Multi Objective Optimization Approach for WSN based on Reinforcement Learning

Faten Hajjej[1,2], Monia Hamdi[1], and Mourad Zaied[1]

[1] Research Team in Intelligent Machines, Gabes University, Tunisia
monia.hamdi@gmail.com
mourad.zaied@univgb.tn
[2] Gabes University Hospital, Public Health Ministry, Tunisia
hajjej.faten@gmail.com
faten.hajjej@rns.tn

Abstract. Wireless sensor networks (WSNs) are gradually invading our daily lives, offering us new services every day. They can be found in applications that affect us more and more. First used to monitor the environment and urban areas, they then provided support for first-aid and military surveillance activities. Now they are appearing in applications even closer to home to improve our lifestyle, such as guiding us to available parking spaces or informing us about air quality. The wide range of applications for wireless sensor networks has prompted several researchers to work towards a WSN with lower deployment costs while maximizing network lifetime and coverage. In this paper, an optimization approach-based Q-learning algorithm for optimal coverage of heterogeneous sensor networks is proposed. The findings of the simulation prove that the proposed approach maintains network coverage while using the minimum amount of energy, compared with other approaches.

Keywords: WSN · Coverage · Energy consumption · Q-Learning.

1 Introduction

Wireless sensor networks (WSNs) are essentially collections of independent nodes dispersed around an area of interest (AoI) that jointly sense and exchange physical and environmental parameters. Today, the use of solutions based on Wireless Sensor Networks (WSNs) is growing rapidly [1]. The WSNs deploy a very large number of small intelligent devices that compose distributed data management networks to collect detailed information about the environment. In typical scenarios, these networks are widely deployed in Areas of Interest (AoI), such as inaccessible terrain or high-risk locations, for the monitoring of different types of applications [1].

Despite miniaturization and reducing manufacturing costs, these sensors generally have limited resources in terms of power transmission, data processing, bandwidth, storage capacity, and energy. These material constraints have influenced many research issues in the field. Energy consumption is a key challenge in these networks since most low-power devices have batteries with a limited lifetime, and the battery replacement on tens of thousands of these devices deployed in areas with difficult access is often impractical and even impossible in most cases [2] [3]. Given limited network resources, the establishment of a powerful and efficient deployment topology is crucial to overcome problems deduced from the limited resources of sensor nodes and to enhance the network Quality of service (QoS). The problem is to find an optimal sensor deployment such that the given coverage and connectivity requirements are satisfied. This kind of system must be robust and self-healing because it has high costs and difficult maintenance to do redeployment. Thus, using optimization techniques during the deployment process is important to guarantee good performance and QoS. Network deployment topology directly affects QoS metrics such as monitoring quality, connectivity, and energy consumption. Consequently, the design of a WSN deployment strategy generally aims to achieve many objectives simultaneously, for example, maximizing the total covered zone and reducing network energy consumption. We often encounter contradictory objectives. Thus, using multi-objective-optimization approaches for deployment issues could find the best trade-off between conflicting objectives [4].

From the side of machine learning and data mining, WSN covers the learning field because the learning method, especially the distributed one, can be used to optimize WSN while solving many problems in WSN itself. All of the learning methods themselves can be modeled as optimization problems.

The current study addresses the coverage of the WSN problem by developing a novel distributed technique. Here, a distributed algorithm that may improve network coverage in a dynamic, autonomous, and decentralized way is proposed. The suggested approach is a multi-objective optimization approach based on a Q-learning algorithm. The main contributions of this paper can be summarized as follows:

- A novel distributed approach for topology control-based Q-learning aims to minimize network energy consumption and maximize overall coverage.
- Multiple simulations and comparative analyses are employed to demonstrate how the suggested method improves network performance while handling success.

The remainder of this paper is structured as follows: A related paper regarding various placement methods is provided in Section 2. Section 3 presents the problem formulation. Section 4 outlines the suggested deployment strategy. Results and a discussion are presented in Section 5. Finally, a brief conclusion is presented in Section 6.

2 Related works

Numerous techniques have been explored in recent WSN studies to optimize networks in terms of QoS measures. These recent research efforts are centered on elevating WSN coverage, reducing energy consumption, and enhancing network lifetime through optimization strategies. A new relocation strategy known as the Distributed Self-Spreading Algorithm (DSSA) is presented in [5] and is based on the radial virtual force. Numerous performance indicators were taken into consideration, including coverage, homogeneity, timing, and convergence. The placement of sensor nodes for Directional Sensor Networks (DSNs) was tackled by Zhang et al. [6]. In order to increase the monitoring field's coverage probability, they proposed a novel deployment strategy that made use of the best PSO algorithm [6].

In [14] optimal deployment model based on a modified version of NSGA-II was proposed for target monitoring applications in WSNs. The conventional dominance method was enhanced for better performance and selection of the chromosomes. The authors in [14] dealt with three objective functions, namely, connectivity, coverage, and the number of sensors.

The authors in [7] introduced a strategy that leverages the Voronoi Glow-worm Swarm Optimization K means algorithm to enhance network lifetime and reduce energy consumption. Wang et al. investigated swarm sensing algorithms for optimizing energy consumption and quality of service in WSN, achieving enhancements in coverage rates through optimization techniques [8].

The authors in [9] proposed an enhanced approach using a swarm optimizer to optimize coverage area and network energy. The proposed approach integrates the virtual force algorithm combined with a boundary mechanism to control the locations of sensor nodes and the Voronoi diagram to extract network data simultaneously. Zhu et al. employ an improved sparrow search algorithm (SSA) method to tackle problems of low node utilization, node redundancy, and low coverage. To improve population diversity and disperse the sparrow population more evenly, the good point-set population initialization approach is first implemented to replace the random population initialization in the original SSA. Then, an enhanced adaptive learning factor is added to SSA to update each person's position even more. Refraction reverse learning is then used to disturb people and prevent them from entering a difficult-to-leave local optimal region [10]. To find the best cluster centers while balancing five objectives: minimizing the number of cluster heads, maximizing overall cluster head energy, minimizing cluster compactness, minimizing energy consumption from non-cluster head to cluster head transmission, and maximizing cluster separation, the authors of [11] proposed a multi-objective binary Grey Wolf optimization technique.

3 Problem Formulation

Maximizing the total coverage area of the network is one of the main objectives of the area coverage problem. The coverage efficiency is immediately influenced by the choice of the deployment strategy. To ensure a powerful deployment strategy, it is recommended to deploy sensors deterministically [12]. Nonetheless, for applications involving dangerous areas of natural disasters and hostile environments, deterministic deployment is not always possible. Therefore, nodes may rather be deployed all at once, from an aircraft, for example, in a random way and at high density. The initial random topology cannot be assumed to be optimal since it may have serious overlapped zones and wide coverage gaps. The coverage metric of WSN essentially depends on the energy consumption of sensor nodes, which are self-powered, and for their entire lives, batteries are the only source of power. Generally, these sources are limited and not rechargeable. Sensing, processing, moving, and communicating are sensor functions that require energy consumption. Thus, once the sensor battery runs out of energy, the sensor node is not useful anymore. This can considerably degrade the covered area and the QoS for the entire network. In this work, we look to optimize the non-equitable initial distribution of nodes in terms of coverage, energy consumption, and connectivity metrics. The objective of the present work is to find the optimal locations of sensor nodes in terms of: (i) minimizing the total energy consumed by the network; (ii) maximizing the total covered area; and (iii) maintaining the network connectivity.

4 Proposed Approach

In this work, a distributed deployment technique based on a multi-agent Q-learning algorithm is suggested to enhance overall coverage while using a minimal amount of energy consumption for mobile WSNs. We combine both node replacement and sensing range adjustment actions to reduce coverage holes in the AoI.

4.1 Q-Learning Alghorithm

Q-learning is a reinforcement learning technique based on Markov Decision Processes (MDP), which mostly depend on the agent's repeated interactions with the environment.

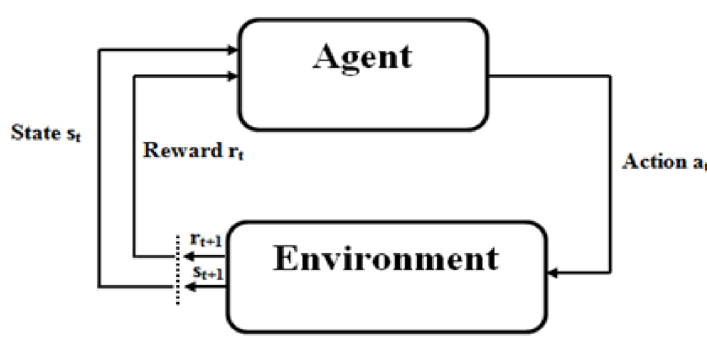


Fig. 1: Mono agent-environment interface in Q-Learning algorithm

In Q-learning, an agent decides, at a given instant t and from the present state s_t, to execute an action that results in a new state s_{t+1} and yields the reward r_t (see figure 1).

Therefore, it involves using previously gained experience to learn the actions that need to be taken in light of the current situation. A strategy known as the policy pi(s) directs the agent to choose every feasible state (position). Because of the mutual interactions between the environment and the learning system, the agent policy is gradually getting better.

Nevertheless, most of the reinforcement learning algorithms employ the iterative approximation of the evaluation function Q(s, a). The Q(s, a) value is specified as the estimated future benefits. Thus, an optimal strategy is to choose the action that corresponds to the maximum Q-value once these values are learned. A state's utility for a mono-agent system is determined by taking into account all

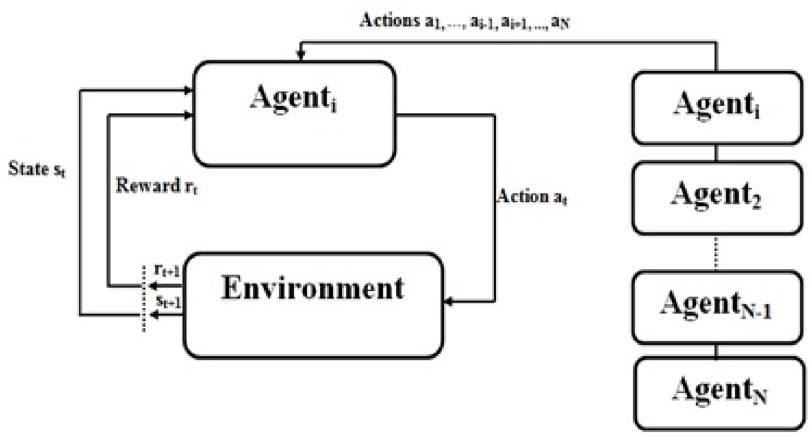


Fig. 2: Multi agent-environment interface in Q-Learning algorithm

conceivable actions and finding the state's maximum Qvalue. However, because the optimality criterion for a multi-agent system depends on the coordinated actions of other participants, it cannot take into account just the individual agent activities. As a result, when extending Q-learning to a multi-agent framework (consisting of N agents), it is necessary to include the behaviors of other agents when formulating the function.

4.2 Actions of agents

At each time step $t > 0$, the set of strategies that node $i \in \tau$ may select is the set of its possible actions $a_i \in \Lambda_i$. The node's actions are the adjusting position and sensing power transmission adjustment. Thus, each mobile node $i \in \tau$ can choose a combined action of changing its location p_i and its power transmission (the sensing range r_i): $a_i = (p_i, r_i), \forall a_i \in \Lambda_i$

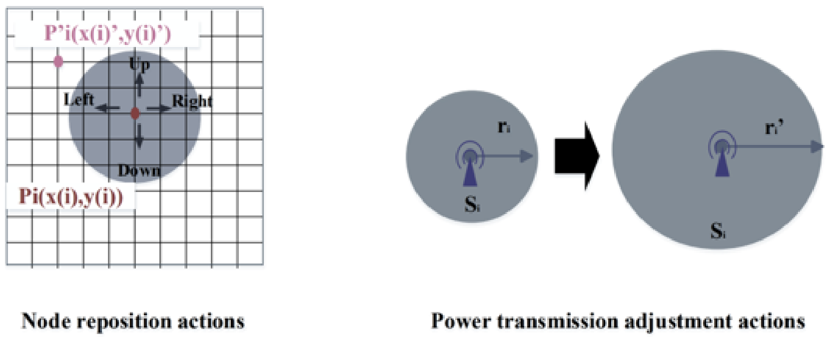


Fig. 3: Sensor node actions

i) Let us suppose that $p_i = (p_{x_i}, p_{y_i})$ is the current position of node i in the bi-dimensional space: $p_i \to p'_i$ is the change of its positions, with $p'_i = (p'_{x_i}, p'_{y_i})$ is the next position of node i (see figure 3).
ii) Let us suppose that $r_i, r'_i \in [R_{\min}, R_{\max}]$ are respectively the current and the next sensing range of node $i : r_i \to r_i{}'$ is the change of the sensing range, with R_{min} and R_{max} are respectively the minimum and maximum sensing ranges of the sensor node (see figure 3).

4.3 The utility function

Each node i has a utility or payoff function $\upsilon_i : \Lambda \to \mathbb{R}$, defined as:

$$\upsilon_i(S_i) = \mu_\alpha Profit(S_i) - \mu_\beta Cost(S_i) \quad (1)$$

Where $P(S_i)$ and $C(S_i)$ depict respectively the profit and the cost of the chosen strategy by node i. μ_α and μ_β are the weights that balance between the profit and the cost of the utility function. At each time t, agents simultaneously select their action strategies and receive their utility functions, which are specified by how much should be gained and how much should be paid. Every sensor node has a Non-Overlapped Sensing Area (NOSA), which is the area that is exclusively covered by that particular node (see figure 4).

The sensor node should minimize overlapped zones with its neighboring nodes to enhance the overall coverage area and decrease coverage gaps; as a result, maximizing the profit function value is thought to maximize the NOSA value.

$$P(S_i, S_{-i}) = Disk_i \setminus \bigcup_{j \in N_i} Disk_j \quad (2)$$

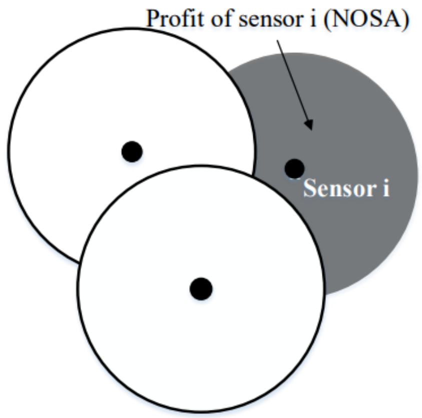


Fig. 4: Profit of sensor i

where N_i is the set of neighbor of node i and $Disk_i$ is the coverage disk of sensor i. The cost represents the energy consumed by the sensor nodes while recovering Coverage Holes. This cost mainly consists of:

- The consumed energy while changing the sensing power denoted by e_{Rs}
- The consumed energy while exercising a motion to change its positions denoted by e_{motion}.

The energy-related to changing the sensing power of node i is specified by:

$$e_{Rs} = \sum_{t} \sum_{i=1}^{n} (((4\pi . R_{S_i}/\lambda)^2) \Delta t) \tag{3}$$

The energy related to the node motions is specified by:

$$e_{motion} = \sum_{t} \sum_{i=1}^{n} (motion(i) \Delta t) \tag{4}$$

Where R_{Si} is the sensing range, Δt is the time duration/span.
The total energy consumed E_{tot} is regarded as the sum of the consumed energy taking into account the sensor motion and the sensing power change:

$$E_{tot} = e_{Rs} + e_{motion} \tag{5}$$

The network connectivity is assumed to be full if the communication radius is larger than the double of the sensing radius [13].

4.4 Deployment approach-based Q-Learning

In this work, we propose a novel cooperative approach based on multi-agent reinforcement learning, namely, the Q-Learning algorithm, to optimize node deployment in terms of coverage and energy consumption. Here, we deal with a multi-agent reinforcement learning concept. The figure 5 illustrates the steps of our deployment approach for one iteration.

We begin with the initialization of different parameters, such as the Q table. Each sensor has its own Q table that measures the quality of an action performed in a given state and whether it leads to another new state. The actions are categorized into two types, namely, changing position by moving in four directions (up, down, left, and right) and modifying the sensing radius. The states are the grid points of the AoI. Then, for each sensor, the algorithm switches between the exploration and exploitation phases. The former is to choose a random action to perform. The latter is to find the best Q value and perform the corresponding action. Finally, the Q value of the current agent was updated. The algorithm repeats those steps until its stopping criteria are reached.

The proposed algorithm can be described in the flowchart presented in Figure 6.

NbC represents the estimator of the action strategy of the opponent's agent. The n counts the number of visits to a given state s by sensors. The state is the vector of the state that contains the coordinates of each state in AoI. The Q-learning algorithm switches between exploring and exploiting phases to find the best action sequence. The exploration phase assists nodes in discovering the unknown environment rapidly by choosing arbitrary action; whereas the exploitation one enables them to keep effective actions by evaluating the flowing expression:

$$\arg\max_{a_i} \sum_{a_{-i}} \tfrac{NbC(s,a_{-i})}{n(s)} Q(s, a_i, a_{-i})$$

The metric ε controls the trade-off between the exploitation of the system's previously acquired knowledge and the exploration of the environment.

After choosing an action a_i, sensor Si acts by changing its position and its sensing range (power transmission) Rs. Then, the agent learns the obtained strategy (action, state) by updating its Q-table to use in the next iteration when choosing a new action. The Q tables are updated using the following mathematical formula:

$$Q(s, a_i, a_{-i}) \leftarrow (1-\alpha)Q(s, a_i, a_{-i}) + \alpha[R(s, a_i) + \gamma\pi(s')]$$

where

$$\pi(s') = max_{a_i} \sum_{a_i} \frac{NbC(s', a_{-i})}{n(s')} Q(s', a_i, a_{-i}) \tag{6}$$

With a_{-i} are the opponents' actions, $R(s, a_i)$ is the reward (calculated using the utility function), $C(s, a_{-i})$ is the number of times the opponent has played

action a_{-i} in state s, and s' is the next state reached by performing the chosen action a_i. Let α in $[0, 1]$ be the initial learning rate.
When the network's overall coverage level is reached, the algorithm satisfies the stopping criterion. If this still had not been satisfied, algorithms would stop running when the number of iterations reached 500.

5 Simulation experiments and results

To validate the mathematical formulations defined in previous sections, we needed to implement our approach and compare the obtained results with those reached by the PSO [8] algorithm, the NSGA-II [14] algorithm, as well as those of DSSA [5]. For a reliable comparison, we extended and simulated the three previously mentioned algorithms. The overall network coverage, the energy consumed while changing sensing power, the energy consumed while changing node motions, and the overall energy consumption were the performance criteria that were defined in this paper. The suggested method's effectiveness was evaluated in several topological scenarios. The simulation trials were performed using Matlab.

For our simulations, we considered a square region with a $100-meter-long$ edge divided into a predefined number of squares. The size of each square was equal to $1m^2$, and its center was the point to be covered. Nodes were randomly deployed in the AoI with a uniform distribution. The transmission range was fixed at $30m$, and the sensing ranges were chosen randomly between $7m$ and $15m$ Other simulation parameters are described as follows:

- xm (Maximum width of AoI) = $100m$
- ym(Maximum length of AoI) = $100m$
- IE (Initial energy for each sensor) = $10Ah$
- ε (Exploration rate) = $[0, 0.5]$

The metric ε controls the trade-off between the exploitation of the system's previously acquired knowledge and the exploration of the environment. ε was used to balance the exploration and exploitation processes in the algorithm. Using a random value within the interval given above, the Q-learning algorithm converges towards an optimal solution

Figure 7 shows a snapshot sample to prove the ability of our proposed Q-Learning approach to enhance the overall coverage of the AoI. This figure clearly shows the efficiency of our topology optimization method to improve the overall coverage to reach the coverage sill. In our case, we have selected a coverage sill value of 96%. Sensor nodes adjust both sensing range and position to quickly and accurately optimize network coverage.

Figure 8 and table 1 present the proposed approach's AoI coverage percentage, PSO, NSGA-II, and DSSA obtained. The figure clearly shows that our approach outperforms all the other algorithms' optimizations.

Figures 10, 11, and 9 show respectively the consumed energy related to the change of the sensing power, the consumed energy related to node motions, and the total energy. The results prove that our approach outperforms the three

Table 1: Comparison of total coverage percentage for 50-nodes topology.

	Node=20	**Node=30**	**Node=40**	**Node=50**
PSO	35.61	50.64	63.3	70.9
NSGA-II	45.3	64.9	71.9	81.9
DSSA	56.98	71.36	79.32	88.03
Proposed Approach	62.34	78	82.9	96.46

Table 2: Total Log network energy consumption (in mAh) Vs. number of sensor nodes.

	Node=20	**Node=30**	**Node=40**	**Node=50**
PSO	13.67	14.1	14.51	15.48
NSGA-II	13.37	13.67	14.33	15.18
DSSA	10.59	10.69	11.16	13.24
Proposed Approach	9.46	10.19	10.44	12.01

approaches in terms of energy consumption (see Table 2). As presented in Figure 10 the energy consumed related to power adjustment is more important than that related to node motion for the proposed approach. These findings imply that sensor nodes mitigate the uncovered zone with fewer possible movements and privilege the power adjustment action. Thus, our approach succeeded in extending the network's lifespan by dissipating less overall energy (see Figure 11).

In contrast to alternative methods (research based on simulations or mathematical models with set experimental parameters), this solution may adjust to changes in the network's status. Moreover, this method can adjust to the newly learned information because of the reliable machine learning algorithm, Q-learning. Thus, the Q-learning-based reward function facilitates a faster approach to the predefined overall coverage threshold of the detection region (refer to figure 7).

Furthermore, Q-learning considers the anticipated neighbor's behaviors in addition to the agent's (sensor) best past actions. Such behavior enables sensors to more accurately predict the optimal course of action and make the most successful decision possible, improving overall coverage and energy usage while recovering coverage holes. As a result, our suggested approach's topology control scheme performs better than alternative methods in terms of energy usage and coverage for an AoI.

6 Conclusion

The research aims to solve the problem of insufficient coverage and poor energy resources in wireless sensor networks. By employing Q-Learning algorithm performances, this study designed a powerful wireless sensor network optimization algorithm and applied it to enhance the sensor node deployment process in AoI. This method makes it possible to find a solution that adapts to the needs of WSN users by finding a trade-off between several objectives, such as coverage, energy consumption, and connectivity. The introduced approach was comprehensively validated. Experiments are carefully designed by programming real-life scenarios. The results indicate that the proposed methodologies can assist WSN designers in efficiently planning reliable and optimized WSN topology for area coverage scenarios, compared with other state-of-the-art algorithms.

References

1. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci: Wireless sensor networks: a survey. Computer Networks **38**, 393–422 (2002)
2. E. Shih, S. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan: Physical layer-driven protocol and algorithm design for energy-efficient wireless sensor. In: MobiCom001: Proceedings of the 7th annual international conference on Mobile computing and networking, pp. 272–287. ACM, Roma, Italy (2001)
3. B.J. Chen, K. Jamieson, H. Balakrishnan, and R. Morris: Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks. Wireless Networks **8**(5) (2002)
4. Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, L. Hanzo: A Survey of Multi-Objective Optimization in Wireless Sensor Networks: Metrics, Algorithms and Open Problems. IEEE Communications Surveys & Tutorials **(**19), 550–586 (2017)
5. Z. Junguo, L. Yutong, C. Chen, L. Fantao: Directional Probability Perceived Nodes Deployment Based on Particle Swarm Optimization. International Journal of Distributed Sensor Networks **(**12), 1–6 (2016)
6. N. Heo, P.K. Varshney: A distributed self-spreading algorithm for mobile wireless sensor networks. In: IEEE Wireless Communications and Networking, pp. 1597–1602. IEEE, USA (2003)
7. A. Chowdhury, D. De, A. Raychaudhuri, M. Bakshi: Corrigendum to Energy-Efficient Coverage Optimization in Wireless Sensor Networks based on Voronoi-Glowworm Swarm Optimization-K-means algorithm. AdHoc Networks, **(**122),102907 (2021)
8. G. Wang, J. Feng, G. Li, J. Song, D. Jia: Energy Consumption and QoS Optimization Coverage Mechanism in Wireless Sensor Networks Based on Swarm Sensing Algorithm. Journal of Sensors **7**,1-11, (2022)
9. P. Musikawan, Y. Kongsorot, P. Muneesawang, and C. So-In: An enhanced obstacle-aware deployment scheme with an opposition-based competitive swarm optimizer for mobile WSNs. Expert Systems with Applications **189**(5),116035 (2022)
10. J. Wang, D. Zhu, Z. Ding and Y. Gong: WSN Coverage Optimization based on Improved Sparrow Search Algorithm. In: 15th International Conference on Advanced Computational Intelligence (ICACI), pp. 1–8, IEEE, Seoul, Korea (2023)
11. R. Pal, M. Saraswat, S. Kumar; Energy efficient multi-criterion binary grey wolf optimizer based clustering for heterogeneous wireless sensor networks. Soft Comput **(**28), 3251–3265 (2024)
12. T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan and K. K. Saluja: Sensor deployment strategy for detection of targets traversing a region. Mobile Networks and Applications **(**8), 453–461 (2003)
13. A. Ghosh, and S. K. Das: Coverage and connectivity issues in wireless sensor networks: A survey. Pervasive and Mobile Computing **(**4), 303-334 (2008)

14. S. Harizan, P. Kuila: A novel NSGA-II for coverage and connectivity aware sensor node scheduling in industrial wireless sensor networks. Digital Signal Processing (105), 102753 (2020)

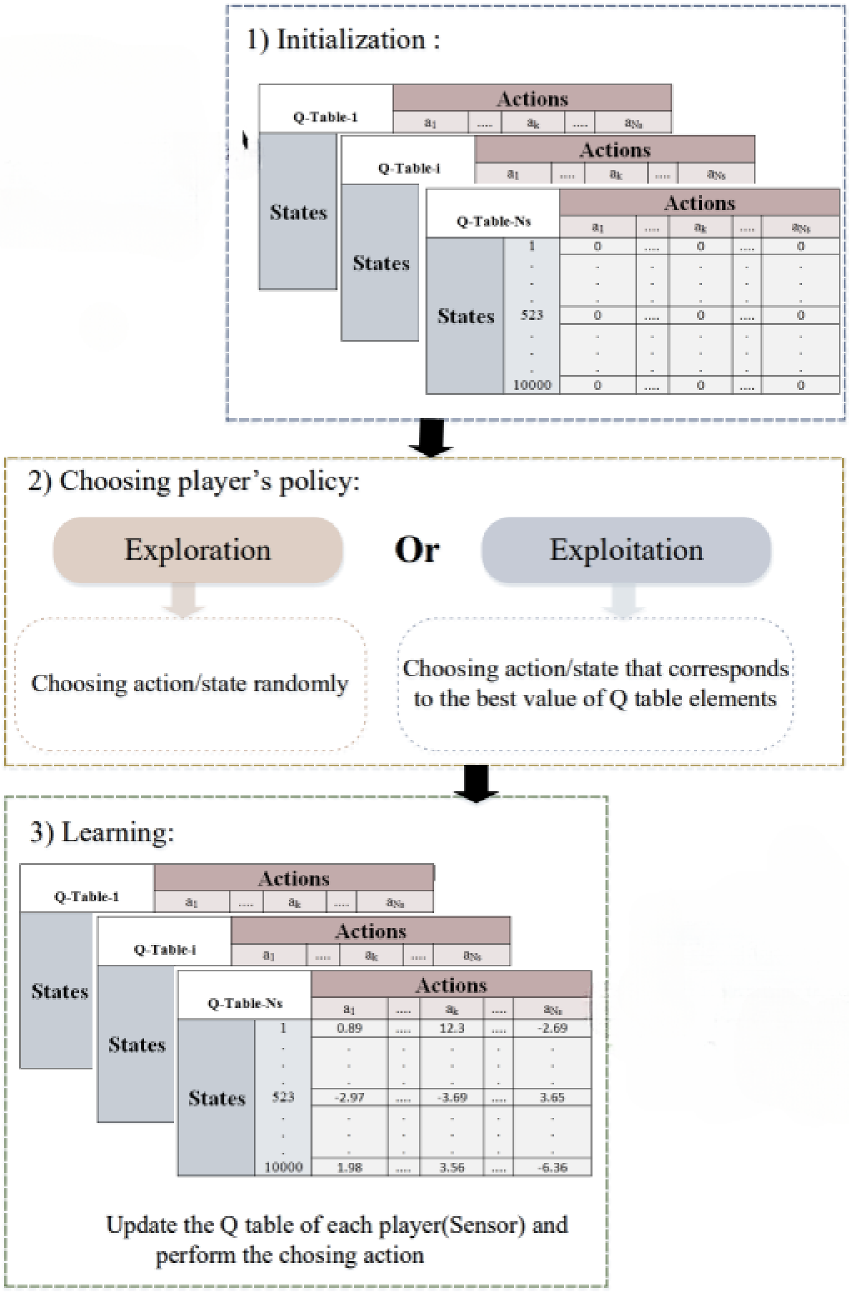


Fig. 5: The schematization of the WSN deployment approach based on Q-learning for one iteration.

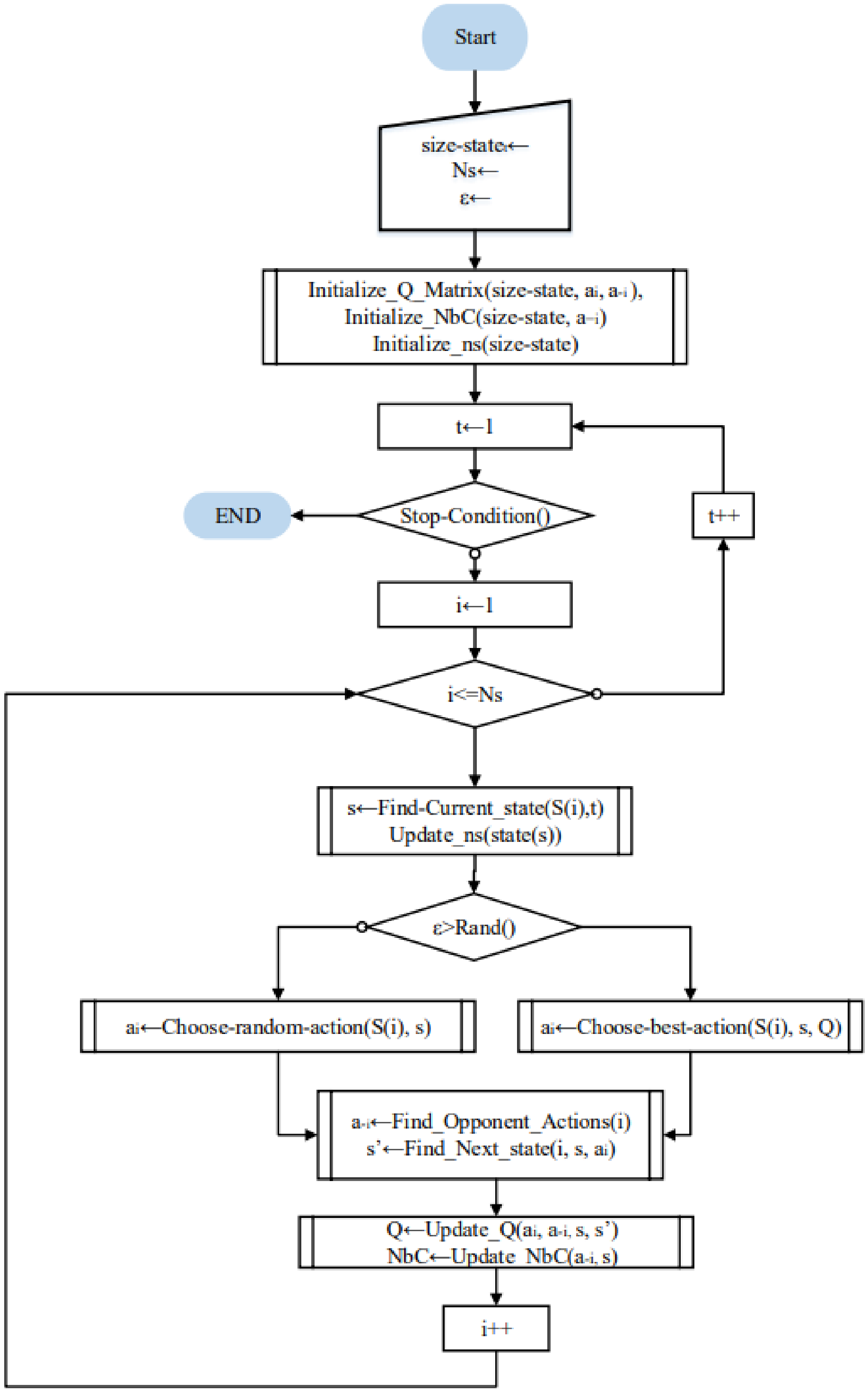
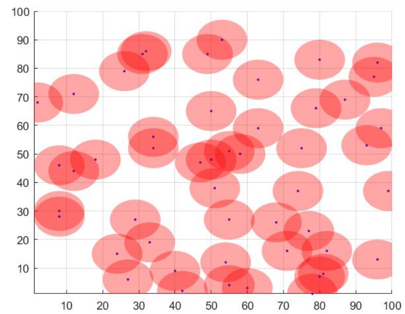
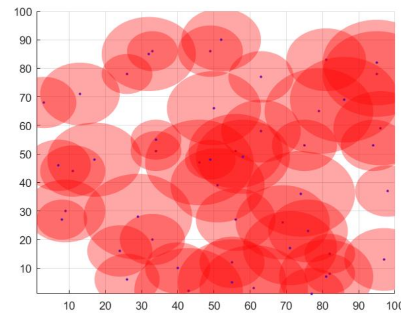


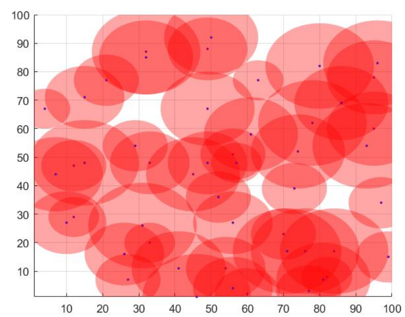
Fig. 6: The proposed deployment approach flowchart.



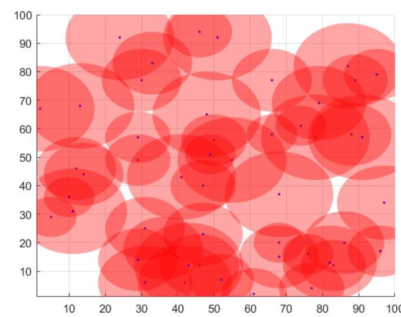
(a) Initial Deployment



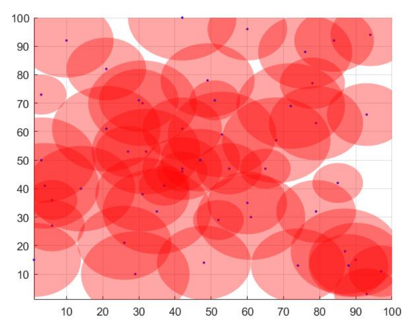
(b) iteration 1



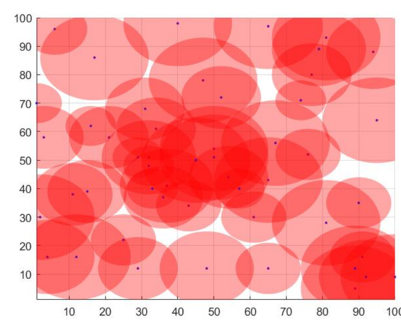
(c) iterations 7



(d) iterations 74



(e) iterations 126



(f) iterations 212

Fig. 7: Overall coverage optimization for 50 nodes topology.

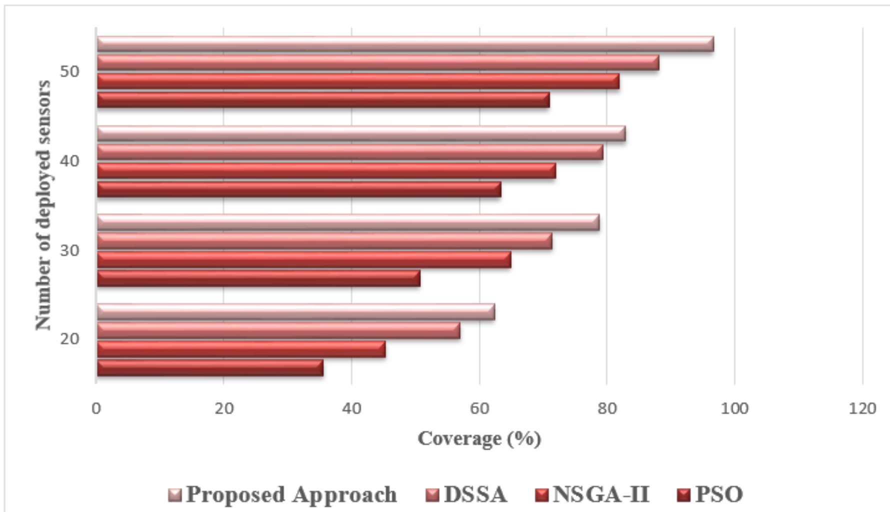


Fig. 8: Percentage of overall coverage Vs. the number of nodes.

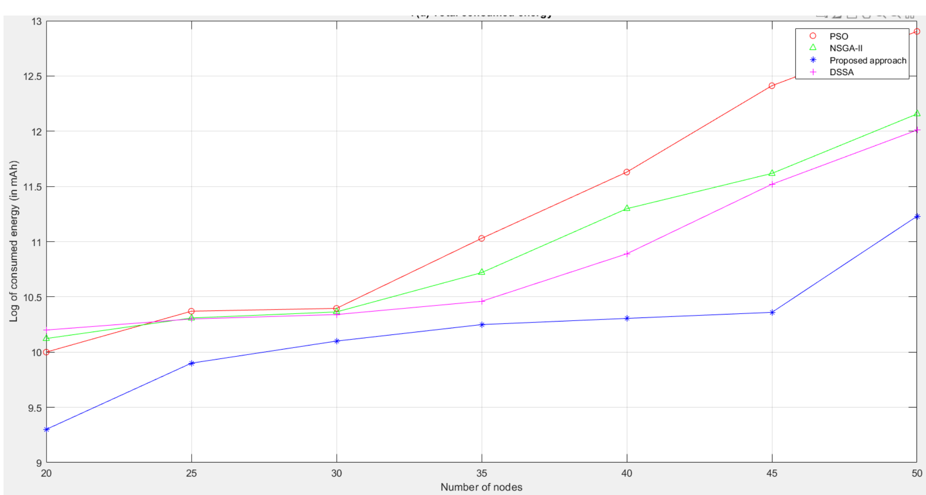


Fig. 9: The total consumed energy

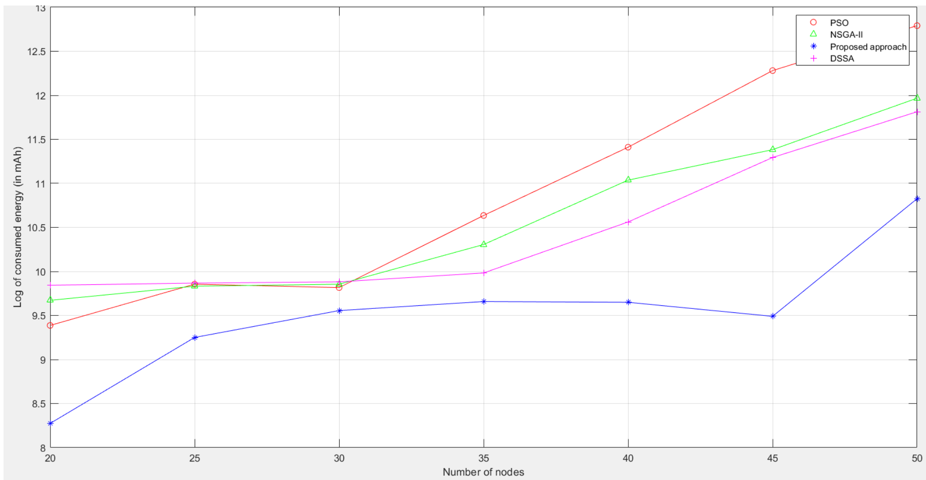


Fig. 10: The consumed energy related to the change in the sensing power.

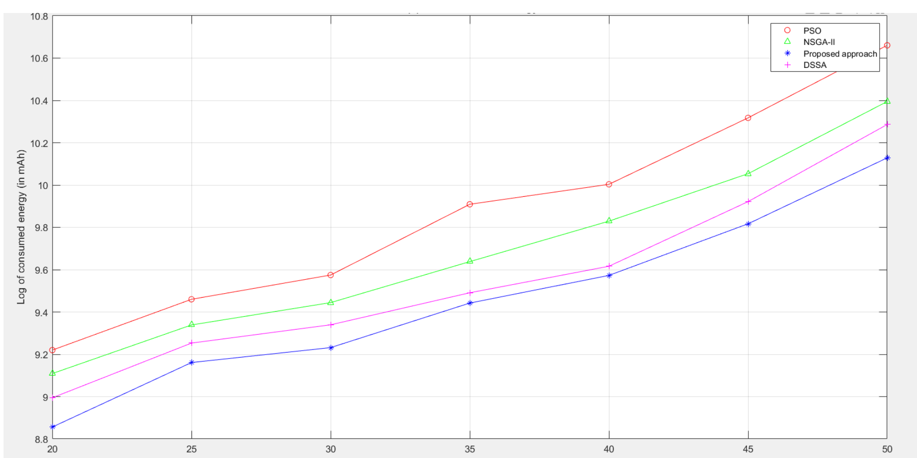


Fig. 11: The energy consumed related to node motions.