proposed algorithm fields a markedly smaller UAV fleet relative to the competing methods. This observation confirms that the proposed algorithm is capable of sustaining a high satisfaction rate even as it economizes on UAV deployments.

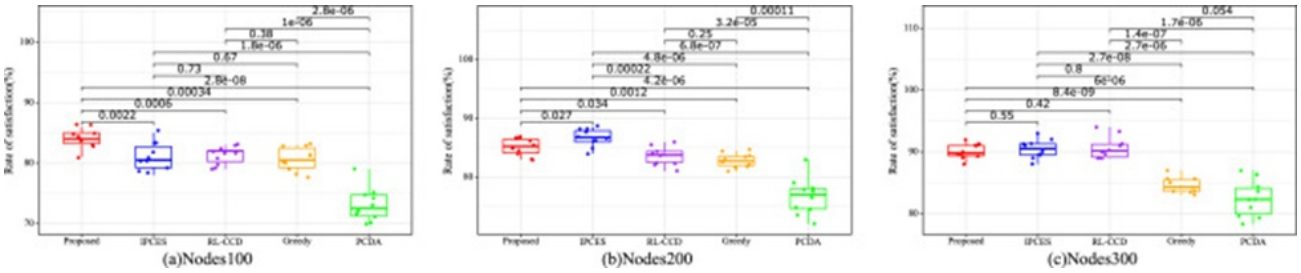


Figure 14. Network satisfaction ratio of each algorithm with different number of nodes.

The experimental outcomes signal that the network's overall charging efficiency consistently outstrips the energy draw dictated by its topology. Hence, room remains to further elevate network performance through a reworking of the communication topology. Within the original setup, communication links are reshuffled for nodes that remain unfulfilled under the current connectivity arrangement. The Dijkstra algorithm is brought in to locate suboptimal routing paths for these nodes. One must note, however, that the problem scale balloons enormously if every feasible path for each unsatisfied node is enumerated. For this reason, a ceiling of five suboptimal paths per unsatisfied node is imposed; these paths are sorted by ascending total transmission path length and tested for connectivity in sequence. Should all nodes populating a given path be able to reach satisfaction following reconnection, the link is established. Otherwise, the next suboptimal path is attempted. If none of the suboptimal candidates proves viable, the node is bypassed without altering its initial communication link. Holding all else equal, the communication links of the optimized deployed network are rebuilt and the resulting satisfaction rate is juxtaposed with that of the original configuration.

As shown in **Figure 15**, after topology reconstruction, the satisfaction rate reliably stays above 95%. Of particular note, if the pool of suboptimal candidate paths for reconstruction were left unbounded, there exists a substantial probability that the satisfaction rate could climb to 100%, provided the overall charging efficiency surpasses the aggregate energy consumption rate. Moreover, were UAV charging deployment anchored in particular communication or routing frameworks —for instance, those woven into the canonical LEACH [47] or HEED [48] protocols—the energy consumption across the network could be more evenly balanced, thereby prolonging the network's functional lifespan. **Table 6** catalogues the network parameters and the topological modifications observed for a 100-node instance following our reconstruction procedure. Post-reconstruction, a larger fraction of the dispensed charging energy gets productively used, which in turn drives up the energy utilization rate.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

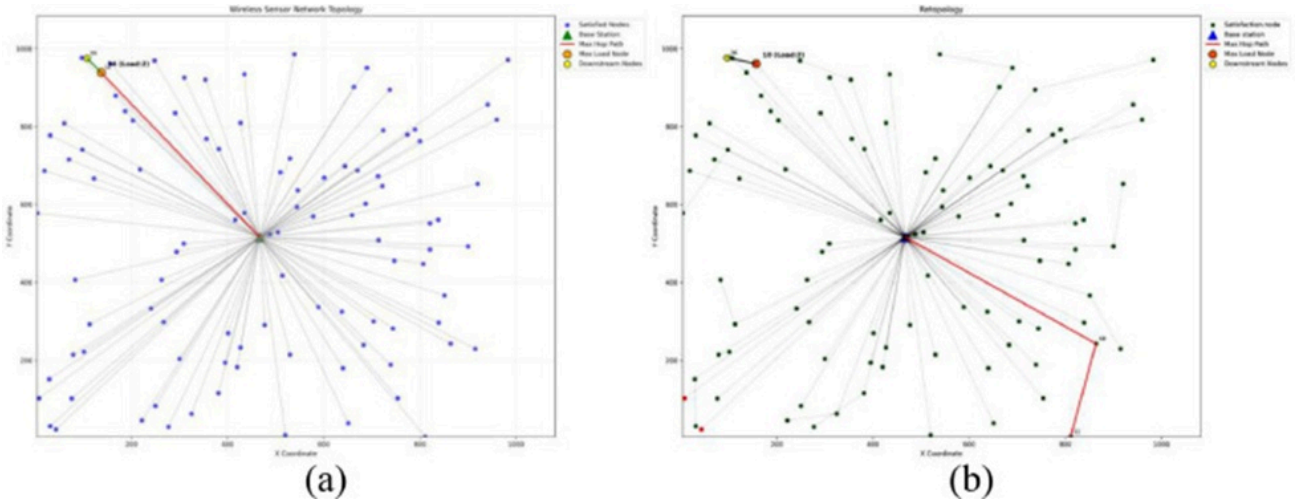


Figure 15. (a) Initial communication topology; (b) reconstruct the communication topology.

Table 6. The changes in the network after topology reconstruction.

Nodes	Satisfaction average rate after Retopology	Initial satisfaction average ratio	Total charging efficiency	Retopology energy consumption	Initial energy consumption
100	100 %	83 %	10.129W	8.276W	6.613W
150	96.7 %	86 %	14.481W	11.305W	9.908W
200	96.5 %	88 %	20.762W	15.737W	13.231W
250	97.6 %	89 %	25.658W	20.275W	16.468W
300	98.3 %	90 %	34.159W	24.101W	19.454W

To probe the resilience of the proposed algorithm, the network extent was enlarged to Z = 200 m, and 300 sensor nodes were sprinkled at random for testing. Of significance, as the network footprint swelled, the mean data transmission span for the nodes likewise doubled, triggering a sharp escalation in energy drain. With the UAV's radiated power held at 3 W, the collective charging efficiency proved insufficient to cover the network's total energy appetite. Consequently, the radiated power was raised to 6 W, with all remaining network parameters left untouched. As displayed in **Figure 16a**, the UAV fleet deployed across the network expanded commensurate with the enlarged network footprint. Subsequent to this concluding deployment, the network registered a node satisfaction rate approaching 94%.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

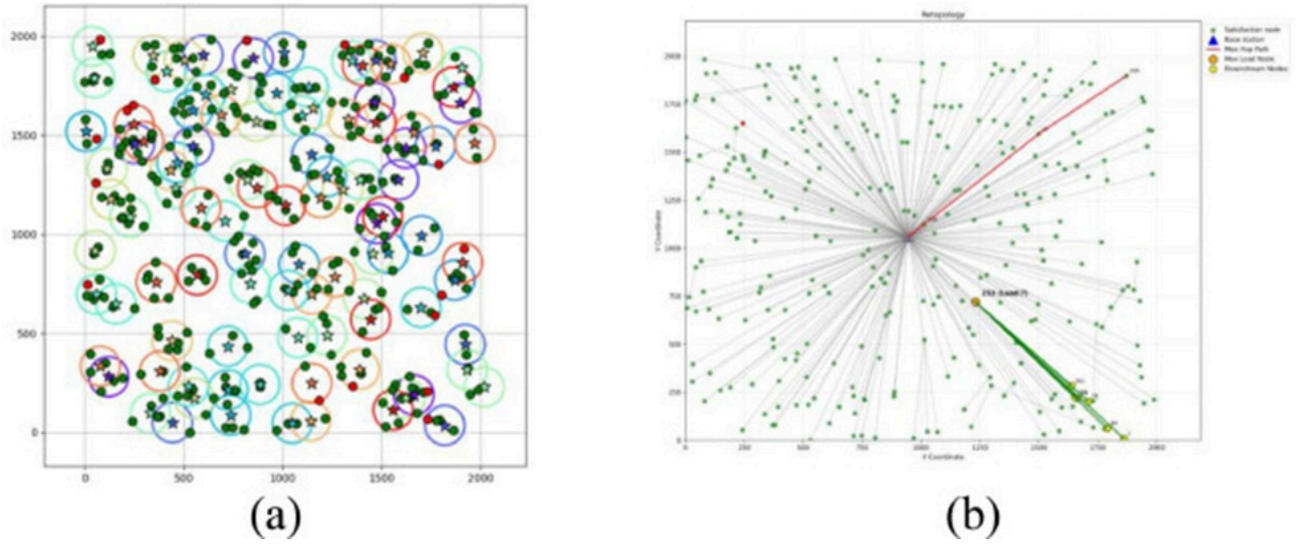


Figure 16. (a) Deployment of UAVs in the network; (b) the reconstructed communication topology.

Analysis experiment on the execution time of two-phase reinforcement learning algorithms

The execution durations of the dual-phase reinforcement learning procedure were examined on both CPU and GPU hardware platforms. For these tests, the network span was set to 100 m, the node count to 300, and the episode budget for each phase of the RL algorithm was fixed at 10, with a step ceiling of 200. It bears pointing out that once the network attains a sufficient level of optimization, the algorithm converges. Thus, not every step translates into a meaningful advance. Steps that yield no gain can be bypassed entirely, potentially saving a considerable chunk of time. For the purposes of a more thorough analysis of temporal cost, however, these steps were retained. **Figures 17a** and **17b** respectively portray the execution time profiles of the dual-phase RL algorithm on CPU and GPU. The average line (shown in red) reveals that the per-step execution time remains under the 1-second mark, irrespective of whether computation proceeds on CPU or GPU, a cost that falls within acceptable bounds. In comparative terms, shifting execution to the GPU delivers only a modest performance gain, with time savings remaining limited. This stems from the fact that the bulk of the computation undertaken by the proposed algorithm lies in energy and energy-consumption arithmetic as well as iterative scanning processes, a workload that is CPU-bound and serial in character. The RL operations themselves—Q-table lookups and updates—entail a negligible computational footprint and do not dominate the total runtime. The computational capacity of the GPU goes largely untapped.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

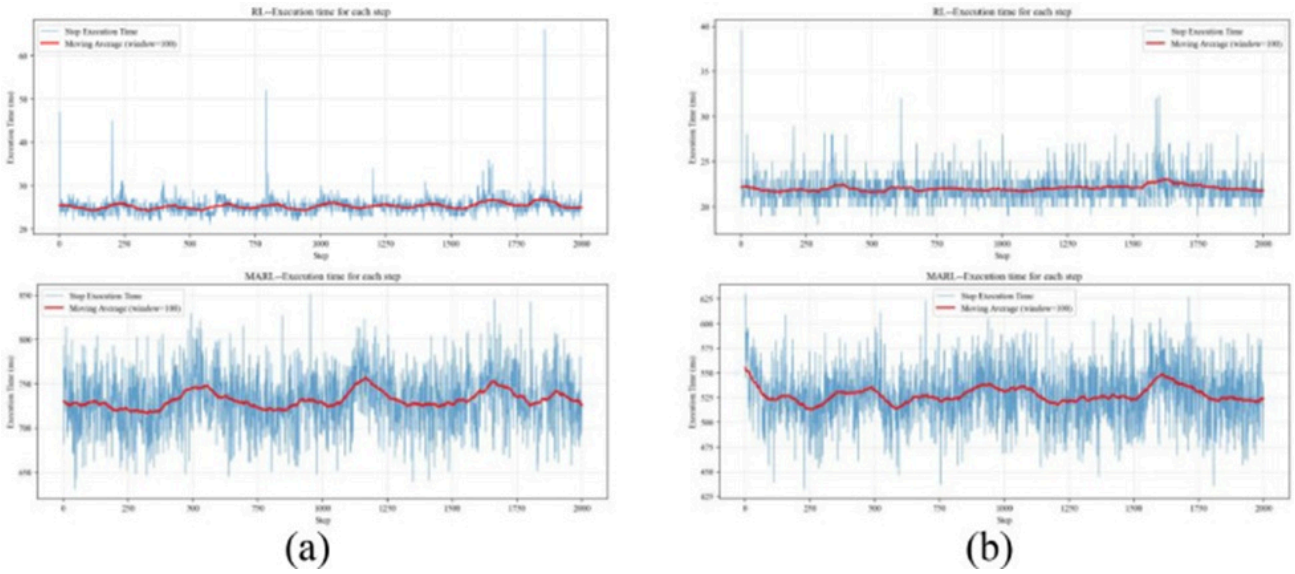


Figure 17. (a) Execution time of reinforcement learning using CPU, (b) Execution time of reinforcement learning using GPU.

Conclusion

This work puts forward a MUAVs-OCCDP custom-built for WRSNs deployments, driven by the ambition to satisfy network requisites while fielding the smallest attainable UAV contingent. A thorough decomposition of the MUAVs-OCCDP permits the primary objective to be broken into three priority-ordered sub-objectives, which are tackled one after another. The devised algorithm begins with the design of a minimal routing protocol that exploits the self-organizing tendencies of WRSNs to erect the network topology, with nodal energy drain rates serving as surrogates for charging demand. An improved NSGA-II variant is then introduced, blending the first and third sub-objectives into a unified multi-objective optimization construct. This construct operates under the stringent condition of guaranteeing full node coverage, thereby fixing the initial UAV deployment coordinates. Following this, a two-phase RL-based adaptive readjustment procedure is crafted to polish the seed deployment coordinates through learned exploration. In phase 1, under the guarantee that every node falls within at least one UAV footprint, freshly covered nodes and charging efficiency act as incentive signals to refine deployment sites. The twofold aim is to minimize the UAV count while boosting average charging efficiency. Phase 2 entails a further repositioning of the sites inherited from phase 1, targeting the maximal fulfilment of network requirements. Additionally, a multi-agent experience-sharing scheme is embedded to hone the movement strategies of agents and speed algorithmic convergence. The framework institutes both global and local reward components, complemented by a cooperation coefficient, to foster collaborative conduct among agents during learning, thereby effectively mitigating competitive deadlock. It supplies a fresh avenue for cooperative multi-UAV operation. Simulation outputs attest to its superiority over existing advanced methods across multiple dimensions. This paper confines its scope to charger placement within static WRSNs. In real-world practice, WRSNs operating under dynamic conditions enjoy far wider deployment. Hence, looking ahead, we intend to address dynamic WRSNs scenarios, crafting multi-UAV charging path plans anchored in desired deployment waypoints, coupled with mobility models and the spatiotemporal signatures of sensor nodes, to realize the sustained, self-perpetuating functioning of dynamic networks.

Acknowledgements

None

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Conflict of interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Reports a relationship that includes: Has a patent pending. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Financial support

This work was partly supported by the National Natural Science Foundation of China [grant No 62002180], the Key Scientific Research Projects of Colleges and Universities in Henan Province [grant Nos. 24A520030, 24A520031, 24A520032], the Training Plan for Young Backbone Teachers in Higher Education Institutions in Henan Province [grant No 2023GGJS120], the Key Scientific and Technological Research Projects in Henan Province [grant Nos. 252102210247].

Ethics statement

None

Received: 05 Jan 2022 Revised: 11 Mar 2022 Accepted: 28 Mar 2022
Published online: 15 November 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Srivastava A, Prakash J. Internet of Low-altitude UAVs (IoLoUA): a methodical modeling on integration of Internet of “things” with “UAV” possibilities and tests. Artif Intell Rev. 2022;56(3):2279-394.

2. Majid M, Habib S, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Applications of wireless sensor networks and internet of things frameworks in the industry revolution 4.0: a systematic literature review. Sensors. 2022;22:2087.

3. Al Khatib I, Shamayleh A, Ndiaye M. Healthcare and the internet of medical things: applications, trends, key challenges, and proposed resolutions. Informatics. 2024;11:47.

4. Dong Y, Li S, Bao G, Wang C. An efficient combined charging strategy for large-scale wireless rechargeable sensor networks. IEEE Sens J. 2020;20:10306-15.

5. Liu F, Jiang C, Xiao W. Multistep prediction-based adaptive dynamic programming sensor scheduling approach for collaborative target tracking in energy harvesting wireless sensor networks. IEEE Trans Autom Sci Eng. 2021;18(2):693-704.

6. Jin Z, Geng Y, Zhu C, Xia Y, Deng X, Yi L, et al. Deployment optimization for target perpetual coverage in energy harvesting wireless sensor network. Digit Commun Netw. 2024;10:498-508.

7. Ma D, Lan G, Hassan M, Hu W, Das SK. Sensing, computing, and communications for energy harvesting IoTs: a survey. IEEE Commun Surv Tutorials. 2020;22:1222-50.

8. Gao J, Wu R, Hao J, Xu C, Guo H. SWIPT-based energy scheduling for solar-powered WSN in full-duplex mode. IEEE Sens J. 2022;22:13668-81.

9. Soni S, Shrivastava M. Novel wireless charging algorithms to charge mobile wireless sensor network by using reinforcement learning. SN Appl Sci. 2019;1.

10. Detka K, Górecki K. Wireless power transfer—A review. Energies (Basel). 2022;15:7236.

11. Huda SMA, Arafat MY, Moh S. Wireless power transfer in wirelessly powered sensor networks: a review of recent progress. Sensors. 2022;22:2952.

12. Babu CVS, Pal A. Enhancing security for unmanned aircraft systems in IoT environments. Unmanned Aircraft Syst. 2024;429-76.

13. Hu S, Tsai Y, Kao S, Chang F. Using non-dominated sorting genetic algorithm-II for charging station deployment in wireless rechargeable sensor networks. Int J Appl Sci Eng. 2022;19(2):1-13.

14. Wang H, Li J, Xiao W. Reinforcement learning-based charging cluster determination algorithm for optimal charger placement in wireless rechargeable sensor networks. Ad Hoc Netw. 2024;164:103605.

15. Ding X, Guo J, Li D, Wu W. Optimal wireless charger placement with individual energy requirement. Theor Comput Sci. 2020;857:16-28.

16. Ding X, Wang Y, Sun G, Luo C, Li D, Chen W, et al. Optimal charger placement for wireless power transfer. Comput Netw. 2020;170:107123.

17. Lian J, Yao H. Joint deployment of sensors and chargers in wireless rechargeable sensor networks. Energies (Basel). 2024;17:3130.

18. Jin D, Jiao K. Charging infrastructure intellectualization and future of different automotive powertrains. Joule. 2020;4:1634-6.

19. Wan P, Cheng Y, Wu B, Wang G. An algorithm to optimize deployment of charging base stations for WRSN. EURASIP J Wirel Commun Netw. 2019;2019.

20. Liang S, Fang Z, Sun G, Lin C, Li J, Li S, et al. Charging UAV deployment for improving charging performance of wireless rechargeable sensor networks via joint optimization approach. Comput Netw. 2021;201:108573.

21. Yang M, Wang A, Sun G, Zhang Y. Deploying charging nodes in wireless rechargeable sensor networks based on improved firefly algorithm. Comput Electr Eng. 2018;72:719-31.

22. Tian Y, Cheng R, Zhang X, Su Y, Jin Y. A strengthened dominance relation considering convergence and diversity for evolutionary many-objective optimization. IEEE Trans Evol Comput. 2019;23(2):331-45.

23. Cho H-H, Chien W-C, Tseng F-H, Chao H-C. Artificial-intelligence-based charger deployment in wireless rechargeable sensor networks. Future Internet. 2023;15:117.

24. Deng T, Tang S, Chen Y, Lin F. A hybrid charging scheme for minimizing the number of energy-exhausted nodes in wireless rechargeable sensor networks. IEEE Sens J. 2024;24:17169-82.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

25. Lee C, Na W, Jang G, Lee C, Cho S. Energy-efficient and delay-minimizing charging method with a multiple directional mobile charger. IEEE Internet Things J. 2020;8:8291-303.

26. Jia R, Wu J, Wang X, Lu J, Lin F, Zheng Z, et al. Energy cost minimization in wireless rechargeable sensor networks. IEEE/ACM Trans Netw. 2023;31:2345-60.

27. Liu Y, Pan H, Sun G, Wang A, Li J, Liang S. Joint scheduling and trajectory optimization of charging UAV in wireless rechargeable sensor networks. IEEE Internet Things J. 2021;9:11796-813.

28. You W, Ren M, Ma Y, Wu D, Yang J, Liu X, et al. Practical charger placement scheme for wireless rechargeable sensor networks with obstacles. ACM Trans Sens Netw. 2023;20:1-23.

29. Dai H, Wang X, Liu A, Ma H, Ma G, Dou W. Wireless charger placement for directional charging. IEEE/ACM Trans Netw. 2018;26(4):1865-78.

30. Liao JH, Hong CM, Jiang JR. An adaptive algorithm for charger deployment optimization in wireless rechargeable sensor networks. Intell Syst Appl. 2014;2080-9.

31. Dai H, Wang X, Liu AX, Ma H, Chen G. Optimizing wireless charger placement for directional charging. IEEE Conf Comput Commun. 2017;1-9.

32. Hou G, Wang J, Fan Y. Wind power forecasting method of large-scale wind turbine clusters based on DBSCAN clustering and an enhanced hunter-prey optimization algorithm. Energy Convers Manag. 2024;307:118341.

33. Yao H, Zheng C, Fu X, Yang Y, Ungurean I. Charger and receiver deployment with delay constraint in mobile wireless rechargeable sensor networks. Ad Hoc Netw. 2022;126:102756.

34. Lee D, Lee C, Jang G, Na W, Cho S. Energy-efficient directional charging strategy for wireless rechargeable sensor networks. IEEE Internet Things J. 2022;9:19034-48.

35. Wang M, Chen H, Wang Y, Chen W. Improved Soft-k-means clustering charging based on node collaborative scheduling in wireless sensor networks. Wirel Pers Commun. 2024;137:2487-513.

36. Yu Y, Cheng Q. Charging strategy and scheduling algorithm for directional wireless power transfer in WRSNs. Alexandria Eng J. 2022;61:8315-24.

37. Liu T, Wu B, Zhang S, Peng J, Xu W. An effective multi-node charging scheme for wireless rechargeable sensor networks. IEEE INFOCOM 2022 - IEEE Conference on Computer Communications. 2020.

38. Chen Z, Tian H, Shen H. Improve the quality of charging services for rechargeable wireless sensor networks by deploying a mobile vehicle with multiple removable chargers. Wirel Netw. 2022;28:2805-19.

39. Mallikarjuna M, Amgoth T. A hybrid charging scheme for efficient operation in wireless sensor network. Wirel Netw. 2024;31:1301-20.

40. Xu W, Liang W, Jia X, Kan H, Xu Y, Zhang X. Minimizing the maximum charging delay of multiple mobile chargers under the Multi-node energy charging scheme. IEEE Trans Mob Comput. 2021;20:1846-61.

41. Wang N, Wu J, Dai H. Bundle charging: wireless charging energy minimization in dense wireless sensor networks. 2019.

42. Chen J, Yu CW, Cheng R-H. Collaborative hybrid charging scheduling in wireless rechargeable sensor networks. IEEE Trans Veh Technol. 2022;71:8994-9010.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

43. Ijemaru GK, Ang LM, Seng KP. Optimizing energy consumption and provisioning for wireless charging and data collection in large-scale WRSNs with mobile elements. IEEE Internet Things J. 2023;10:17585-602.

44. Radoslav L. Wireless identification and sensing platform. 2012.

45. Ostaffe H. RF-based wireless charging and energy harvesting enables new applications and improves product design. Powercast Corporation. 2017.

46. Fu L, Cheng P, Gu Y, Chen J, He T. Optimal charging in wireless rechargeable sensor networks. IEEE Trans Veh Technol. 2016;65:278-91.

47. Singh SK, Kumar P, Singh JP. A survey on successors of LEACH protocol. IEEE Access. 2017;5:4298-328.

48. Younis O, Fahmy S. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mob Comput. 2004;3:366-79.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press