Ad Hoc Networks 101 (2020) 102082

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

A distributed coverage hole recovery approach based on reinforcement learning for Wireless Sensor Networks☆,☆☆

Faten Hajjej[a,*], Monia Hamdi[b,a], Ridha Ejbali[a], Mourad Zaied[a]

[a] Research Team in Intelligent Machines (RTIM), Gabès University, Tunisia
[b] Information Technology Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 26 December 2018
Revised 17 January 2020
Accepted 24 January 2020
Available online 4 February 2020

Keywords:
WSN
Coverage hole recovery
Game theory
Reinforcement learning
Energy consumption
Coverage

ABSTRACT

In Wireless Sensor Networks (WSNs), various anomalies may arise and reduce their reliability and efficiency. For example, Coverage Hole can occur in such networks due to several causes, such as damaging events, sensors battery exhaustion, hardware failure, and software bugs. Modern trends to use relocation of deployed sensor nodes when the manual addition of nodes is neither doable nor economical in many applications have attracted attention. The lack of central supervision and control in harsh and hostile environments have encouraged researchers to shift from centralized to distributed node relocation schemes. In this paper, a new game theory approach based on reinforcement learning to recover Coverage Holes in a distributed way is proposed. For the formulated potential game, sensor nodes can recover Coverage Holes using only local acquaintances. To reduce the coverage gaps, the combined action of node reposition and sensing range adjustment is chosen by each sensor node. The simulation results prove that, unlike previous methods, the proposed approach can sustain a network overall coverage in the presence of random damage events.

1. Introduction

Over the last decade, continuous and omnipresent monitoring has become possible thanks to large mobile networks and wireless sensors.

Currently, the use of solutions based on Wireless Sensor Networks (WSNs) has gained extensive interest [1]. WSNs deploy a very large number of small intelligent devices that form distributed ad hoc networks for data collection and management. In typical scenarios, these networks are widely deployed in inaccessible and dangerous areas for monitoring different classes of applications [1]. Therefore, self-organization and mutual cooperation of nodes, are essential as they extend the lifetime of the network and reduce congestion by avoiding redundant data [2]. Wireless communications play a crucial role in computer networks. Their usefulness lies in the open solutions they offer to provide mobility and essential services where infrastructure installation is not possible [3].

WSNs have applications in various fields, such as environmental and real-world habitat monitoring, machine surveillance, precision farming, indoor control, intelligent alarms, and military applications. Monitoring is one of the most important and critical applications of WSNs. In fact, such networks perform this task using a large number of electronic devices deployed across the area of interest (AoI) in order to sense and communicate environmental or physical parameters cooperatively [1,4–6]. Generally, the sensor nodes' initial positions are not predetermined or preconceived, but rather randomly generated. Indeed, random positioning also means that the monitored areas may have Coverage Holes and serious overlapping areas, that powerfully degrade the efficiency, reliability, and quality of service (QoS) [7–9]. Random deployment is not the only reason for coverage gaps damaging events, sensors battery exhaustion, hardware failures, software bugs, and security attacks could also be other possible reasons [8].

In most cases, it may be impractical and even impossible to replace or recharge the damaged sensor nodes by technicians, especially for WSNs settled in hostile and/or remote environments. However, the gathered data are often scientifically or strategically important, and any discontinuity in the data collection process can seriously decrease the efficiency and robustness of the network. A quick and autonomous network recovery can avert the WSN services break. Therefore, the ability of a network to heal itself with

☆ Fully documented templates are available in the elsarticle package on CTAN.
☆☆ Since 1880.

* Corresponding author.
E-mail address: hajjej.faten@gmail.com (F. Hajjej).

https://doi.org/10.1016/j.adhoc.2020.102082
1570-8705/

minimal human supervision is important to ensure rapid responses in case of emergency events [10,11].

In [12], three distributed approaches were proposed for coverage control in WSNs without information on the location. The authors in [12] present a decentralized scheme based on Laplacian flows to process a generator of Coverage Holes. The coverage hole is spotted based on the first homology classes of the Rips complex. A self-healing distributed approach based on a fuzzy-logic was proposed in [13] to detect and repair the Coverage Holes. For mobile sensors, the density control technique was used to minimize the overlapping area in order to eliminate Coverage Holes. In [14], the authors introduce a distributed method that detects and eliminates Coverage Holes with the minimum number of redundant nodes using only connectivity information in a high-density network. An autonomous distributed coverage hole detection approach was proposed in [15]. To recover holes, this approach calculates the available mobility distance before moving to eliminate holes. To maintain the network coverage, a distributed algorithm based on the Voronoi diagram was presented in [16]. The authors propose to control the size of each Voronoi cell to eliminate the Coverage Holes. In this case, the communication range must be long enough otherwise the Voronoi cell of each node may not be judged accurately.

Intense research activities have been devoted to the development of a variety of network recuperation and topology control schemes, such as sensing power adjustment, sensor node reposition [10,11], and clustering [17,18].

Hybrid topology control schemes, such as altering both sensing power and spatial locations of mobile nodes, have also been elaborated to overcome the limitations of previous schemes. Power adjustment aims to dynamically modify the node sensing range to maintain network connectivity and minimize the amount of dissipated energy per node. In certain situations, the power adjustment is more appropriate than nodes reposition mechanisms. In fact, the dimension of the damaged area could be reduced by a simple power adjustment and thus economize energy consumption by avoiding unnecessary movements. Physical node reposition can react more easily, especially in the proximity of damaged zones. Therefore, a harmonic balancing between both mechanisms using a hybrid system permits the network's healing process to be more resilient and economical in terms of energy consumption [10].

Various coverage control topologies were implemented for WSNs, but all of them focus mainly on two models of topologies: centralized topology and distributed topology. The former consists of a collection of nodes that sends new information and forwards it to sink nodes where all the data is processed. It is established on the central nodes and locations of the other sensor nodes. However, the latter has no central management by the sink nodes. It consists of a set of independent nodes with equal roles. Because of the random behavior of Coverage Holes, accurate temporal and geographical information on the Coverage Hole are not constantly available and thus a dynamic and repeated adjustment of the deployment parameters becomes necessary. Therefore, Coverage Hole recovery using any centralized approach would result in computational complexity and energy consumption [10]. Distributed hybrid topology seems to be the most appropriate solution that could heal Coverage Holes optimally in terms of QoS metrics (i.e. energy consumption, network lifetime, total coverage rate etc.). Indeed, the distributed approaches would enable the sensor nodes to react promptly and autonomously to unforeseen events with a minimal degree of dependency and human supervision. Each node could limit its interactions with other nodes and communicate only with its neighbor nodes. Besides, it could measure the local uncovered area to take the appropriate actions later, which would reduce the consumed energy and enhance the network survivability [19]. A distributed topology control approach can handle the problem from the game theory (GT) perspective. In fact, since the game theory is used to make strategic decisions in uncertain environments, its relevance to a distributed control is already recognized [20].

Recently, researchers have shown significant interest in the extension of multi-agent reinforcement learning [21] in stochastic games [22] which are presented as a relevant framework for multi-agent learning [23]. This type of games extends both game theory [20] and the formal framework of Markov decision processes, in which the action-state binding learning algorithm also called the Q-learning algorithm [24], is defined. A repeated game is a sub-specialty of stochastic games. The players repeatedly perform the game and, at any iteration, the current game necessarily depends on the results of the previous one.

In the present work, a novel distributed approach is designed to tackle the problem of Coverage Holes for WSN. Here, a hybrid algorithm that is able to mitigate the coverage gaps in a decentralized, dynamic and autonomous way is designed. The proposed approach merges two coverage control schemes, namely the nodes reposition and power transmission adjustment, using the game theory concept based on the Q-learning algorithm. Q-learning is a reinforcement learning technique that relies mainly on the iterated interaction between the learning system and the environment. Reinforcement learning techniques were used to solve a variety of WSNs issues, such as, power management [25], security [26] and routing problems [27]. To the best of our knowledge, this is the first work that uses reinforcement learning (Q-learning) to expand distributed approaches for Coverage Holes problems in mobile sensor networks. The main contributions of this paper can be summarized as follows:

- A novel hybrid approach for topology control based on the game theory concept and Q-learning aiming to recover Coverage Holes optimally by finding trade-offs between setting the sensor positions and the sensing power.
- A novel potential game where nodes actions are decided in an autonomous and decentralized way, which qualifies the proposed control scheme to be used for networks settled in harsh and hostile environments.
- Extensive simulations and comparison studies are used to prove that the proposed approach enhances the network performances when dealing with successive random Coverage Holes.

The remainder of this paper is organized as follows: Section 2 reviews the related work about different control schemes for WSNs. The system model is identified in Section 3. Section 4 describes the proposed approach. In section 5, we present and discuss the study results. Finally, the main conclusion of our study is drawn in Section 6.

2. Related work

The main research field of topology control schemes can be categorized into two types: node relocation-based schemes and power transmission-based ones [10,11]. Sensor nodes relocation provides a promising solution to emerging and challenging coverage problems in WSNs [28–30]. Relocation schemes can be classified into virtual force-based schemes, Voronoi-based schemes, and flip-based movement schemes. In [28], a new relocation approach based on the radial virtual force called the Distributed Self Spreading Algorithm (DSSA) is proposed. Several performance metrics such as coverage, uniformity, time, and convergence were considered. Despite the good performance of DSSA, especially for sensors movements and small-scale coverage hole recovery, the results presented in [31] prove that it is not as effective in large scale coverage hole since it needs a huge number of iterations to converge. Nevertheless, nodes relocation can produce new small Coverage Holes due to independent and distributed motions. Besides, in certain cases, some nodes motions could be useless in terms of coverage maximization and may deplete the sensor battery supplies and reduce their ability to reach and heal the Coverage Holes. Topology control schemes based on adjusting power transmission have been proven to be a powerful control scheme for small-sized Coverage Holes [32,33]. The individual sensor nodes require higher transmission power to recover wide coverage gaps which rapidly exhausts the battery supplies and widens the Coverage Hole areas. Recently several research studies have exploited the hybrid topology control scheme for hole recovery; however the results presented in [31] prove that it does not overcome the deficiencies of the previously mentioned approaches [34,35], and [36]. In [35], an exact potential game was developed for coverage optimization in a WSN consisting of two distributed based learning algorithms. A sensing model based on footprints and finite angle of views were considered. Each sensor node decides to move or to adjust its angle of view/range in terms of its residual energy and neighbor nodes' actions. In spite of the good performance of these models, the authors of [36] showed that they are restricted to specific wireless network applications due to their limited sensing model (finite angle of view). The same authors investigated an omnidirectional sensor model which can be employed in a wider range of applications thanks to its larger sensing scope. However, a learning algorithm similar to that of [35] was used for the proposed potential game.

Lately, there has been a substantial interest in extending single agent reinforcement learning [37] to stochastic games [38] which are presented as a relevant framework for multi-agent learning. Such games extend both the formal framework of Markov decision-making processes and the game theory. Nonetheless, one of the major problems in extending mono-agent learning to multi-agent systems is the interaction between agents: individual actions can no longer be considered in isolation from the actions of the other agents, as their consequences are interdependent. A single agent using reinforcement learning can converge to an optimal policy towards stationary agents since the stationarity of opposing agents can be included in the environment model. In this case, the problem reverts to a single agent environment.

Several previous works have focused on the possible contributions ranging from game theory to multi-agent learning. Littman [39] proposed the minimax-Q learning algorithm, in which he proved the convergence of purely competitive games (i.e. opposite rewards). The authors in [40] have introduced the Nash Q-learning algorithm as part of non-cooperative stochastic games, with decorrelated rewards. In the same context, Greenwald et al. [41] proposed a similar version to Nash Q-learning, called correlated equilibria learning that used the value of correlated balances (in case of correlated rewards) rather than Nash's balances. Littman [42] redefined Nash Q-learning in the Friend-or-Foe Q-learning algorithm, as the combination of a cooperative algorithm and a competitive and a proven convergence towards Nash equilibrium for different classes of stochastic games. For their part, Claus et al. [43] introduced the concept of a joint-action learner (JAL) which learns the value of joint actions rather than just the value of its own actions and proved the convergence towards a Nash equilibrium in the purely cooperative games (identical rewards). In this work, the concept of JAL is considered. Moreover, a new distributed approach based on the game theory and multi-agent reinforcement learning (Q-learning) is proposed for the Coverage Holes recovery. A study of the proposed approach behavior dealing with a sequence of random destructive events (i.e. Coverage Holes), is first presented. Compared to [36], the present work extends the payoff function to a more intelligent and efficient function using the Q-learning algorithm.

3. System model

A WSN can be regarded as a connected graph $G(V, E)$ with V is the set of vertices representing nodes and E is the set of edges representing links between nodes. Each sensor can be designed as a Unit Disk Graph (UDG) in a two-dimensional rectangular region, denoted by $\Delta = [x_{\min}; x_{\max}] \times [y_{\min;} y_{\max}]$ with a sensing radius R_s and a communication radius R_c [44]. For the sake of simplicity, we assume that all nodes in S have the same R_c. Let $S_1 \in S$ and $S_2 \in S$, the couple of nodes (S_1, S_2) are bi-directionally connected if S_1 is within the transmission range of S_2. Let E_i denote the starting energy of sensor node i. We suppose that each node is localized via GPS. Coverage Holes may appear due to the predictable or unpredictable death of the sensor nodes, such as battery power depletion or explosion in the post-placement stage. Coverage Hole i can be regarded as a circle with radius RH_i and center (x_{Hi}, y_{Hi}). This circle depicts the zone where the node failures take place. Coverage Holes with different and complex shape (convex and non-convex) can be easily estimated by combining multiple Coverage Holes of different centers and radii. Let D-nodes denote the set of damaged nodes and U-nodes the set of undamaged nodes, each sensor node $S_i \in S$ situated in the holes zone belongs to D-nodes; otherwise, it is regarded as an undamaged node and belongs to U-nodes [45]. For simplicity reasons, we assume that Coverage Holes locations are precisely detected.

3.1. Game theory

The Game theory (GT) is a mathematical model used to make strategic decisions in an uncertain environment. It is concerned with the modeling of situations to describe the possible behaviors of interdependent agents in order to determine the optimal strategy. Dealing with competitive situations in GT, the outcome of an agent's choice of action depends critically on the actions of other agents. In what follows, we present some mathematical background in game theory.

Definition 1. (Game theory). A game in strategic form $\xi = \{\tau, \Lambda, \upsilon\}$ is defined by three basic components:

(1) a player set $\tau = \{1, \ldots, n\}$, here, n is the number of players. We denote by $-i$ players other than player i.
(2) an action space $\Lambda = \{\Lambda_1, \Lambda_2, \ldots, \Lambda_n\}$, each player i has a strategy Λ_i where Λ_i is a series of players' actions.
(3) a utility function set υ where each player i has an utility function $\upsilon_i : \Lambda \to \mathbb{R}$ used to evaluate the players' payoff.

Denote by Λ_i the action profile $\{a_1, a_2, \ldots, a_m\}$, $\forall i \in \tau$. Denote by the strategy profile other than i, i.e. the set of actions of all players except i: $\{\Lambda_1, \ldots \Lambda_{i-1}, \Lambda_{i+1}, \ldots, \Lambda_n\}$.

Definition 2. (Nash Equilibrium (NA)). is a state in which no player intends to change his strategy given the strategies adopted by other players. Let ξ be a strategic form game, the strategy $(\ddot{a}_i, \ddot{a}_{-i})$ is NE of ξ if, for all $a_i \in \Lambda_i$ and for all $\ddot{a}_{-i} \in \Lambda_{-i}$, $\upsilon(\ddot{a}_i, \ddot{a}_{-i}) \geq \upsilon(a_i, \ddot{a}_{-i}), \forall i \in \tau$.

Definition 3. (exact potential game). A strategic form game Ψ is an exact potential game with potential function $\varphi : \Lambda \to \mathbb{R}$ if, for each $i \in \tau$, for each $a_{-i} \in \Lambda_{-i}$, and for each $a_i, \ddot{a}_i \in \Lambda_i$, we have:

$$\varphi(a_i, a_{-i}) - \varphi(\ddot{a}_i, a_{-i}) = \upsilon_i(a_i, a_{-i}) - \upsilon_i(\ddot{a}_i, a_{-i}) \tag{1}$$

In this paper, the Coverage Holes problem was formulated as a potential repeated multiplayer game. In this game, the sensor nodes represent the players which communicate with each other to heal the coverage gaps.

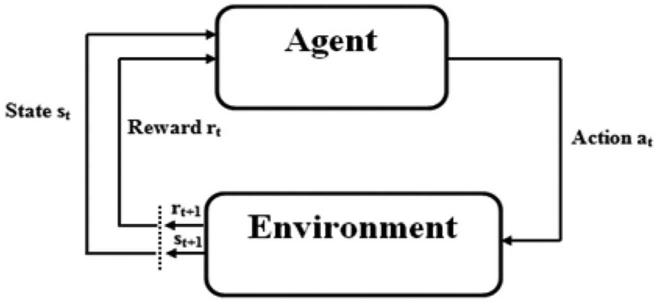


Fig. 1. Agent-environment interface for mono-agent Reinforcement Learning.

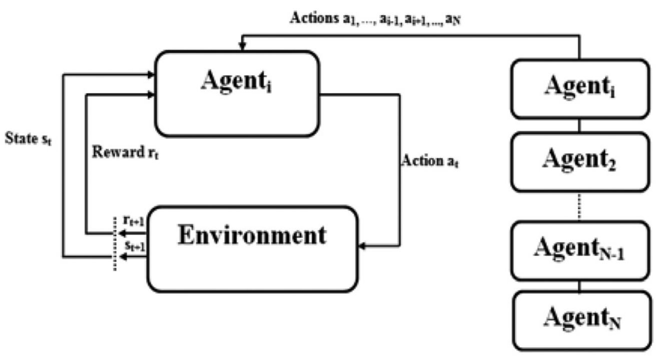


Fig. 2. Agent-environment interface for a joint-action.

3.2. Q-Learning

The Q-learning is a reinforcement learning technique based on the Markov Decision Processes (MDP), which mainly rely on the iterated interaction between the agent and the environment. In the Q-learning, an agent chooses to execute an action a_t at an instant t from the current state s_t which leads to a new state s_{t+1} and that provides the reward r_t (See Fig. 1). Thus, it consists of learning the actions to be performed according to the current state thanks to a previously acquired experience. The policy $pi(s)$ is a plan that instructs the agent (player) which moves to select every possible state (position). The player policy is gradually improved due to the mutual interactions between the learning system and the environment. However, in practice, most of the reinforcement learning algorithms do not directly work on policies; they rather use the iterative approximation of the evaluation function $Q(s, a)$. Thus, the $Q(s, a)$ value is defined as the estimated future benefits. Once these values are learned, the optimal strategy is to choose the action which corresponds to the maximum Q-value.

For the mono-agent system, the utility of a state is calculated as the maximum Q-value of this state regarding all possible actions. Nevertheless, for a multi-agent system, the optimality criterion cannot consider only the individual actions of agents since it depends on the joint actions of other players. Thereby, the extension of the Q-learning to the multi-agent framework (with N agents) should take into account other players' actions in the formulation of function (See Fig. 2).

In fact, Q-learning algorithms behave well with discrete states and actions. However, in real-world issues, the space is generally continuous with infinite states and action combinations. Q-learning cannot be directly applicable to the continuous variables. Space often needs to be discretized using approximation function such as gradient descent, linear combinations of features and neural network [46,47].

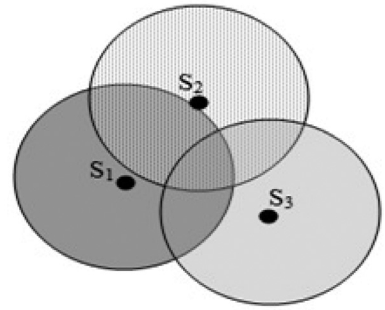


Fig. 3. Profit of sensor i.

4. Proposed hybrid recovery algorithm

In what follows we will present a novel game to heal Coverage Holes for mobile WSNs. Our distributed approach allows the sensor nodes to decide the appropriate topology control actions in an autonomous and decentralized way. The proposed game is performed using a Q-learning algorithm and is proven to be an exact potential game. In previous works, we optimized the initial deployment topology using multi-objective optimization approaches [48–50].

4.1. Coverage problem formulation

Based on the aforementioned issue, the players are the sensor nodes that can communicate simultaneously with each other to enhance their own coverage zones with minimal power consumption. Moreover, we suppose that all players are rational and each one seeks to increase his expected utility function according to its local vision of the network and its own prevision on actions that may be taken by other players. Reciprocal interactions between this set of players in order to maximize their payoff function (utility function) can be modeled as a repeated multi-player game. *Set of actions* At each time step $t > 0$, the set of strategies that node $i \in \tau$ may choose are the set of its possible actions $a_i \in \Lambda_i$. As mentioned earlier, an hybrid topology control uses both nodes reposition and power transmission adjustment. Consequently, each mobile sensor $i \in \tau$ can choose a combined action of changing its location p_i and its power transmission (the sensing range r_i): $a_i = (p_i, r_i), \forall a_i \in \Lambda_i$ i) Let us suppose that $p_i = (p_{x_i}, p_{y_i})$ is the current positions 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. ii) Let us suppose that $r_i, r'_i \in [R_{\min}, R_{\max}]$ are respectively the current and the next time 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. *Utility functions* Each node i has a utility or payoff function $\upsilon_i : \Lambda \to \mathbb{R}$, defined as:

$$\upsilon_i(a_i, a_{-i}) = \mu_\alpha P(a_i, a_{-i}) - \mu_\beta C(a_i) \quad (2)$$

Where $P(a_i, a_{-i})$ and $C(a_i)$ 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 game. At each time t, in a repeated game, players simultaneously select their action strategies and receive their utility functions which are specified by how much benefit should be gained and how much should be paid? When all the players properly predict the opponent players strategies and play the best response to their predictions, the emerging strategy profile is a Nash equilibrium state.

Each node has a Non-Overlapped Sensing Area (NOSA), defined as the zone covered by only the node in question (See Fig. 3). To maximize the total coverage area and reduce the coverage gaps, the sensor node should reduce overlapped zones with its neighboring nodes; therefore, maximizing profit function value is regarded as maximizing the NOSA value.

$$P(s_i, s_{-i}) = Disk_i \setminus \bigcup_{j \in N_i} Disk_j \tag{3}$$

Where N_i is the set of neighbor of node i. The cost represents the energy consumed by the sensor nodes while recovering Coverage Holes. This cost mainly consists of two types:

-The consumed energy while changing the sensing power denoted by E_{pow}

-The consumed energy while exercising a motion to change its positions denoted by E_{mot}. The energy related to changing the sensing power of node i is specified by:

$$E_{pow} = \sum_t \sum_{i=1}^{n} (((4\pi . R_{S_i}/\lambda)^2)\Delta t) \tag{4}$$

The energy related to the node motions is specified by:

$$E_{mot} = \sum_t \sum_{i=1}^{n} (motion(i)\Delta t) \tag{5}$$

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

$$E_{tot} = E_{pow} + E_{mot} \tag{6}$$

Lemma1 The specified coverage game ξ_{cov} is an exact potential game if there is a function $\varphi : \Lambda \to \mathbb{R}$, such that $\forall a_i, \ddot{a}_i \in \Lambda_i, \forall a_{-i} \in \Lambda_{-i}$:

$$\varphi(a_i, a_{-i}) - \varphi(\ddot{a}_i, a_{-i}) = \upsilon(a_i, a_{-i}) - \upsilon(\ddot{a}_i, a_{-i}) \tag{7}$$

With φ defined as follows:

$$\varphi(a_i, a_{-i}) = \mu_\alpha P_{tot}(a_i, a_{-i}) - \mu_\beta C(a_i, a_{-i}) \tag{8}$$

Where P_{tot} is the network overall coverage and C the sensor node energy consumption.

Proof. Let us consider any strategy $a := (a_i, a_{-i}) \in \Lambda, \forall i \in \tau$. We select an arbitrary strategy $\ddot{a} := (\ddot{a}_i, a_{-i})$ from Λ. The coverage game ξ_{cov} is defined to be the exact potential game. In order to prove this, we change the strategy of node i from a_i to $\ddot{a}_i$ without changing the others.

We observe that:

$$\begin{aligned}
&\varphi(a_i, a_{-i}) - \varphi(\ddot{a}_i, a_{-i}) \\
&= (\mu_\alpha P_{tot}(a_i, a_{-i}) - \mu_\beta C(a_i, a_{-i})) - (\mu_\alpha P_{tot}(\ddot{a}_i, a_{-i}) - \mu_\beta C(\ddot{a}_i, a_{-i})) \\
&= \mu_\alpha (P_{tot}(a_i, a_{-i}) - P_{tot}(\ddot{a}_i, a_{-i})) - \mu_\beta (C(a_i, a_{-i}) - C(\ddot{a}_i, a_{-i})) \\
&= \mu_\alpha (P_{tot}(a_i, a_{-i}) - P_{tot}(\ddot{a}_i, a_{-i})) - \mu_\beta (C(a_i) - C(\ddot{a}_i))
\end{aligned}$$

Normally, any change in the overlapped covered area of node i does not have any impact on the overall net coverage since it is already covered by at least another sensor node (See equation).

$P_{tot}(a_i, a_{-i}) - P_{tot}(\ddot{a}_i, a_{-i}) = P(a_i, a_{-i}) - P(\ddot{a}_i, a_{-i})$

So, we have:

$$\begin{aligned}
&\varphi(a_i, a_{-i}) - \varphi(\ddot{a}_i, a_{-i}) \\
&= \mu_\alpha (P(a_i, a_{-i}) - P(\ddot{a}_i, a_{-i})) - \mu_\beta (C(a_i) - C(\ddot{a}_i)) \\
&= \upsilon(a_i, a_{-i}) - \upsilon(\ddot{a}_i, a_{-i})
\end{aligned}$$

Therefore, when player i switches from strategy a_i to strategy $ddota_i$, the variation in the potential function φ equals to the variation in the payoff/ utility function. By referring to definition 3, we can affirm that the proposed coverage game is proven to be an exact potential game. □

Algorithm 1 Distributed Payoff-based Q-learning algorithm.

1: **Initialization:** At $t = 1$, all sensor nodes maintain their initial random positions (initial states).Initialize $Q(s, a_i, a_{-i})$ arbitrary, $NbC(s, a_{-i}) \leftarrow 0$ and $n(s) \leftarrow 0$; $\forall s \in S$; $\forall a_i \in \Lambda_i$ and $\forall a_{-i} \in \Lambda_{-i}$.Let $\alpha \in [0, 1]$ be the initial learning rate and ε be the initial exploration rate.

2: **Choose action and update state:** At each time $t \geq 2$, each sensor node $i$updates its state respecting the following rules:

i observe the current s: $n(s) \leftarrow n(s) + 1$

ii choose the exploration rate ε

- Exploration: with probability ε return a random action
- Exploitation: otherwise, with probability $(1 - \varepsilon)$ select an action a_i by solving:$\arg\max_{a_i} \sum_{a_{-i}} \frac{NbC(s,a_{-i})}{n(s)} Q(s, a_i, a_{-i})$

3: **Learning:** Observing the opponents' actions a_{-i}, the reward $R(s, a_i)$ and the next state:$Q(s, a_i, a_{-i}) \leftarrow (1 - \alpha)Q(s, a_i, a_{-i}) + \alpha[R(s, a_i) + \gamma\pi(s')]NbC(s, a_{-i}) \leftarrow NbC(s, a_{-i}) + 1$where$\pi(s') = max_{a_i} \sum_{a_i} \frac{NbC(s',a_{-i})}{n(s')} Q(s', a_i, a_{-i})andC(s, a_{-i})$ is the number of times the opponent has played action a_{-i} in state s.

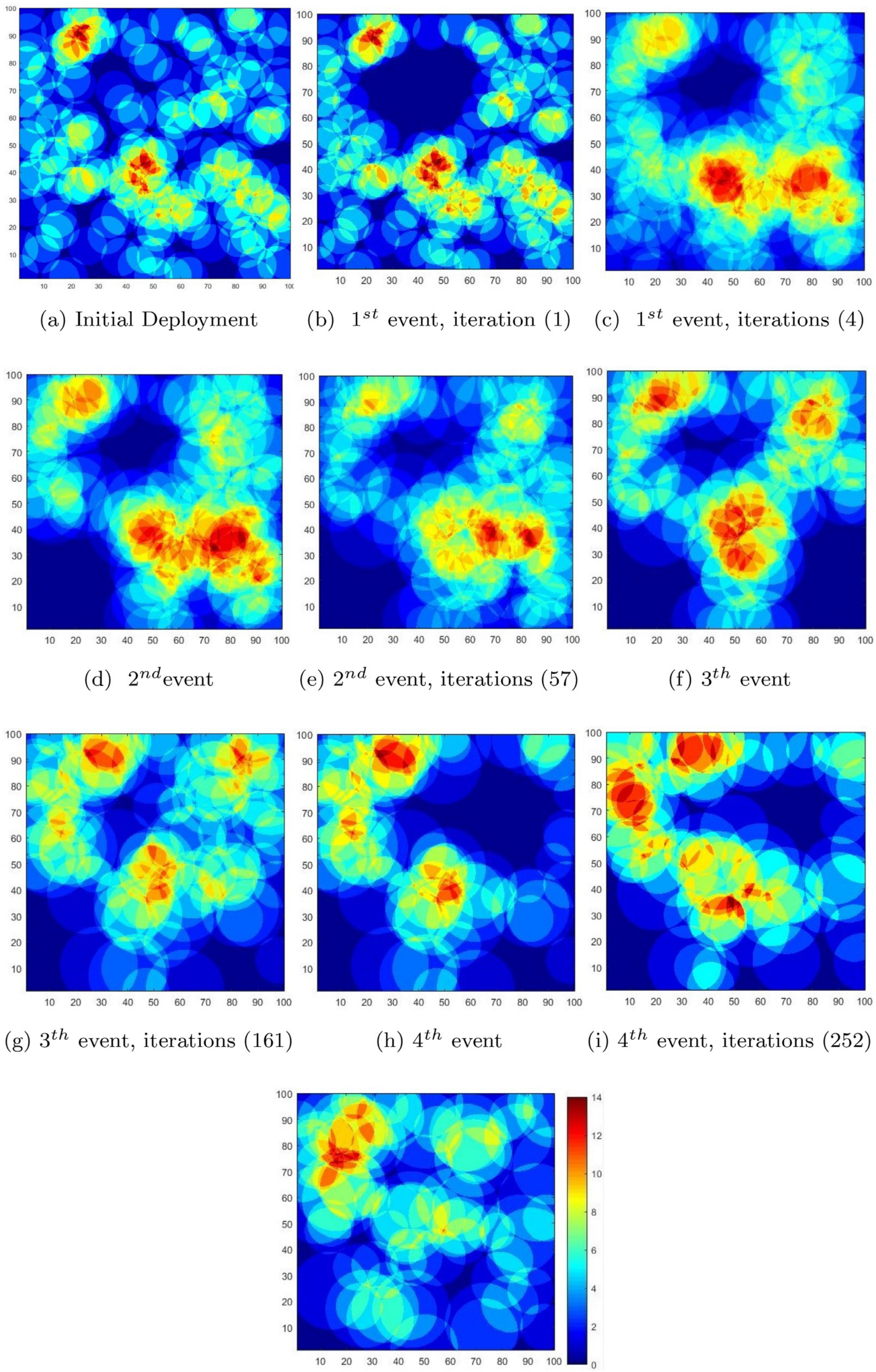
4: **Repeat:** Sensor i executes step 2 and step 3 until the ending condition is met.

4.2. Distributed payoff-based Q-learning algorithm

The payoff function depends not only on the profile of the sensor node actions but also on the actions of its neighbors. Consequently, a distributed learning approach is the most appropriate solution since it needs only the payoff received from the previous steps. The players adapt their behaviors to change of other players' behaviors. In this work, we propose a new distributed payoff-based Q-learning algorithm to perform the previously formulated coverage game. Here, we deal with a multi-players game; therefore, a multi-agent reinforcement learning concept is used.

The proposed algorithm can be described in pseudo-code 1.

The payoff function switches between exploring and exploiting processes, to find the best action sequence. In fact, the exploration process assists nodes to discover the unknown environment rapidly, whereas the exploitation one enables them to keep the effective scenarios. The metric ε controls the trade-off between the exploitation of the system's previously acquired knowledge and the exploration of the environment. This work extends the potential game presented in [36] by improving the distributed payoff based learning, which only considers the two last most successful actions of mobile nodes. For the previously considered topology control scheme, sensor nodes were meant to behave not only in an autonomous and decentralized way but also in a rational way. The new payoff function based on the Q-learning algorithm takes into account the node environment interaction and opponents' behavior, which gives the sensor nodes a more targeted and effective response while healing the newly-formed Coverage Holes. Besides, the Q-learning algorithm ensures a good balance between the acquisition of information about new interactions (exploration) and the choice of actions that seems to lead to viable rewards (exploitation). Contrary to [36], the temporary strategy is not considered here. The temporary strategy means that we do not have to change both sensing power and node positions over iterations; one of these metrics can be kept. It should also be noted that the proposed approach was proven to be a potential game, which, theoretically, can ensure its convergence towards a unique Nash equilibrium. In the following section, we will describe the extensive simulations that were performed in order to validate the proposed algorithm and confirm its efficiency.



(a) Initial Deployment (b) 1^{st} event, iteration (1) (c) 1^{st} event, iterations (4)

(d) 2^{nd}event (e) 2^{nd} event, iterations (57) (f) 3^{th} event

(g) 3^{th} event, iterations (161) (h) 4^{th} event (i) 4^{th} event, iterations (252)

(j) 4^{th} event, iterations (500)

Fig. 4. Overall coverage dealing with random successive Coverage Holes Value of Stopping Criterion for N = 250 Nodes for an AoI without obstacle.

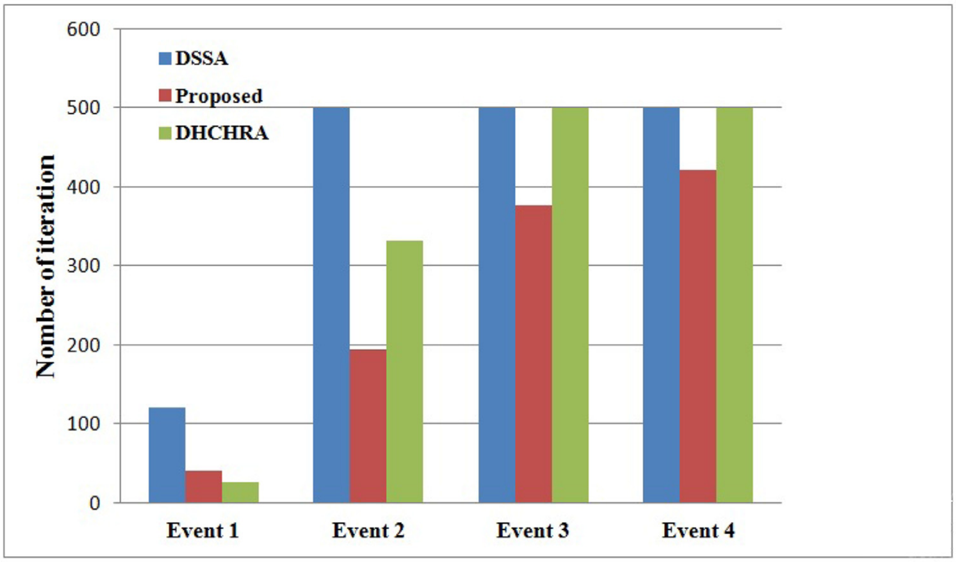


Fig. 5. Value of Stopping Criterion for N = 250 Nodes for an AoI without obstacle.

Table 1
Simulation parameters setting.

Variables names	Variables values
Wireless Sensor Network Parameters	
Ns : Number of sensor Nodes	[50, 250]
R_S^i: Sensing radius	U ~ [7, 15] m
Rc : Communication radius	30 m
Δ: Area of interest	[1, 100] × [1, 100]m_2
Algorithm parameters	
ε: Exploration rate	[0, 0.5]
Coverage holes parameters	
Number of Damages	4
R_H^i	U ~ [10, 20] m

5. Simulations and results

To validate the previously defined mathematical formulations, we needed to implement our approach and compare the obtained results with those of the previously proposed approaches. In this work, performance criteria were specified in terms of the total coverage proportion, energy related to the change of sensing power, the energy related to node motions, and the total energy consumption. The performance of the proposed approach was tested on three topology scenarios: the no obstacle scenario, the rectangular obstacle scenario, and the T shaped obstacle scenario. The power transmission adjustment is not possible in all WSNs applications. Thus, we considered another simulation scenario that uses sensor nodes which support only positions adjustment in no obstacle AoI. The obtained results were compared with those reached by the DSSA algorithm [28] as well as those of [36].

5.1. Simulation parameters

For our simulations, we considered a square region δ with a 100 m long edge divided into a predefined number of squares. The size of each square was equal to 1 m^2 and its center was the point to be covered. The simulation trials were performed using Matlab. Nodes were randomly deployed in the AoI δ with a uniform distribution. The transmission range was fixed to 30 m and the sensing ranges were chosen randomly between 7 m and 15 m (see Table 1). The network connectivity is assumed to be full if the communication radius is larger than the double of the sensing radius [51].

As explained above, Q-learning is not appropriate for continuous space and an approximation function is needed to discretized the space. In this work, a simple form of discretization was used. In fact, the agent (node) performs two actions, namely changing the location and the power transmission. For each algorithm iteration, the node changes its location by moving one step in the following directions: left, right, up, or down. The sensor sensing radius takes on a value in this set: 7, 9, 11, 13 or 15. For the rest of this paper, the Distributed Hybrid Coverage Hole Recovery Approach proposed in [36] is abbreviated as (DHCHRA).

ε was used to balance the exploration and the exploitation processes in the algorithm. Using a value interval given in table helps the Q-learning algorithm to converge towards an optimal solution [46,47].

For a reliable comparison, we extended and simulated the two previously mentioned algorithms. In this paper, we considered the energy consumed for the movement of the sensor and that of change of its sensing range so that the sensor was proposed to raise its sensing range to avoid needlessly prevalent movements, especially with the frequently arising Coverage Holes. Four damage events with different sizes were generated for each scenario. The algorithms remained active, after each new emerging Coverage Hole until an overall coverage of 95% was reached. If this threshold had not been satisfied, algorithms would stop running when the number of iterations reach 500. The stopping criterion, i.e. the maximal number of iterations was set to 500 to prevent the collective death of sensor nodes due to the exhaustion of the residual energy in their batteries.

5.2. Simulation results

In this section, we present and analyze the obtained results. Our results are compared with those given by DSSA and DHCHRA. All the algorithms instances were tested on the same initial node distributions using four different scenarios.

5.2.1. Area of interest without obstacle scenario

In this scenario, a detection area without physical obstacles was presented and four damaged events were generated. The damage event time sequences were 1 s, 14 s, 57 s, and 161 s. Fig. 4 shows a snapshot sample to prove the ability of the proposed game theory-based Q-learning algorithm to recover Coverage Holes. The color in

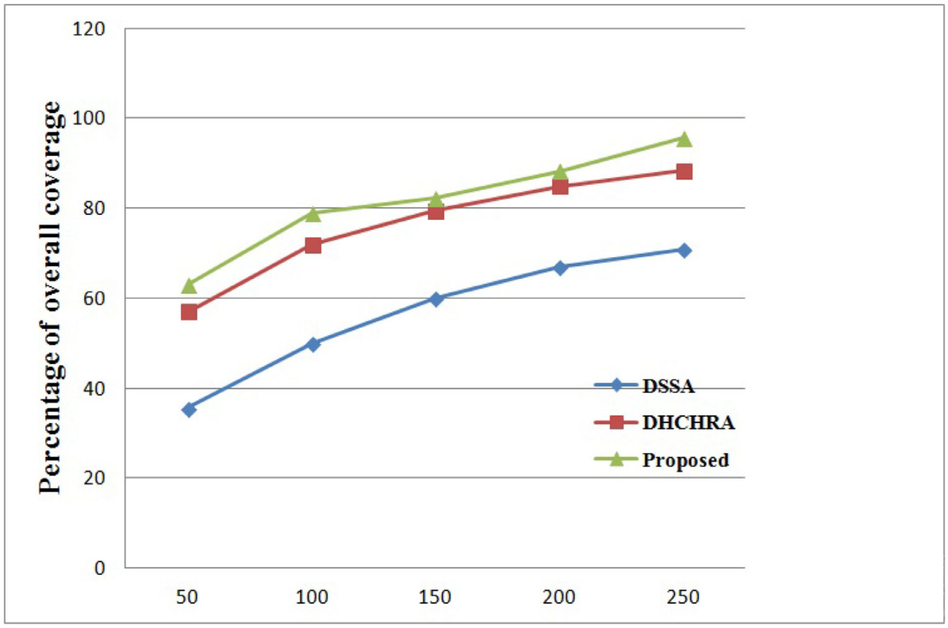
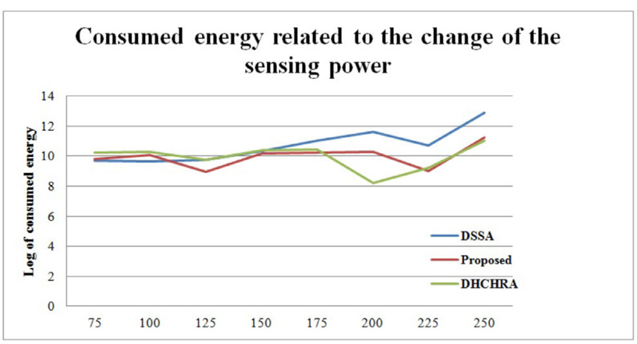
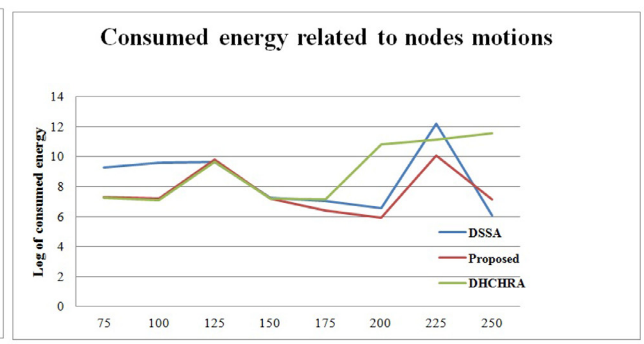


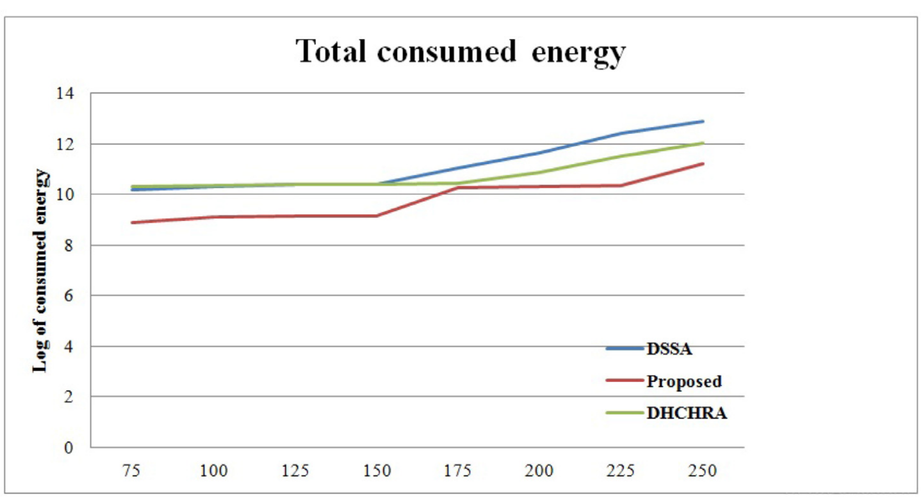
Fig. 6. Percentage of overall coverage vs. the number of nodes.



(a) The consumed energy related to the change of the sensing power

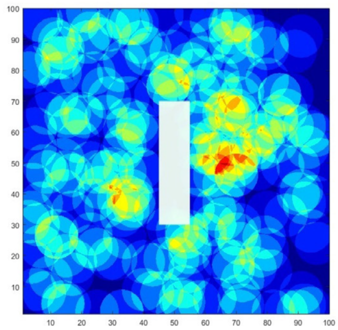


(b) The consumed energy related to node motions

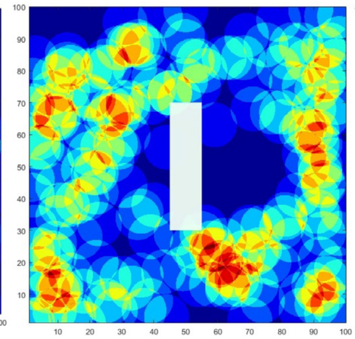


(c) The total consumed energy

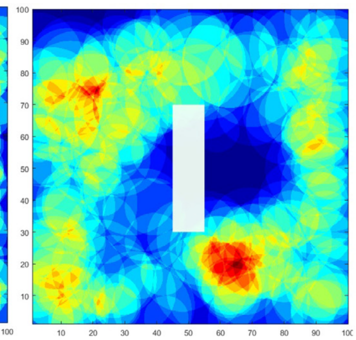
Fig. 7. Network consumed energy vs. the number of nodes for an AoI without obstacle.



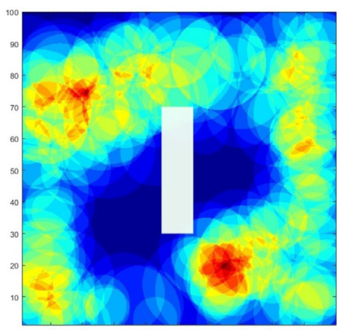
(a) Initial Deployment



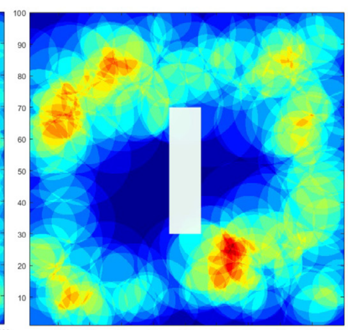
(b) 1^{st} event, iteration (1)



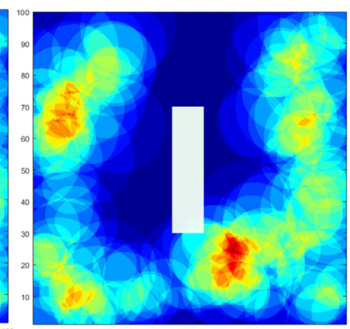
(c) 1^{st} event, iterations (5)



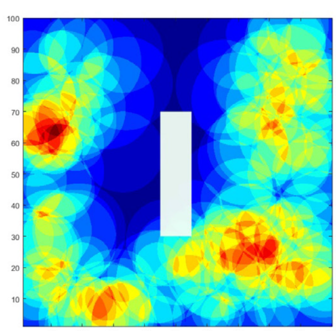
(d) 2^{nd}event



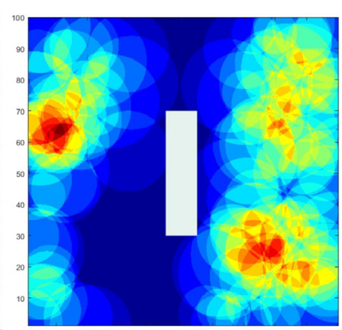
(e) 2^{nd} event, iterations (69)



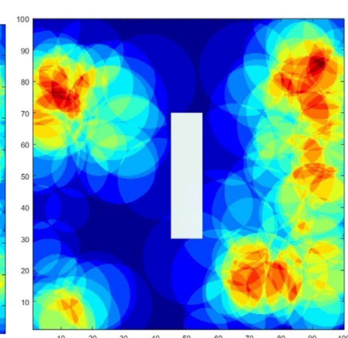
(f) 3^{th} event



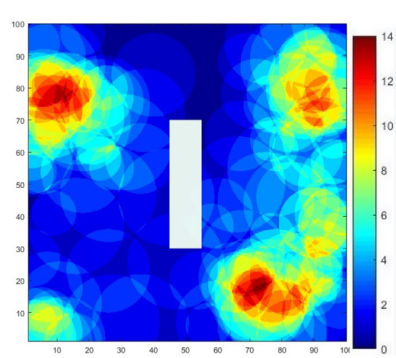
(g) 3^{th} event, iterations (189)



(h) 4^{th} event

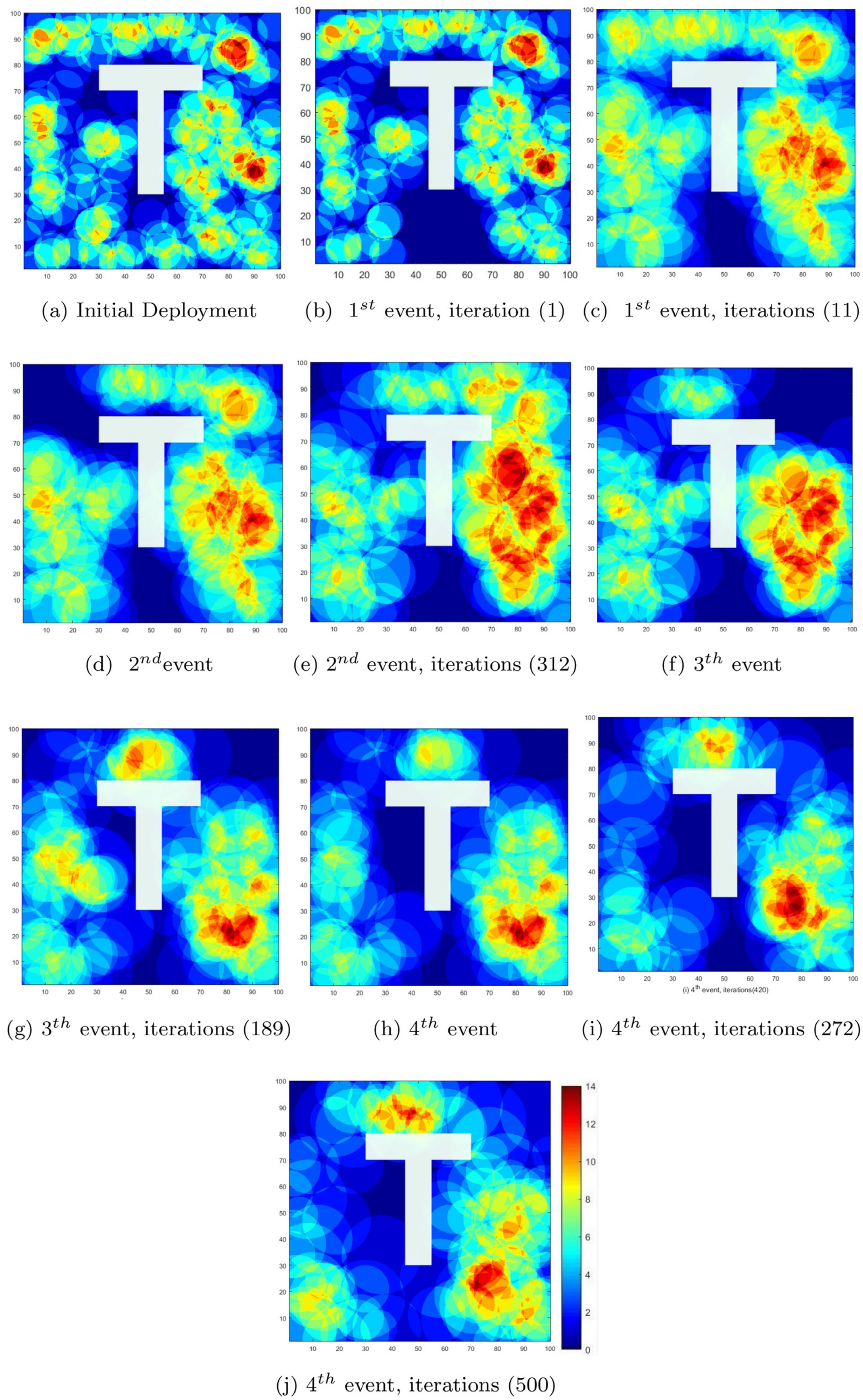


(i) 4^{th} event, iterations (272)



(j) 4^{th} event, iterations (500)

Fig. 8. Overall coverage dealing with random successive Coverage Holes for an AoI with rectangular shape obstacle.



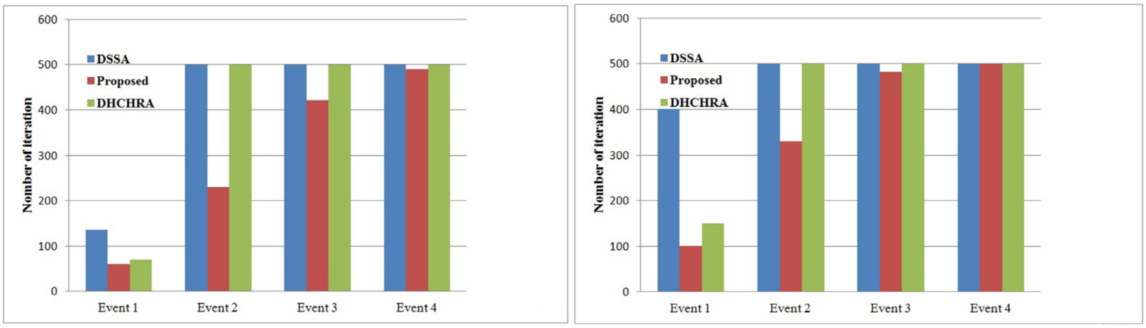
(a) Initial Deployment (b) 1^{st} event, iteration (1) (c) 1^{st} event, iterations (11)

(d) 2^{nd}event (e) 2^{nd} event, iterations (312) (f) 3^{th} event

(g) 3^{th} event, iterations (189) (h) 4^{th} event (i) 4^{th} event, iterations (272)

(j) 4^{th} event, iterations (500)

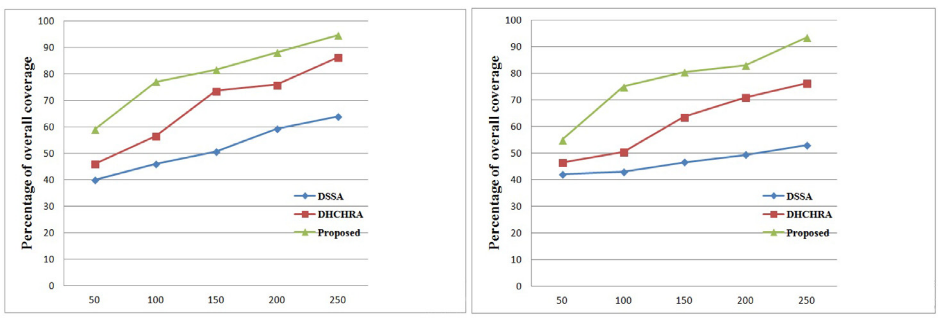
Fig. 9. Overall coverage dealing with random successive Coverage Holes for an AoI with rectangular shape obstacle.



(a) Rectangular shape obstacle.

(b) T shape obstacle.

Fig. 10. Network consumed energy vs. the number of nodes for an AoI with obstacles.



(a) Rectangular shape obstacle.

(b) T shape obstacle.

Fig. 11. Percentage of overall coverage vs. the number of nodes for an AoI with obstacles.

the figure specifies the number of sensors that can cover a zone. Precisely, the number of sensors in the zone decreases as the color changes to blue; or it increases as the color changes to red.

The dark blue, however, indicates that a zone is not covered by any sensor node. Fig. 4b,d,f and h show the new emerging Coverage Holes owing to successive destructive events. The behavior of our potential game dealing with the four Coverage Holes is illustrated in Fig. 4c, e, g and j. These figures clearly show the efficiency of our proposed topology control scheme to heal the network and maintain the highest coverage threshold, after successive Coverage Hole events. Sensor nodes adjust both sensing range and position in order to quickly and accurately heal the coverage gaps.

Fig. 5 shows the average of the required number of iterations after the appearance of each coverage hole for$NS = 250$ nodes for the given algorithms. The number of iterations increases with the appearance of successive events, for the three algorithms. The stopping criterion (number of iterations equal to $\succ$500) was reached by DSSA and the DHCHRA after event 2 and event 3, respectively. The proposed approach converged faster compared to the prior counterparts. Indeed, it succeeded to recover the four Coverage Holes and reach the aforementioned coverage threshold (95% of overall coverage area) before reaching the stopping criterion.

Fig. 6 presents the percentage of the overall coverage for DHCHRA, DSSA, and the proposed algorithm considering the number of the sensor nodes (50 to 250 nodes) deployed in the AoI. The overall coverage is evaluated during the occurrence of the consecutive four Coverage Holes. As can be seen in this figure, the proposed algorithm outperforms the above-mentioned algorithms in terms of overall network coverage. We simulated the efficiency of DSSA, the DHCHRA and the proposed approach in terms of energy consumption.

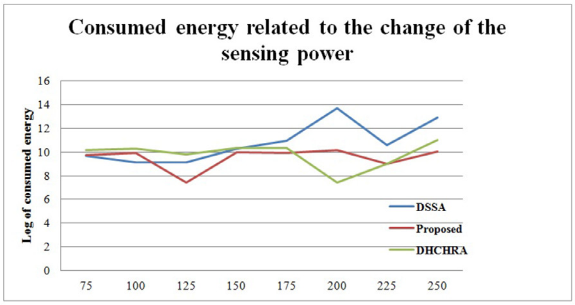
Fig. 7 a–c show the consumed energy related to the change of the sensing power, the consumed energy related to node motions, and the total consumed energy, respectively. The results prove that our approach outperforms the two approaches in terms of energy consumption. As presented above (see Fig. 5), the proposed approach has a rapid response to the consecutive Coverage Holes compared to the DSSA and the DHCHRA. These results mean that sensor nodes take their optimal positions in terms of coverage with fewer possible movements, which empowers the network to dissipate fewer amounts of energy (see Fig. 7) and, thus, extend the network survivability.

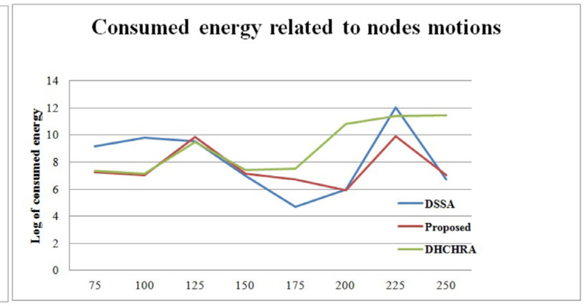
5.2.2. Area of Interest with obstacles scenario

To bring our simulations closer to reality, two obstacles shapes were considered in our simulation scenarios: a rectangular shape and T shape. The obstacles were placed in the center of the AoI. The sensor nodes can be placed on the border of the obstacles. The results of those two scenarios will be presented as follows: Figs. 8 and 9 show a snapshot sample to prove the ability of the proposed game theory-based Q-learning algorithm to recover Coverage Holes in an area of detection that contains a rectangular and T shaped obstacle, respectively.

Figs. 8b,d,f,h and 9b,d,f,h show the new emerging Coverage Holes owing to successive destructive events for rectangular obstacle scenario and T shaped obstacle scenario, respectively. The damage event time sequences were 1 s, 5 s, 70 s, and 273 s, and 1 s, 11 s, 169 s, and 312 s for rectangular obstacle scenario and T shaped obstacle scenario, respectively. Other sub-figures show the behavior of the approach dealing with the four Coverage Holes in a detection area with rectangular obstacle shape. The figures present the capacity of the proposed approach to eliminate Coverage Holes, even in the presence of obstacles.

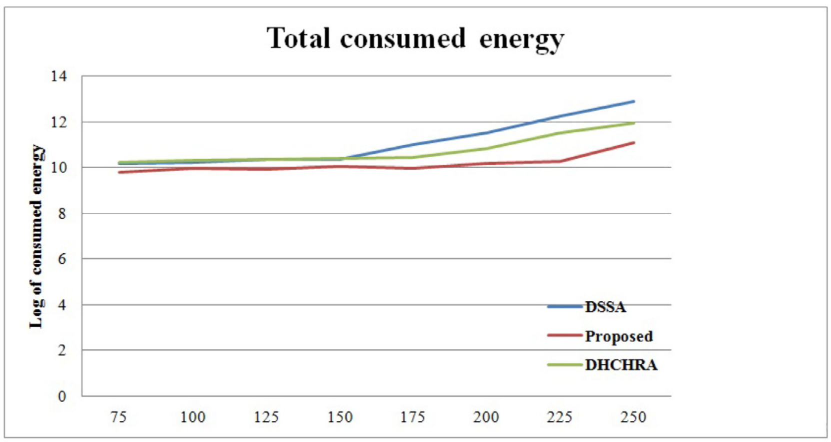
Fig. 10 shows the average of the number of iterations required to recover each coverage hole for the proposed obstacle scenarios. For the rectangular obstacle scenario, the stopping criterion was reached by the DSSA and the DHCHRA after event 2. The four Cov-





(a) The consumed energy related to the change of the sensing power

(b) The consumed energy related to node motions



(c) The total consumed energy

Fig. 12. Network consumed energy vs. the number of nodes for an AoI with rectangular shape obstacle.

erage Holes were recovered by our proposed approach, which succeeded to achieve the coverage threshold (95% of overall coverage area) before reaching the stopping criterion. For the T shaped obstacle scenario, the stopping criterion was reached by the DSSA and the DHCHRA after event 1 and event 2, respectively. However, the proposed approach recovers the four Coverage Holes before reaching 92% of overall coverage area.

Fig. 11 presents the percentage of the overall coverage for the DHCHRA, blue the DSSA, and the proposed algorithm varying the number of the sensor nodes (from 50 to 250 nodes) deployed in a detection area with obstacles. The overall coverage is evaluated during the occurrence of the consecutive four Coverage Holes. As shown in this figure, the proposed algorithm outperforms other algorithms in terms of the overall network coverage, even in the presence of the obstacles.

Figs. 12 and 13 present the efficiency of the DSSA, the DHCHRA and the proposed approach in terms of energy consumption for detection area with obstacles. Figs. 12a,b,c and 13a,b,c show the consumed energy related to the change of the sensing power, the consumed energy related to node motions and the total consumed energy, respectively. The results prove that our approach performs better than both of the two approaches in terms of energy consumption.

5.2.3. Topology control scheme without power adjustment

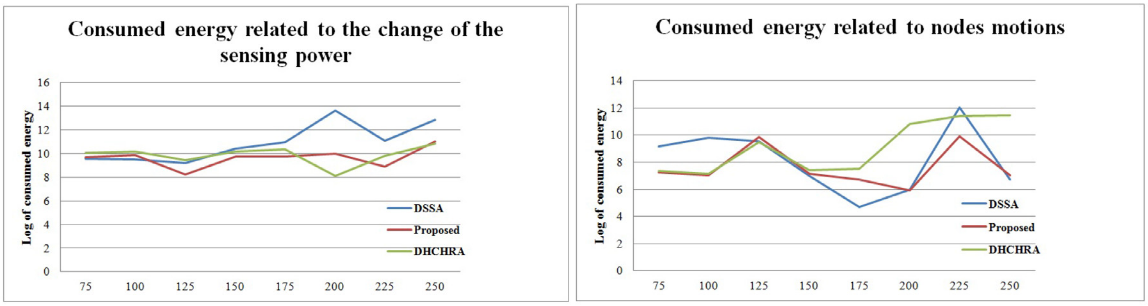
Many WSN applications do not support power transmission adjustment. Therefore, in this section, a topology control strategy without power control tested in an AoI without obstacle is presented. The action of the nodes is limited to the position adjustment.

Fig. 14 shows the percentage of the total coverage area of the proposed approach for the different suggested scenarios, namely an AoI with no obstacle, an AoI with rectangular shape obstacle, an AoI with T shaped obstacle, and a control topology scheme with no power adjustment for an AoI without obstacle. The percentage of the total coverage was evaluated varying the number of the sensor nodes (from 50 to 250 nodes). This figure proves that the proposed approach reaches a satisfactory coverage rate with 50, 100, 150, and 200 nodes for each suggested scenario. Also, it reaches the coverage threshold (95% of overall coverage area) with 250 nodes, except for the T shaped obstacle scenario that reaches only 93% of overall area coverage.

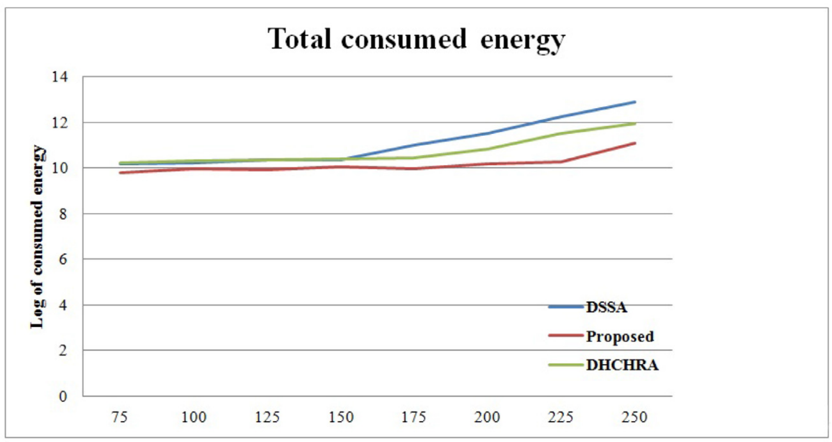
Fig. 15 shows the energy consumption results of our approach for all the proposed scenarios. For the no obstacle scenario, the rectangular obstacle scenario, and the T shaped obstacle scenario the energy consumption value is convergent. However, the energy consumption of the proposed approach without power adjustment obviously increased.

5.3. Results and discussions

We have demonstrated, through a wide range of experimental scenarios, that the proposed Q learning-based approach for WSN topology control gives better results compared to other approaches. The advantages of using such an approach can be summarized as follows: Unlike other approaches (studies based on mathematical models or simulations, assuming fixed experimental parameters), this solution is adaptable to variations in the state of the network. Besides, thanks to the robust machine learning algorithm, namely the Q learning, this approach can calibrate itself



(a) The consumed energy related to the change of the sensing power (b) The consumed energy related to node motions



(c) The total consumed energy

Fig. 13. Network consumed energy vs. the number of nodes for an AoI with T shape obstacle.

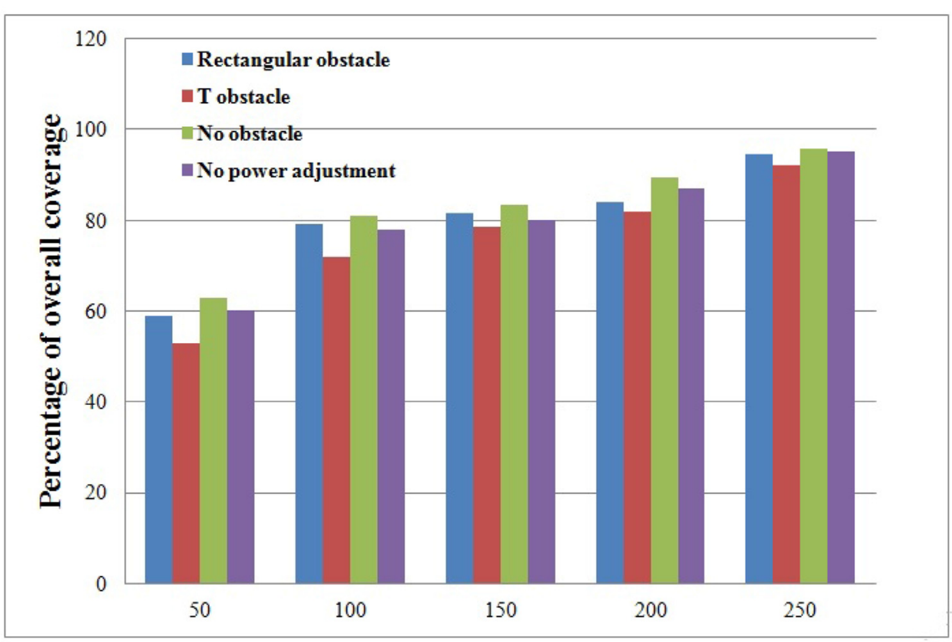


Fig. 14. Percentage of overall coverage vs. the number of nodes of the proposed approach for the different proposed scenarios.

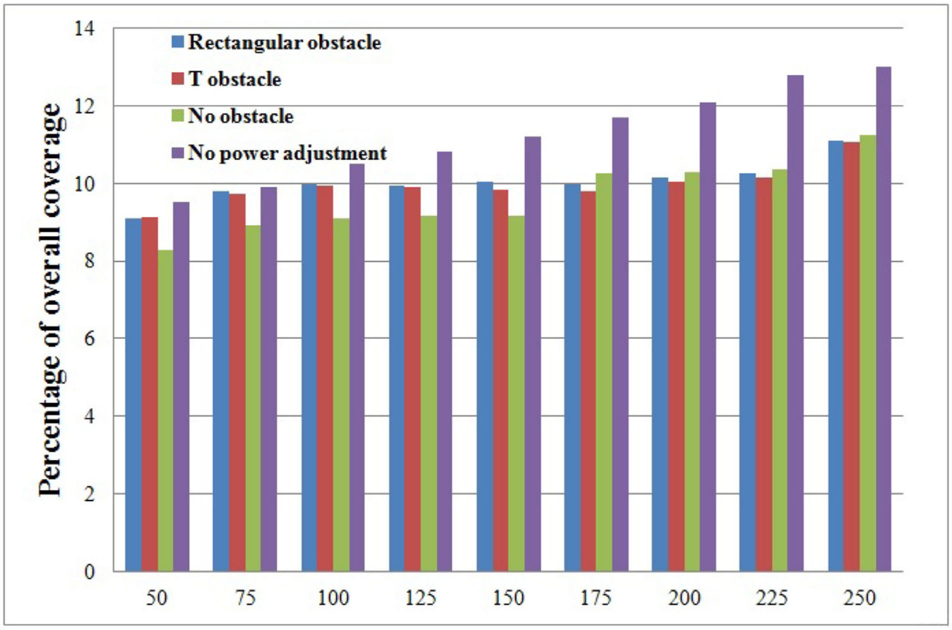


Fig. 15. The total consumed energy vs. the number of nodes of the proposed approach for the different proposed scenarios.

to the newly acquired knowledge. Consequently, Q-learning based payoff function helps the proposed coverage game to find the Nash equilibrium and reach the predefined overall coverage threshold in a detection area without obstacles more rapidly (see Fig. 5). Even in an environment which contains different shapes of obstacles, our approach succeeds to overcome its counterparts in terms of overall coverage (see Fig. 11 a and b). In addition, Q-learning takes into account not only the best previous actions of an agent (sensor), but also the predicted neighbors' actions. Such behavior helps sensors to effectively forecast the best strategy that should be taken and perform the most accurate and successful decision that enhances both the overall coverage and energy consumption while recovering Coverage Holes. Therefore, the topology control scheme induced by our proposed approach outperforms other approaches in terms of energy consumption and coverage for an AoI without obstacles scenario and also for an AoI with obstacles scenario (see Figs. 6 and 7). However, the speed of our approach to recovers the Coverage Holes while respecting the predefined coverage sill (95% of overall coverage area) decreased in scenarios with obstacles. This coverage sill could not be reached with more complicated obstacle scenarios (with T shaped obstacle only 92% of overall coverage area was achieved). In fact, those results can be explained by the effect of obstacle on the ease of movement for some nodes, spacially sensor nodes located in the proximity of the obstacle border (see Figs. 5 and 10).

The coverage rate of the proposed approach with no power adjustment is convergent compared to that of the original version. However, energy consumption rose compared to the original version. Indeed, such type of topology control scheme has limited the actions of the nodes to reposition. Some coverage gaps do not need the movement of sensors to recover it, a power adjustment of the hole neighbor's nodes can solve the problem and cover the zone with a minimum amount of energy consumption (see Figs. 14 and 15).

6. Conclusion

In the present work, we proposed a new hybrid Coverage Hole recovery approach for WSNs which relies on both the sensing power control and node relocation using a game theory based on Q-learning algorithm. The efficiency of the proposed algorithm was tested on a network topology and Coverage Hole events via an elaborated simulation study. The simulation results prove that the proposed approach outperforms the other Coverage Hole recovery algorithms in terms of overall coverage and energy consumption. As part of future works, we are planning to enhance the technical maturity of the proposed approach by taking into consideration the approximation functions, such as neural networks. We also think of applying adequate communication protocols to improve the proposed approach. For better results, we plan to design a payoff function based on deep Q-learning.

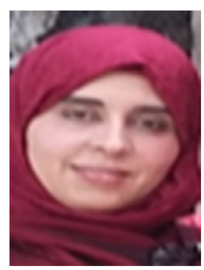
Acknowledgment

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program for the funds.

References

[1] I.F. Akyildiz, Wireless sensor networks: a survey, Comput. Netw. 38 (4) (2002) 393–422, doi:10.1016/S1389-1286(01)00302-4.
[2] B. Doina, Self-organizing and self-healing schemes in wireless sensor networks, Comput. Commun. Netw. (2009) 293–304, doi:10.1007/978-1-84882-218-4_11.
[3] B.H. C. Benjie J. Kyle, M. Robert, Span: an energy efficient coordination algorithm for topology maintenance in ad hoc wireless networks, ACM Wireless Netw. J. (2002) 481–494, doi:10.1145/381677.381686.
[4] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Networks 52 (12) (2008) 2292–2330, doi:10.1016/j.comnet.2008.04.002.
[5] I.F. Akyildiz, T. Melodia, K.R. Chowdury, Wireless sensor network survey, Comput. Networks 52 (6) (2008) 888–896, doi:10.1016/j.comnet.2008.04.002.
[6] M. Younisa, I.F. Senturkb, K. Akkayab, S. Leec, F. Senel, Topology management techniques for tolerating node failures in wireless sensor networks: a survey, Comput. Netw. 53 (2014) 254–283, doi:10.1016/j.comnet.2013.08.021.
[7] N. Jabeura, N. Sahliblljaz, M. Khanc, Survey on sensor holes: a cause-effect-solution perspective, Procedia Comput Sci 19 (2013) 1074–1080, doi:10.1016/j.procs.2013.06.151.
[8] C. Ma, J. He, H.H. Chen, Z. Tang, Coverage overlapping problems in applications of IEEE 802.15.4 wireless sensor networks, IEEE WCNC, 2013. 4364–369
[9] H. Grari, F. Hajjej, R. Ejbali, M. Zaied, An optimal approach for deployment sensors in WSN, JCM 13 (3) (2018) 124–128, doi:10.1016/j.procs.2013.06.151.
[10] M.A. Labrador, P.M. Wightman, Topology control in wireless sensor networks: With a companion simulation tool for teaching and research, Signals Commun., 2009.

[11] P. Santi, Topology control in wireless ad hoc and sensor networks, ACM Computing Surveys (CSUR) 37 (2) (2005) 164–194, doi:10.1145/1089733.1089736.
[12] M.Y. Donmeza, R. Kosar, C. Ersoy, An analytical approach to the deployment quality of surveillance wireless sensor networks considering the effect of jammers and coverage holes, Int. J. Comput.Telecommun. Netw. 54 (2010) 3449–3466.
[13] X.J. Du, M. Zhang, K.E. Nygard, S. Guizani, H.H. Chen, Self-healing sensor networks with distributed decision making, Int. J. Sens. Netw. 2 (2007) 289–298.
[14] X. Huang, D.K. Hunter, Distributed coordinate-free hole recovery, in: IEEE ICC Workshops., 2008, pp. 189–194.
[15] P.K. Sahoo, J.P. Sheu, Limited mobility coverage and connectivity maintenance protocols for wireless sensor networks, Comput. Netw. 55 (2011) 2856–2872.
[16] G.W. G. Cao, T.F. Porta, Movement-assisted sensor deployment, IEEE Trans. Mob. Comput. 5 (2006) 640–652.
[17] A.A. Abbasi, M. Younis, A survey on clustering algorithms for wireless sensor networks. computer communications, ACM Comput. Surv. (CSUR) 30 (1415) (2007) 2826–2841, doi:10.1016/j.comcom.2007.05.024.
[18] N. Xu, A. Huang, T.W. Hou, H.H. Chen, Coverage and connectivity guaranteed topology control algorithm for cluster-based wireless sensor networks, Wireless Commun. Mob. Comput. 12 (1) (2012) 23–32, doi:10.1002/wcm.887.
[19] J. Corts, S. Martnez, F. Bullo, Spatially-distributed coverage optimization and control with limited-range interactions, ESAIM: Control Optim. CalculusVari. 11 (4) (2005) 691–719, doi:10.1051/cocv:2005024.
[20] D. Fudenberg, J. Tirole, Game Theory, MIT Press, 1994.
[21] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning : a survey, J. Artif. Intell. Res. 4 (1996) 237–285, doi:10.1613/jair.301.
[22] E. Solan, N. Vieille, Stochastic games, in: National Academy of Science of the United States of America, 2015, pp. 327–332.
[23] M.L. Littman, Markov games as a framework for multi-agent reinforcement learning, in: 11th International Conference on Machine Learning, 1994, pp. 157–163.
[24] J. Hu, M.P. Wellman, Q-Learning in general-sum stochastic games, J. Mach. Learn. 19 (2003) 1039–1069.
[25] L. Zhou, H. Tang, H.L. Q. Jiang, Dynamic power management strategies for a sensor node optimized by reinforcement learning, Int. J. Comput. Sci. Eng. 13 (1) (2016) 24–37.
[26] S. Shamshirban, A. Patel, N.B. Anuar, M.L.M. Kiah, A. Abraham, Cooperative game theoretic approach using fuzzy q-learning for detecting and preventing intrusions in wireless sensor networks, Eng. Appl. Artif. Intell. 32 (2014) 228–241, doi:10.1016/j.engappai.2014.02.001.
[27] G. Liu, X. Wang, X. Li, J. Hao, Z. Feng, Esrq: An efficient secure routing method in wireless sensor networks based on q-learning, in: IEEE International Conference On Trust, Security And Privacy In Computing And Communications12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018, pp. 1–6.
[28] N. Heo, P.K. Varshney, A distributed self spreading algorithm for mobile wireless sensor networks, in: IEEE Wireless Communications and Networking, 2003, pp. 1597–1602.
[29] Y. Yoo, D. Agrawal, Mobile sensor relocation to prolong the lifetime of wireless sensor networks, in: Vehicular Technology Conference, 2003, pp. 193–197.
[30] Y. Maali, A. Rafiei, M. Abolhasan, D. Franklin, F. Safaei, A fuzzy logic node relocation model in wsns, in: 6th International Conference on Signal Processing and Communication Systems, 2012.
[31] H. Shu, Q. Liang, Fuzzy optimization for distributed sensor deployment, in: IEEE Wireless Communications and Networking Conference, 2005, pp. 1903–1908.
[32] J.A. Torkestani, An energy-efficient topology control mechanism for wireless sensor networks based on transmit power adjustment, Wireless Pers. Commun. 82 (4) (2015) 2537–2556, doi:10.1007/s11277-015-2363-9.
[33] S.L. F. Miao, J. Zhang, G. Zhou, L. Gu, T. He, J.A. Stankovic, S.H. Son, G.J. Pappas, Atpc: adaptive transmission power control for wireless sensor networks, ACM Trans. Sensor Netw. (TOSN) 12 (1) (2016) 1–6, doi:10.1145/2746342.
[34] P.K. Sahoo, W.C. Liao, Hora: a distributed coverage hole repair algorithm for wireless sensor networks, IEEE Trans. Mob. Comput. 14 (7) (2015) 1397–1410, doi:10.1109/TMC.2014.2359651.
[35] M. Zhu, S. Martnez, Distributed coverage games for energy-aware mobile sensor networks, SIAM J. Control Optim. 51 (1) (2013) 1–27, doi:10.1137/100784163.
[36] M. Abolhasan, Y. Maali, A. Rafiei, W. Ni, Distributed hybrid coverage hole recovery in wireless sensor networks, IEEE Sens. J. 16 (23) (2016) 8640–8648, doi:10.1109/JSEN.2016.2613539.
[37] L.P. Kaelbling, M.L. Littman, A.P. Moore, Reinforcement learning : a survey, J. Artif. Intell. Res. 4 (1996) 237–285, doi:10.1613/jair.301.
[38] L.S. Shapley, Stochastic games, in: Proceedings of the National Academy of Science, 1953, pp. 327–332.
[39] M.L. Littman, Markov games as a framework for multi-agent reinforcement learning, in: The 11th International Conference on Machine Learning, 1994, pp. 157–163.
[40] J. Hu, M.P. Wellman, Nash q-learning for general-sum stochastic games, J. Artif. Intell. Res. 4 (2003) 1039–1069.
[41] A. Greenwald, K. Hall, Correlated-q learning, in: The Twentieth International Conference on Machine Learning, 2003, pp. 242–249.
[42] M.L. Littman, Friend-or-foe q-learning in general-sum stochastic games, in: The Eighteenth International Conference on Machine Learning, 2001, pp. 322–328.
[43] C. Claus, C. Boutilier, The dynamics of reinforcement learning in cooperative multiagent systems, in: The fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence, 1998, pp. 746–752.
[44] H. Ren, M.Q.H. Meng, Using game approach to control bioeffects for wireless body sensor networks, in: IEEE International Conference on Robotics and Biomimetics, 2006, pp. 17–20.
[45] X. Li, M.R. Lyu, A novel coalitional game model for security issues in wireless networks, in: IEEE Global Telecommunications Conference, 2008.
[46] Y. Chizhov, Reinforcement learning with function approximation: Survey and practice experience, in: The International Conference Modelling of Business, Industrial and Transport Systems, 2008, pp. 1–7.
[47] G.J. Gordon, Reinforcement learning with function approximation converges to a region, in: The Neural Information Processing Systems Conference., 2000, pp. 996–1002.
[48] F. Hajjej, R. Ejbali, M. Zaied, Multi objective nodes placement approach in WSN based on nature inspired optimisation algorithms, in: The Second International Conference on Advances in Sensors, Actuators, Metering and Sensing, 2017, pp. 1–6.
[49] F. Hajjej, R. Ejbali, M. Zaied, Sensor node deployment approach in wireless sensor network based on multi-objective flower pollination algorithm, Sensors Transducers 6 (2017) 1–8, doi:10.1109/JSEN.2016.2613539.
[50] F. Hajjej, R. Ejbali, M. Zaied, A new optimal deployment model of internet of things based on wireless sensor networks, in: 15th International Wireless Communications & Mobile Computing Conference (IWCMC), 2019, pp. 2092–2097.
[51] A. Ghosh, S.K. Das, Coverage and connectivity issues in wireless sensor networks: a survey, pervasive and mobile computing, Pervasive Mob. Comput. 4 (2008) 303–334, doi:10.1016/j.pmcj.2008.02.001.



Faten Hajjej received her graduate degree in Computer Engineering from the National School of Engineers of Sfax, University of Sfax, in 2009. She has been pursuing the Ph.D. degree with the Research Group on Intelligent Machines (REGIM-Lab) at the University of Sfax under the supervision of Dr. Ridha Ejbali and Prof. Mourad Zaied. Her research interests include the internet of things, wireless sensor networks, multi objective optimization, and nature inspired optimization algorithms.

Monia Hamdi received her Bachelor of Engineering degree in Information Technology from Telecom SudParis, Paris-Saclay University, France, in 2008 and obtained in the same year the Master of Science degree in Telecommunications and Networks from Institut National Polytechnique, Toulouse, France. She received the Ph.D. degree in Computer Science from the University Rennes 1, France in 2012. She worked from 2012 to 2017 as an assistant professor in the Higher Institute of Computer Science and Multimedia, Gabs University, Tunisia. From March o August 2015, she was a Visiting Researcher with the Department of Science and Technology, Link ping University, Sweden. She is currently an assistant professor in the College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, KSA. She has been a member of the technical program committees of several IEEE conferences such as IEEE International Conference on Sciences of Electronics, Technologies of Information and Telecommunications 2016 and 2018, IEEE International Wireless Communications and Mobile Computing Conference 2017. Her research interests include cooperative communications networks and learning and self-organising cellular system networks.

Ridha Ejbali: received his Ph.D. degree in Computer Engineering, Master's degree and computer engineer degree from the National Engineering School of Sfax Tunisia (ENIS) in 2012, 2006 and 2004, respectively. He has been an assistant technologist at the Higher Institute of Technological Studies, Kebili Tunisia since 2005. He joined the Faculty of Sciences of Gabes Tunisia (FSG) where he has been an assistant in the Department computer sciences since 2012. His research area is now in pattern recognition and machine learning using Wavelets and Wavelet networks theories.

Mourad Zaied: received his HDR, Ph.D. degree in Computer Engineering and Master's of Science from the National Engineering School of Sfax in 2013, 2008 and 2003, respectively. He obtained a degree in Computer Engineering from the National Engineering School of Monastir in 1995. Since 1997 he has served in several institutes and faculties at the University of Gabes as a teaching assistant. He joined the National Engineering School of Gabes (ENIG) in 2007 where he is currently an Associate Professor in the Department of Electrical Engineering. He has been a member of the REsearch Group on Intelligent Machines laboratory (REGIM) http://www.regim.org in the National Engineering School of Sfax (ENIS) since 2001. His research interests include computer vision and image, and video analysis. These research activities are centered on Wavelets and Wavelet networks and their applications to data classification and approximation, pattern recognition and image, audio and video coding, and indexing