Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7560167, 13 pages
https://doi.org/10.1155/2018/7560167

Research Article

Novel Learning Algorithms for Efficient Mobile Sink Data Collection Using Reinforcement Learning in Wireless Sensor Network

Santosh Soni[1] **and Manish Shrivastava**[2]

[1] *Department of Information Technology, School of Studies (Engineering & Technology), Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh 495009, India*
[2] *Department of Computer Science and Engineering (Supervisor), School of Studies (Engineering & Technology), Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh 495009, India*

Correspondence should be addressed to Santosh Soni; santoshsoni.77@gmail.com

Received 2 April 2018; Revised 1 July 2018; Accepted 14 July 2018; Published 16 August 2018

Academic Editor: Dajana Cassioli

Generally, wireless sensor network is a group of sensor nodes which is used to continuously monitor and record the various physical, environmental, and critical real time application data. Data traffic received by sink in WSN decreases the energy of nearby sensor nodes as compared to other sensor nodes. This problem is known as hot spot problem in wireless sensor network. In this research study, two novel algorithms are proposed based upon reinforcement learning to solve hot spot problem in wireless sensor network. The first proposed algorithm RLBCA, created cluster heads to reduce the energy consumption and save about 40% of battery power. In the second proposed algorithm ODMST, mobile sink is used to collect the data from cluster heads as per the demand/request generated from cluster heads. Here mobile sink is used to keep record of incoming request from cluster heads in a routing table and visits accordingly. These algorithms did not create the extra overhead on mobile sink and save the energy as well. Finally, the proposed algorithms are compared with existing algorithms like CLIQUE, TTDD, DBRkM, EPMS, RLLO, and RL-CRC to better prove this research study.

1. Introduction

This research study started with a valid question of how to enhance the network lifetime of WSN with better energy optimization of sensor nodes by using reinforcement learning. The solution of this question lies in improving energy efficient WSN algorithms which is a key research area already addressed by various literatures in the last decades. Therefore, it is expected that sensor nodes perform sleep and wake-up mechanism to better utilize their energy for enhancement of network lifetime. The concept of clustering also works very well in this manner.

Due to multihop communication, generally the sensor nodes which are near to base stations always become overloaded because they are intermediate nodes between base station and remaining wireless sensor network for data forwarding to the base station [1]. This situation happens to be a hot spot problem [2, 3] where SNs near to sink node send their own data as well as other nodes data. This leads to decrease the performance of wireless sensor network significantly. Therefore we have motivated towards research in sink mobility which has emerged in WSNs to properly handle the hot spot problem and to decrease the energy communication overheads [4]. Traditionally, mobile sink [5, 6] needs to visit every cluster head [2] to collect the data, leading to longer mobile sink traversal path which in turns creates data delivery latency [7, 8] and higher energy consumption. For this reason, we proposed RLBCA and ODMST algorithm upon reinforcement learning. We proposed the visit of mobile sink only to interested cluster heads by receiving a request message packet for collection of data. However, design of such on-demand mobile sink traversal path [3] is a challenging task as it highly depends upon coverage of network, data delivery, energy efficiency, and lifetime of network.

Reinforcement learning (RL) techniques [9] are used here, being an unsupervised class of learning in the field of machine learning which permits an agent to learn the behaviour in new environment. The prime goal of the agent is to generate actions which increase the rewards in future. Later on these rewards lead to formulate optimal policy. The elements of RL can be formalized using the Markov decision process (MDP) framework [9–11]. MDPs [9] consist of states, actions, and transitions between states and reward function definition. Thus, the use of RL techniques can largely improve the WSNs performance significantly.

Finally, our contribution in this research study is as follows:

(i) Proposed reinforcement learning based clustering algorithm (RLBCA) to form cluster heads

(ii) Proposed on-demand mobile sink traversal (ODMST) algorithm to collect data.

(iii) Comparison of these above-mentioned algorithms with existing algorithm like CLIQUE [12], TTDD [13, 14], EPMS [5], DBRkM [3], RLLO [15], and RL-CRC [16] to better prove our simulation results.

2. Related Works

This section presented the review of recent research studies including energy efficient routing, network lifetime enhancement, coverage, clustering, and reinforcement learning based WSN solutions. In [17], the authors proposed geometric model for mobile sink which has performed very well on various performance matrices. In [1, 13], TTDD protocol is designed where WSN is partitioned into virtual grids based upon the mobile sink node. The path for mobile sink is based upon the grid node [18, 19] which eliminated the hot spot problem. However this process of developing grid consumes more energy of SNs. In [5], here author focused on energy efficient routing and clustering based on PSO algorithm. Here authors have also presented a technique which extends the network lifetime by eliminating the traffic load of the gateways whose remaining energy is beyond a particular threshold value; however authors have considered only failure of the gateways due to complete energy depletion. The EPMS algorithm [5] performs the virtual clustering by using PSO algorithm to improve the network performance. Here the selection of cluster head depends upon the reception of data to control the movement of mobile sink. However this algorithm did not solve the WSNs transmission coverage problem. In [5], authors focused on the delivery latency minimization problem in WSN along with the deployment of mobile sink on a plane randomly; here the proposed algorithm performs well in terms of shortening data delivery latency and reducing route length. However, transmission issue affected the performance of WSN. In [20–23], authors proposed data dissemination framework which is called tree overlay grid, to handle mobile target detection where multiple mobile sinks appear in WSN to consume less energy along with a longer network lifetime; however implementation of this algorithm on real time WSN created complexity. In [3, 24], author described how information local to each node can be shared without extra overhead as feedback to neighbouring nodes which enabled efficient routing to multiple sinks. Such type of situation arises in WSNs with multiple mobile users collecting data from a monitored area; here authors formulate the problem as a reinforcement learning task and applied Q-Routing techniques to derive a solution. Evaluation of the resulting FROMS protocol demonstrates its ability to significantly decrease the network overhead over existing approaches. Here authors proposed two algorithms RkM and DBRkM for path formation of mobile sink. The RkM algorithm worked to determine a path by joining the SNs through one hop communication where DBRkM generated a delay bound path. However every SN has equal load of data aggregation and the sojourn time of mobile sink is negligible. In [25], authors proposed EAPC method which constructed a data collection path and selected the eligible sensors to work as a collection point head. The EAPC method constructed a minimum spanning tree which is rooted at the base station. This method improves the network lifetime and energy consumption but a little bit lacks throughput while increasing numbers of SNs. In [26], author presented cluster based routing known as I-UMDPC where route delays sensitive data to mobile nodes within a time period. However complexity of this algorithm is a little bit higher than existing approaches. In [4], the problem of selecting an optimal cluster is formulated as MDP which showed good performance and energy consumption minimized by determining an optimal number of clusters for intra- and intercluster communications. In [12, 27], CLIQUE algorithm was used for data clustering which saved cluster head selection energy by using reinforcement learning to enable nodes to independently decide whether or not to act as a cluster head on a per packet basis; however on the setting of nonuniform data dissemination paradigm requires more work. In [24], reinforcement learning based clustering algorithm is proposed to address energy and primary user detection challenges in WSN; here Q-value slows the convergence of the proposed algorithm due to the long learning period. In [28], authors presented survey of multiobjective optimization in WSN which includes various performance metrics along with very useful algorithms. In [29], authors presented proactive way to enhance network lifetime, coverage, and discovery of redundant nodes with well-defined simulation results. In [30], an energy efficient routing algorithm is presented for multiple mobile sink which advocates the presence of less than three mobile sinks for collection of data. In [16, 30], RLLO and RL-CRC algorithms uniformly distributed the consumption of energy and further enhanced the PDR ratio with better topology control.

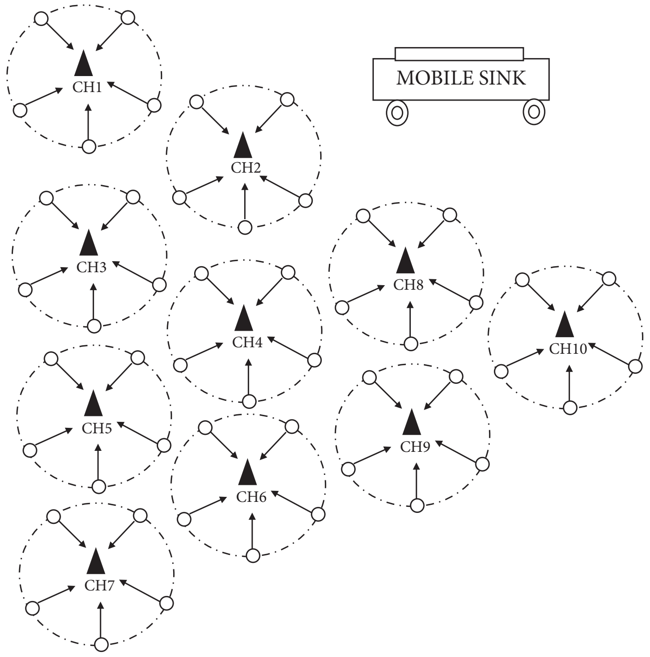


Figure 1: Architecture of mobile sink traversal.

3. System Model and Problem Formulation

This section presented network environment model, basis assumptions, energy model, problem formulation, and our contribution.

3.1. Network Environment Model. We have deployed multiple sensor nodes in random topology [31] to a rectangle area with a radius of R. The basic architecture of mobile sink traversal is shown in Figures 1 and 2. All the senor nodes are static and homogeneous in nature [32]. The entire sensor network environment has equal sectors. Source nodes have the liberty to adjust transmission power as per the distance to target nodes.

3.2. Basic Assumptions. We have made following assumptions in this research study:

(i) All the deployed WSN nodes are static and homogenous in nature.

(ii) All the WSNs nodes are equipped with same amount of initial energy [33].

(iii) Any physical hurdles/obstacle is not present in the network environment.

(iv) The mobile sink [34, 35] is able to collect the data from the cluster heads in proper time.

3.3. Energy Model. In this research study, we considered first radio energy model [23, 30] as energy model for the calculation of energy consumption. Generally energy consumption works in two modes: transmission and reception. Equation (1) shows the transmission of l-bit message (consumption of energy):

$$E_{Tx}(l,d)) = \begin{cases} l.E_{elec} + l.\varepsilon_{fs}.\mathrm{d}^2, & \text{when } \mathrm{d}_0 > d \\ l.E_{elec} + l.\varepsilon_{mp}.\mathrm{d}^4, & \text{when } \mathrm{d}_0 \le \mathrm{d} \end{cases} \quad (1)$$

$$E_{Rx}(l) = l.E_{elec} \quad (2)$$

where E_{elec} represents the consumption of energy. ε_{fs} and ε_{mp} represent the coefficient of free space and multipath fading model. Equation (2) shows the calculation of reception energy consumption.

Table 1: Proposed phases of mobile sink traversal path as per Figure 2.

(i) Advertisement of sink position	(ii) Request from cluster heads to collect the data
(iii) Collection of data from cluster heads	(iv) Distance calculation in the case of multiple request from cluster heads
(v) Mobile sink traversal path	

Table 2: Routing table.

Cluster Head ID	Position	Distance

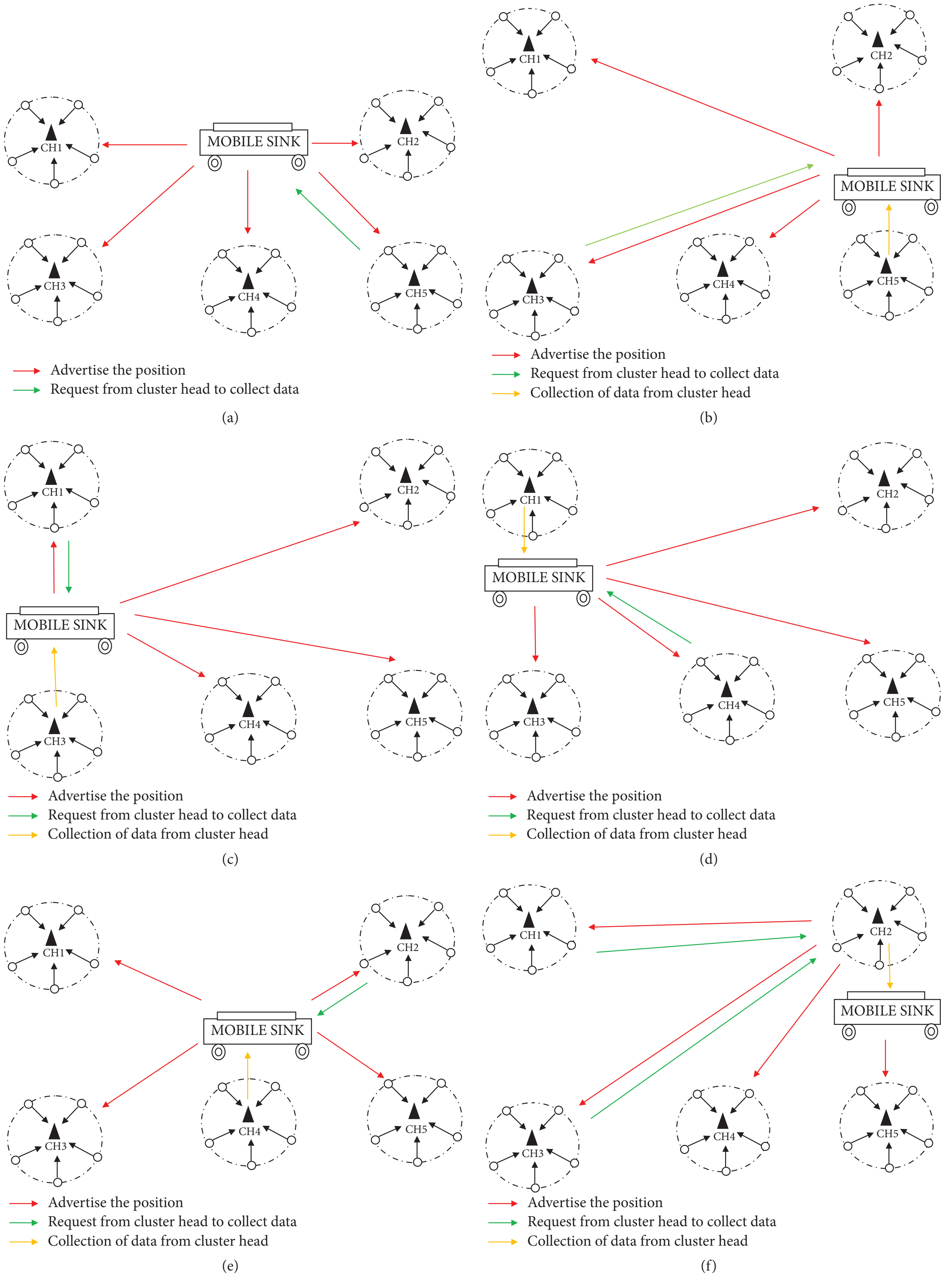
3.4. Problem Formulation and Contribution. The key performance factors of WSN are network lifetime and energy consumption. The lifetime of WSN is counted in terms of whenever first node dies. We have provided following contribution in this research study:

(i) Proposed reinforcement learning based clustering algorithm (RLBCA)

(ii) Proposed novel algorithm for on-demand mobile sink traversal (ODMST).

The proposed mobile sink traversal path is shown in Figure 2 based upon Table 1 and Equation (4).

Figure 2 shows the formation of proposed mobile sink traversal path based upon Tables 1 and 2, (4), and Algorithm 4. It is clear from Figure 2 that initially MS advertises its current position to all CHs. CHs send their request message to MS if any. MS calculates the distance between CHs and MS by using (4) and then creates and executes the mobile sink traversal path. During the traversal of MS, if any CHs send their request again then MS updates the traversal path as per the shortest distance and execute it to collect the data.


MOBILE SINK
CH1
CH2
CH3
CH4
CH5
Advertise the position
Request from cluster head to collect data
Collection of data from cluster head

(a)

(b)

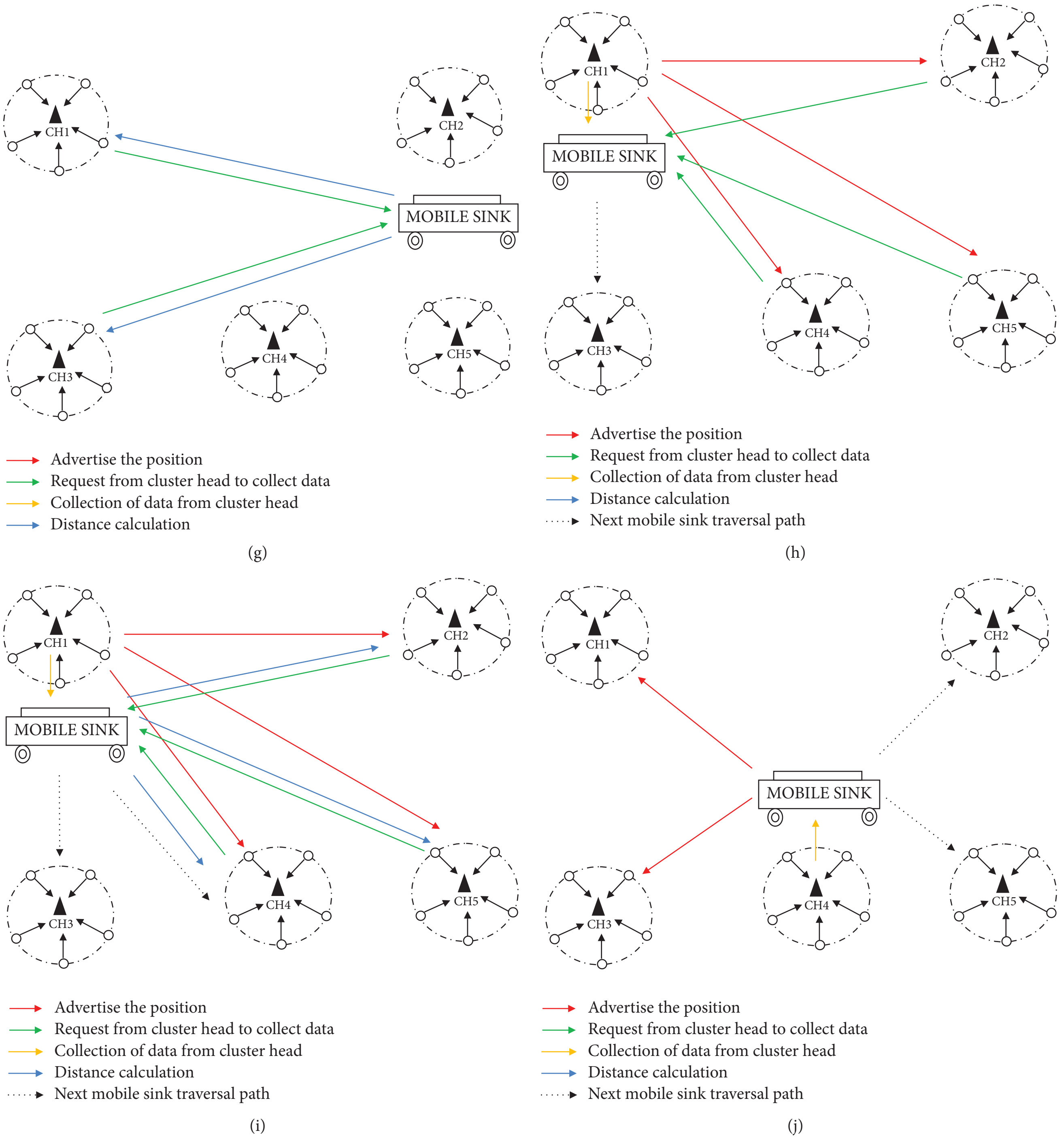
(c)

(d)

(e)

(f)

Figure 2: Continued.


CH1
CH2
MOBILE SINK
CH3
CH4
CH5
Advertise the position
Request from cluster head to collect data
Collection of data from cluster head
Distance calculation
Next mobile sink traversal path

(g) (h) (i) (j)

Figure 2: Proposed mobile sink traversal path.

Finally, as per Figure 2, the first round mobile sink traversal path works as follows:

$$\begin{aligned}&[\mathbf{CH5} \longrightarrow \mathbf{CH3} \longrightarrow \mathbf{CH1} \longrightarrow \mathbf{CH4} \longrightarrow \mathbf{CH2} \longrightarrow \mathbf{CH1}\\ &\quad \longrightarrow \mathbf{CH3} \longrightarrow \mathbf{CH2} \longrightarrow \mathbf{CH4} \longrightarrow \mathbf{CH5} \longrightarrow \mathbf{CH4}\\ &\quad \longrightarrow \mathbf{CH5} \longrightarrow \mathbf{CH2}]\end{aligned} \tag{3}$$

$$\mathrm{d}\left(\mathrm{x},\mathrm{y}\right) = \sqrt{\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \tag{4}$$

4. Reinforcement Learning

The reinforcement learning technique presents what to perform and how to react to present actions for maximizing the

```
For each state-action pair (s, a)
Initialize the table entry Q(s, a) to zero
        Observe the current state s
Do loop:
        Select an action a and execute it
        Receive immediate reward r
        Observe the new state s′
        Update the table entry for Q(s, a) as follows:
         Q(s, a) =r+ ϒ max a′ Q(s′, a′)
        s=s′
Selected action:
        π(s) =arg max a Q(s, a)
Exploration strategy
        $P(ai \mid S) = \frac{kQ(s, a)}{\sum kQ(s, a)}$
End Loop
```

Algorithm 1

```
Step 1. Input policy $\pi_0$
            $I \longleftarrow 0$
Step 2. Repeat
Search $Q^{\pi i}$
   $\pi_{i+1}(s) \longleftarrow \arg\max_{a \epsilon A} Q^{\pi i}$ (state, action) ∀ state
            $I \longleftarrow I + 1$
Step 3. Until $\pi_k = \pi_{k-1}$
Step 4. Output $\pi^* = \pi_k$, $Q^* = Q^{\pi k}$
```

Algorithm 2

Table 3: Representation of RL based clustering elements.

State	Set of neighbouring cluster S = (0, 1, 2...n). A state represents neighbouring cluster head CH which can be selected as $\sum_k^e s = v \in S$ [9–11, 24, 37]
Action	Set of actions in state A = (0, 1, 2...n). Here, an action $\sum_k^e a = 0$ selects cluster head CH and compute rewards [9, 11, 24, 37]
Reward	The computation of cumulative reward $\sum_{k+1}^e max$ is based upon the selection of action and state [9, 11, 24, 37]

reward value to develop the policy [9, 36]. Basically RL has various basic components like agent, action, state, reward, policy, value function, and environment model. Mainly RL is based upon MDP [9] which in turn includes temporal difference and $\boldsymbol{\varepsilon}$-greedy selection approach [24, 37] as a selection and mathematics approach. The basic learning process of RL is shown in Figure 3.

The basic reinforcement learning algorithm works as shown in Algorithm 1.

RL technique also performs policy iteration which has been described in Algorithm 2.

The basic components for clustering by reinforcement learning are given in Table 3.

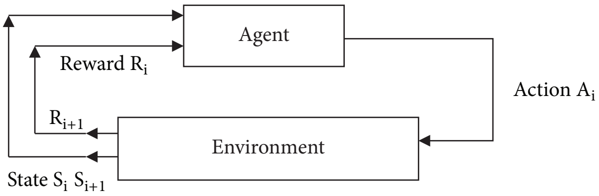


Figure 3: The agent learning environment.

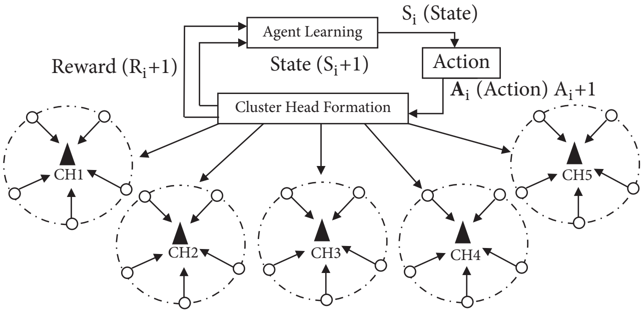


Figure 4: Cluster head formation using RL.

Table 4: RL model for clustering.

Action	$a_i^i \epsilon$ A = {1, 2...j}: every action a_i^i represents a next-hop neighbour node j. Here J represents the number of node i's neighbour nodes [12, 37].
Reward	$a_{t+1}^i a_i^i = 1$ where $a_{t+1}^i a_i^i$ represents the link cost to the next-hop neighbour node [12, 37].

5. The Proposed Algorithms

This section highlights the clustering of SNs and formation of cluster heads based on RLBCA and ODMST algorithms.

5.1. Clustering of SNs by Using Reinforcement Learning.

In this section, we proposed RLBCA in which WSN node works as a learning agent. These learning agents learn the energy level of nearest neighbour to form clusters based upon certain policy. Markov decision process (MDP) [9, 37] is calculated to find cluster. The MDP contains state, action, reward, and policy. The learning agent uses temporal difference method to learn from network environment to draw action policy. The RL model is used for clustering (Table 4).

From Table 4, it is clear that RL model for clustering is encoded with every SN to calculate the cost of a route which goes to cluster head node based upon certain Q-value update $Q_{t+1}^i a_i^i$. The action a_i^i shows the selection of next-hop node j to forward data packets to any cluster head [24]. The reward $r_{t+1}^i a_i^i$ shows the link cost towards next-hop node [9, 11].

The basic elements of Markov decision process (MDP) are [S, T, A, R] where S represents set of states, T represents transition function, A represents set of actions, and R represents the reward function. The learning agent selects an action A with all states S which is shown in Figure 4. The

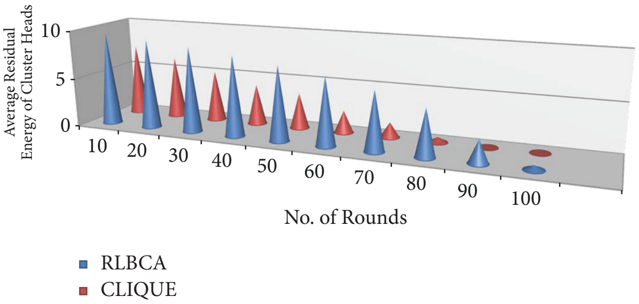


Figure 5: Average residual energy of cluster heads versus number of rounds.

Table 5: Simulation parameters.

Parameter	Value
Simulation Area	200 X 200 m^2
No. Of Sensor Nodes	500
Initial energy of Sensor Nodes	0.5 J
Communication range	50 m
Size of data packet	2500 bits
E_{elec}	50 nj / bit
Speed of mobile sink (V)	2,3,4,5,6 & 7 m/s
Deployment	Random
No of Episodes	4000
Learning parameter(alpha)	0.3, 0.5, 0.7 and 0.9

selected action later on computes the energy consumption for the cluster. Reward R derived from the calculated energy consumption to take proper decision. The formed decision increments current state S to Si+1 and next action A to Ai+1. The learning agent develops the optimal policy Q which increases the reward from learning experience to create optimal cluster heads [9, 11, 24, 37] which are shown in Figure 4.

In MDP, the state transition function P and reward function R are connected with current state and action. The main objective of learning agent is to develop a policy π: S ⟶ A. Learning agent has taken action Ai as per the current state Si, i.e., $\pi(S_i) = A_i$, so that the cumulative value function $V^{\pi}(S_i)$ derived from present/initial state Si worked as follows:

$$V^{\pi}(S_i) = r_i + y r_{i+1} + y^2 r_{i+2} + \cdots = r_i + y + V^{\pi}(S_{i+1}) = \sum_{i=1}^{\infty} y^i r_{i+1} \qquad (5)$$

From (5), r represents the return value and y represents the discount factor. The main objective of learning agent is to develop an intelligent strategy to make $V^{\pi}(S_i)$ highest [9, 11, 24, 37]. This strategy is known as policy and represented by

$$V^{\#} = \arg\max_{\pi} V^{\pi}(S_i) V_s \qquad (6)$$

Finally, to update Q-value the following equation is used:

$$Q_{t+1}(S_t, \alpha_t) = (1-\alpha) Q_t(S_t, \alpha_t) + \alpha[r^{t+1} + y \max Q_t(S_{t+1}, \alpha') - Q_t(S, \alpha_t) \qquad (7)$$

Constantly the Q-table is updated using (7). α and y are learning rate and discount factor. r_t is the return value, max $Q_t(S_{t+1}, \alpha')$ is the highest Q-value, and α' is action taken by learning agent [9, 11, 24, 37]. Based upon (5), (6), and (7), the proposed RLBCA for clustering works as shown in Algorithm 3.

Algorithm 3 is simulated in MATLAB as per the simulation parameters specified in Table 5. The results are compared with CLIQUE algorithm [12] on the basis of average residual energy of cluster heads against number of rounds which are shown in Figure 5. This comparison clearly showed that as the number of rounds increases the average residual energy of CLIQUE algorithm's cluster heads goes down but our proposed RBBCA algorithm performs very well.

5.2. On-Demand Mobile Sink Traversal (ODMST) Algorithm. Initially the cluster heads formed by RLBCA; now ODMST algorithm collected the data from cluster heads in Algorithm 4.

ODMST Algorithm 4 saved the energy of mobile sink by visiting only interested cluster heads. This algorithm also prolongs the lifetime of network. Each round of data transmission cycle is set as T. The method of calculating T is shown in

$$T = \sum_{i=0}^{n-1} \sqrt{(X_{i+1} - X_i)^2} + \frac{(y_{i+1} - y_i)^2}{V} \qquad (8)$$

where V is the moving speed of mobile sink. The average energy Ec is formulated as per the following equation [9, 11, 24, 37]:

$$E_c = \frac{\left[\sum_{i=0}^{n-1} E_i\right]}{n} \qquad (9)$$

where Ei is the residual energy of the node and n is the number of nodes in the cluster.

The return value of the agent node can be calculated as per the following equation [9, 11, 24, 37]:

$$V^{\pi}(j) = \frac{R^j}{E_{i,j}} = \frac{R^j}{\left(2KE_{elec} + K\boldsymbol{\varepsilon}_{amp} d^{\theta}\right) e^{hop(i.s)/H}} \qquad \boldsymbol{\theta} \in [1,2] \qquad (10)$$

Here, the return value of agent node keeps track of remaining residual energy as well as energy consumption. The highest value of Q always leads to optimal path. The mobile sink keeps the updated Q-value and selects the MS traversal path with maximum Q-value [37]. If any issue takes place with SNs then the second maximum Q-value is selected for MS traversal path. Generally this method saves the energy among SNs.

```
Step 1. Initially all sensor nodes sends hello message packet to show their residual energy and current position.
Step 2. The learning agent records the total number of neighbour nodes and their residual energy. Periodically the residual
    energy of each sensor nodes is set and return value of the node is set to zero.
Step 3. Based upon step 2, cluster head formation probability is computed. The base station selects the optimal number of
    cluster heads among the desired cluster heads and creates the list.
Step 4. The base station announces the list of eligible cluster heads.
Step 5. The newly formed cluster heads send advertisement packets to their nearest
Neighbours for communication purpose.
Step 6. The state-action Q-values [10] are updated by reward function (equation (*)) and Q-matrix (equation (**)) to
    achieve the optimal policy (equation (* * *)):
        Reward calculation
        $r^t_{i+1} \longleftarrow \mathrm{Avg}(rwd^t_{E,i+1}, rwd^t_{d,i+1})$   (*)
        Q-matrix updation
        $Q^t_{i+1} \longleftarrow Q^t_{i+1} + \alpha[r^t_{i+1} + y \max \mathrm{Q_t}(Q^t_{i+1} - Q^t_i)]$   (**)
        Optimal policy
        $[\mathrm{H_{opt}}, \mathrm{I}] \longleftarrow \max_{s^t_{i\in s}} \sum_{i=1}^{Es} \mathrm{r^t{}_i(S^t{}_i, a^t{}_i)}$   (* * *)
Step 7. if the current node's residual energy is greater than other neighbour's nodes, the sensor node with higher residual
    energy is elected as a cluster head for next subsequent round.
Step 8. Repeat step 1 to step 7.
```

Algorithm 3: Reinforcement learning based clustering algorithm (RLBCA).

```
Step 1. Obtain Cluster heads by executing Algorithm 3 (RLBCA).
Step 2. Initially mobile sink placed randomly. Mobile sink advertise his position to all cluster heads.
Step 3. Interested cluster heads sends their request to visit message packet to mobile sink.
Step 4. Mobile sink stores these received messages in routing table to calculate distance (as per equation (4)) and visit
    cluster head to collect the data.
Step 5. If multiple request messages are received by mobile sink then
    Step 5.1. Mobile sink calculates distance of SNs as per the equation (4) and store them in routing Table 1.
    Step 5.2. Mobile sink creates the traversal path as per shortest distance and execute it.
    Step 5.3. During this execution of mobile sink traversal, if again any cluster heads send their request
        message then mobile sink used to calculate the shortest distance, update the traversal path and execute it.
Step 6. Go to step 3.
```

Algorithm 4: On-demand mobile sink traversal [ODMST].

6. Performance Evaluations

To properly check the performance of our proposed RLBCA and ODMST algorithm, we have compared these algorithms with other algorithms, namely, TTDD [13], DBRkM [3], EPMS [5], RLLO [15], and RL-CRC [16].

Here the network environment contains 500 sensor nodes in the area of 200 X 200 m^2 area. The initial energy of all sensor nodes is 0.5 J. The extensive simulation has taken place in MATLAB 2012 (A) [38] based upon simulation parameters specified in Table 5.

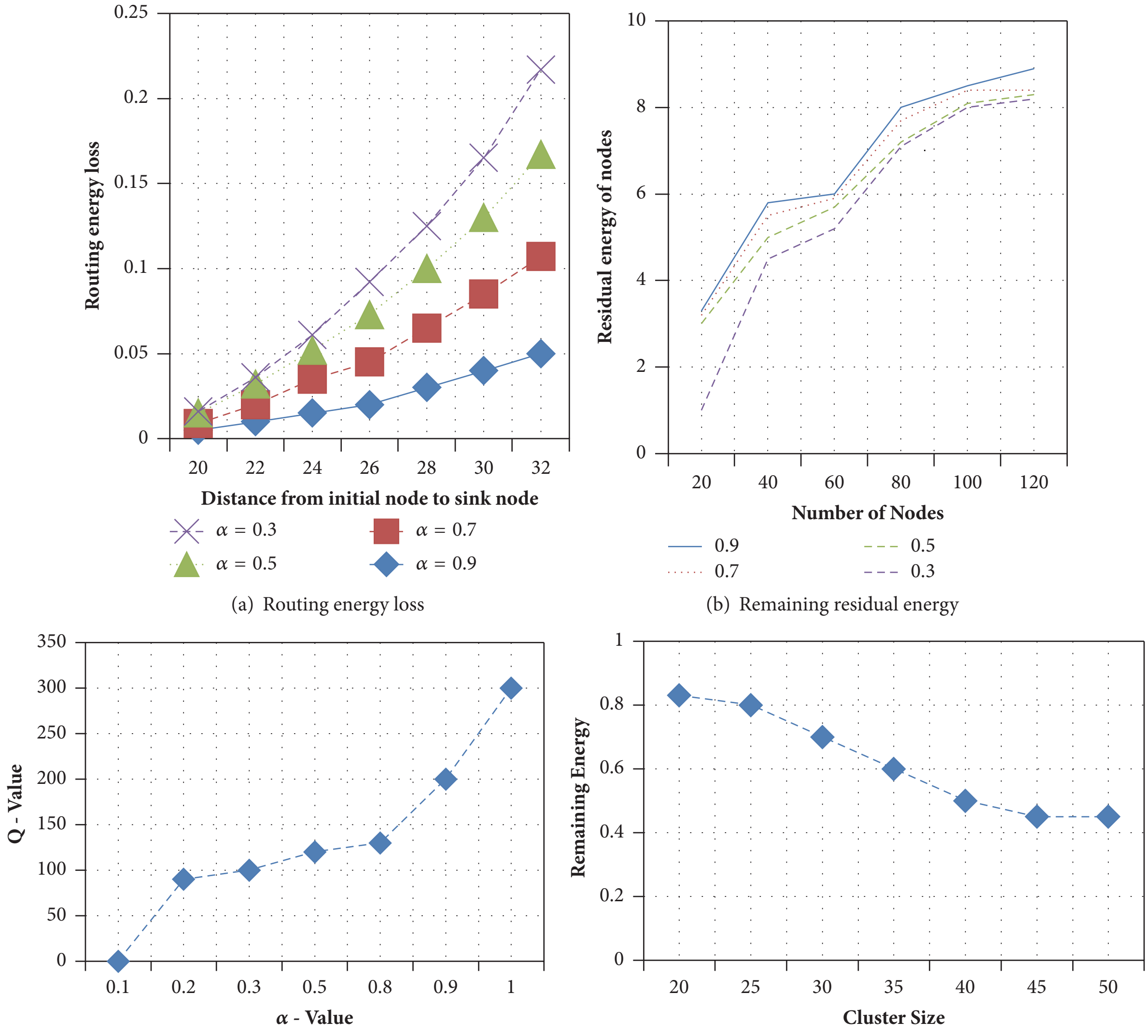
6.1. Result and Discussion. In this section of our research study, the performance of our proposed RLBCA and ODMST algorithm based upon simulation parameters specified in Table 5 using MATLAB is evaluated. Energy consumption and network lifetime are the main performance criteria for our research study. Therefore, it is mandatory to ensure less energy consumption for every cluster heads and mobile sink. The simulation results are compared with other algorithms like TTDD [13], DBRkM [3], EPMS [5], RLLO [15], and RL-CRC [16] on the basis of the following performance metrics:

(i) Routing energy loss: Every cluster head in the WSN consumes certain amount of energy. The cluster heads which are not involved in the routing path of mobile sink are said to be idle, to save their energy.

(ii) Network lifetime: It includes the duration from the starting of WSNs operation until the death of first sensor node.

(iii) Learning time: The sensor nodes learn through learning agent. As the learning (alpha value) increases, the performance of WSN operation also increases along with decrement in routing energy loss.

(iv) Convergence of algorithm: This is the performance of algorithm which is expressed by two ways: rate and order of convergence.



(a) Routing energy loss

(b) Remaining residual energy

(c) Alpha value impact

(d) Remaining energy versus cluster size

Figure 6

(v) Sum of square error (SSE): This presents standard way of analysis within a cluster. The SSE shows the performance of cluster heads as per the energy optimization.

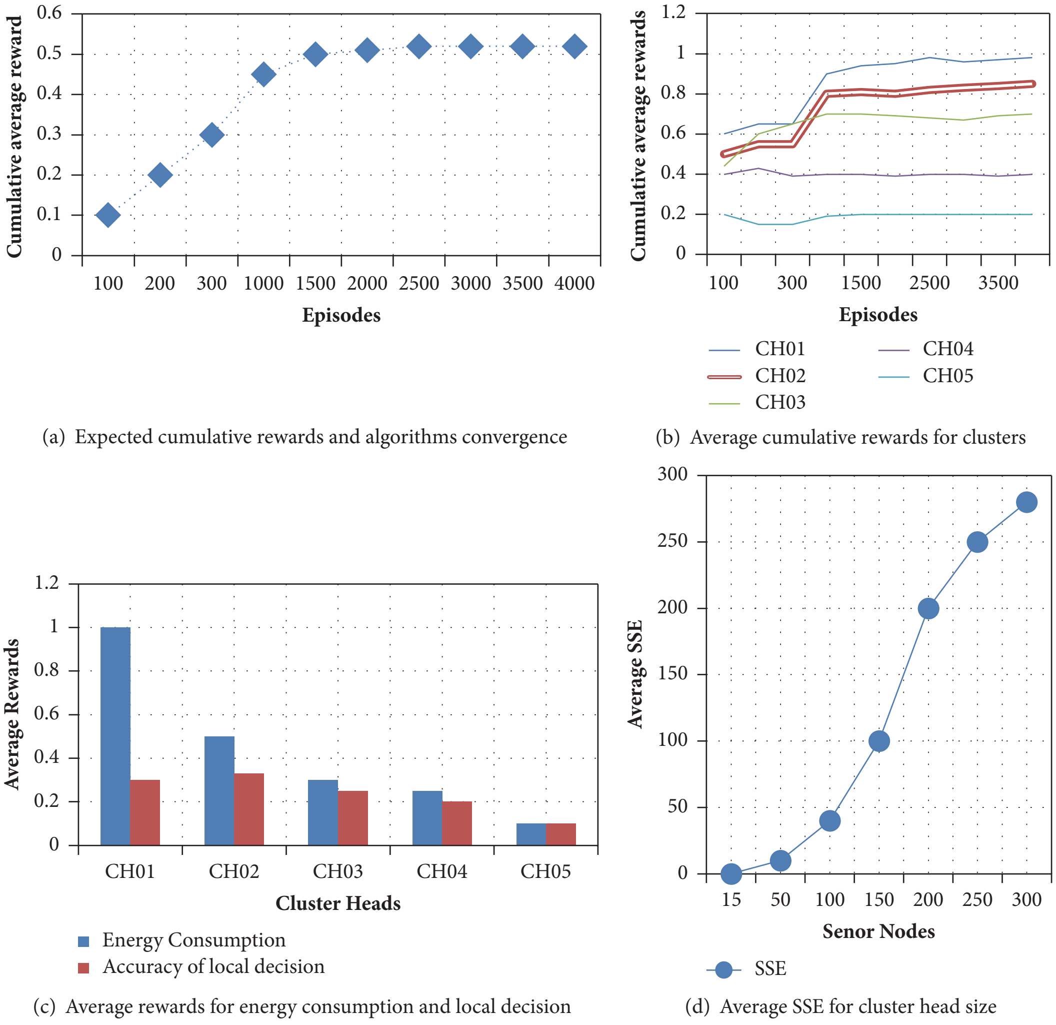
(vi) PDR ratio: This is the difference between generated packets and received packets. Data loss comes under this ratio which is a very important factor in WSN.

(vii) Average end to end delay: Here, average end to end delay is calculated as the duration of time taken by data packets to reach the mobile sink from cluster head.

(viii) Average node degree: This is the number of edges connected to each SN. In WSN, cluster head forms dense network to represent average node degree.

After extensive simulation in MATLAB, we observed that as alpha value (Figure 6(a)) (learning parameter) increased, the routing loss decreased as the distance from initial node to sink node increases. Figure 6(b) showed that as alpha value (learning parameter) is incremented, the residual energy of sensor nodes is also increased. Figure 6(c) showed the comparison between alpha value and q-value which clarify that as the learning parameter (alpha value) increased, the q-value also incremented. Figure 6(d) showed that as cluster size increased the remaining energy of sensor nodes decreased. Finally, we can see that the RL based schemes learned energy dissipation for every cluster by exploration of the clusters to find the perfect cluster.

The RLBCA founds the optimal solution just after state action pair's exploration. This mainly depends upon learning rate, discount factor, and action selection policy. We simulated proposed RLBCA to test the convergence of our algorithm over 4000 episodes and evaluated its performance as presented in Figure 7. The simulation result showed that RLBCA (part of RL) founds the optimized solution only



(a) Expected cumulative rewards and algorithms convergence

(b) Average cumulative rewards for clusters

(c) Average rewards for energy consumption and local decision

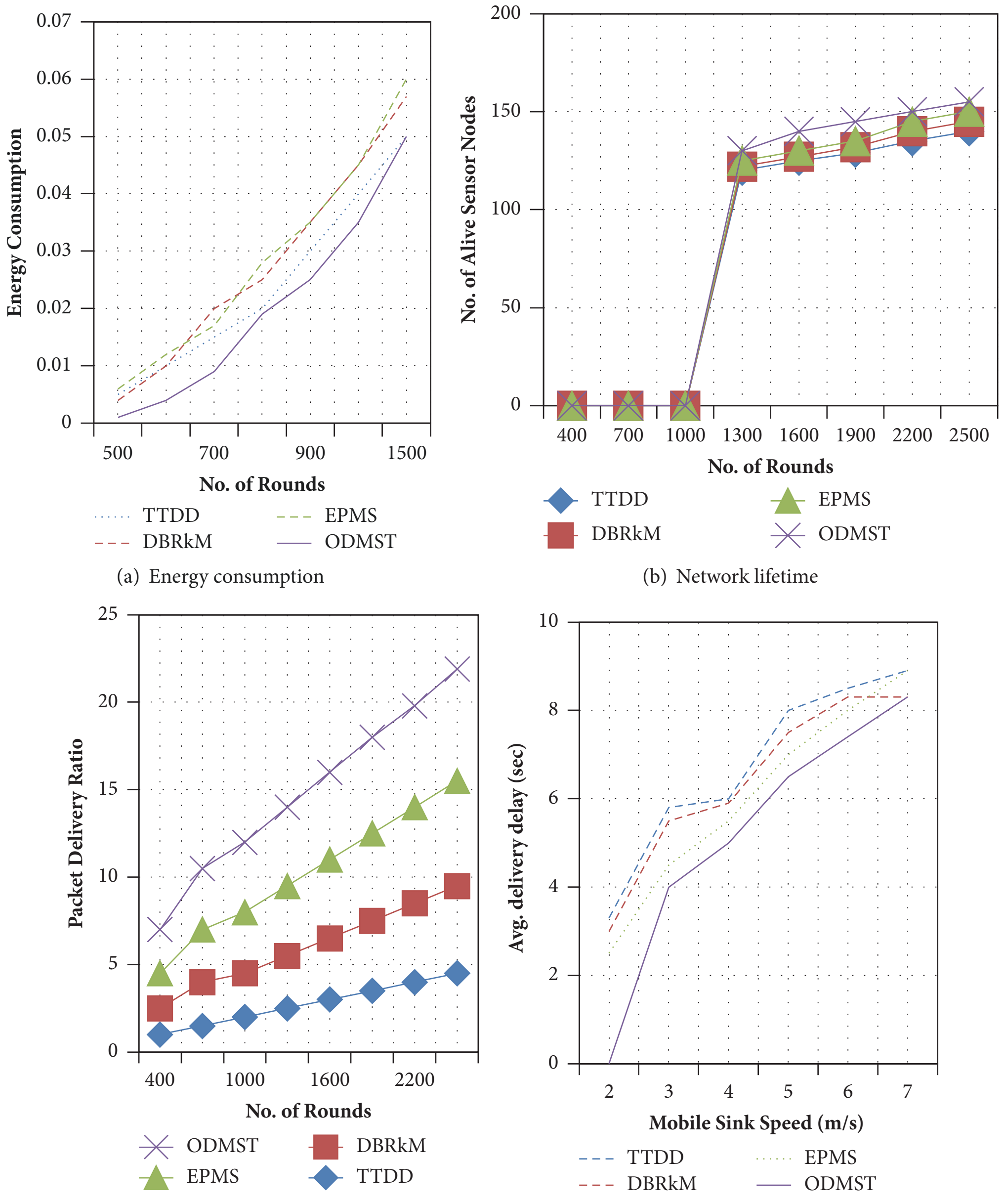
(d) Average SSE for cluster head size

Figure 7

after certain number of episodes. We embedded Q-learning for clustering due to its faster convergence and shorter learning period. Figures 7(a) and 7(b) show cumulative average rewards for five cluster heads CH01, CH02, CH03, CH04, and CH05. This represents that learning agent adapts to environment through neighbour cluster heads. Figure 7(c) showed the performance during learning period and selected the optimal cluster head based upon energy dissipation and accuracy of local decision. Figure 7(d) showed the sum of square error (SSE) for the entire network which is a key component to determine the performance of cluster heads based upon energy optimization.

As per the Figure 8(a), as the number of rounds increased, the energy consumption of ODMST algorithm is comparably lower than other algorithms like TTDD [13], DBRkM [3], and EPMS [5]. This is due to the on-demand mobile sink traversal whereas in TTDD algorithm [13] every data source established a virtual grid network which has consumed more energy. Figure 8(b) showed the network lifetime where in TTDD algorithm [13] first node died after 200 rounds while it is about 400 rounds for DBRkM [3] and 1500 rounds for EPMS algorithm [5] but in our ODMST algorithm, first node died after 2000 rounds which is comparably better than other algorithms. ODMST algorithm worked very well up to 2500 rounds. Packet delivery ratio is shown in Figure 8(c) which clearly justifies that ODMST algorithm provided much better PDR ratio up to 2500 rounds other than the TTDD [13], DBRkM [3], and EPMS [5] algorithms due to better communication link and fewer burdens on mobile sink. The mobile sink speed is shown in Figure 8(d) which reflects the average end to end delay of packets; here ODMST algorithm performs better than other state-of-the-art algorithms because our mobile sink did not suffer from flooding of data packets.

Figure 9 shows that the average node degree is maximum whenever the numbers of sensor nodes are increased from 100 to 500. Simulation result showed that ODMST and RLBCA are able to achieve node degree in even harsh type of node deployment in the WSN. Finally Table 6 shows the overall performance improvement of our proposed algorithms.



(a) Energy consumption

(b) Network lifetime

(c) PDR ratio

(d) Average end to end delay

Figure 8

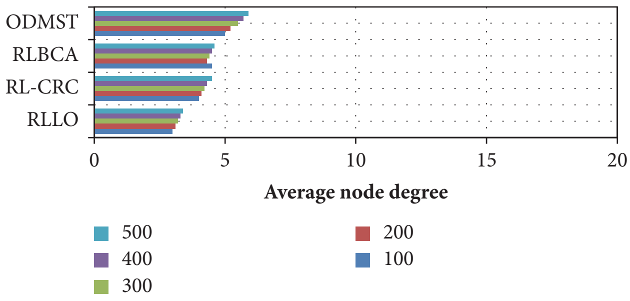


Figure 9: Average node degree.

Table 6: Performance improvement.

Proposed algorithms	% of improvements over existing state of the art algorithms	
	Energy consumption	Network lifetime
RLBCA	36.3%	38.25%
ODMST	39.8%	39%

7. Conclusions and Future Scope

This research study has proposed two novel learning algorithms to properly overcome hot spot problem in WSN by using RL techniques. The main idea is to collect the data from cluster heads as per their demand/request to mainly save the energy consumption of mobile sink and to improve the network lifetime. Simulation results showed that RLBCA created cluster head properly and ODMST algorithm collected the data from cluster heads through mobile sink efficiently compared with the state-of-the-art algorithms. This research study motivated us to further test the scalability and convergence of the proposed algorithms in large scale of WSN.

Abbreviations

WSN:	Wireless sensor network
SNs:	Sensor nodes
MS:	Mobile sink
MDP:	Markov decision process
RLBCA:	Reinforcement learning based cluster algorithm
ODMST:	On-demand mobile sink traversal
CH:	Cluster head
RL:	Reinforcement learning
SSE:	Sum of squares error.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

This research study has no conflicts of interest.

Authors' Contributions

The first author has performed the literature review, simulation, result analysis, and paper writing. The second author has supervised this research study.

References

[1] M. Zareei, A. K. M. M. Islam, C. Vargas-Rosales, N. Mansoor, S. Goudarzi, and M. H. Rehmani, "Mobility-aware medium access control protocols for wireless sensor networks: A survey," *Journal of Network and Computer Applications*, vol. 104, pp. 21–37, 2018.

[2] M. Azharuddin and P. K. Jana, "Particle swarm optimization for maximizing lifetime of wireless sensor networks," *Computers and Electrical Engineering*, vol. 51, pp. 26–42, 2016.

[3] A. Kaswan, K. Nitesh, and P. K. Jana, "Energy efficient path selection for mobile sink and data gathering in wireless sensor networks," *AEÜ - International Journal of Electronics and Communications*, vol. 73, pp. 110–118, 2017.

[4] M. Elshrkawey, S. M. Elsherif, and M. Elsayed Wahed, "An Enhancement Approach for Reducing the Energy Consumption in Wireless Sensor Networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 2, pp. 259–267, 2018.

[5] J. Wang, Y. Cao, B. Li, H.-J. Kim, and S. Lee, "Particle swarm optimization based clustering algorithm with mobile sink for WSNs," *Future Generation Computer Systems*, vol. 76, pp. 452–457, 2017.

[6] M. I. Khan, W. N. Gansterer, and G. Haring, "Static vs. mobile sink: The influence of basic parameters on energy efficiency in wireless sensor networks," *Computer Communications*, vol. 36, no. 9, pp. 965–978, 2013.

[7] Y. Gu, Y. S. Ji, J. Li, F. Ren, and B. Zhao, "EMS: efficient mobile sink scheduling in wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 5, pp. 1556–1570, 2013.

[8] J. Tang, H. Huang, S. Guo, and Y. Yang, "Dellat: Delivery Latency Minimization in Wireless Sensor Networks with Mobile Sink," *Journal of Parallel and Distributed Computing*, vol. 83, pp. 133–142, 2015.

[9] P. Montague, "Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G.," *Trends in Cognitive Sciences*, vol. 3, no. 9, p. 360, 1999.

[10] F. Hu and Q. Hao, *Intelligent Sensor Networks*, CRC Press, 2012.

[11] J. Randlov and P. Alstrom, "Learning to drive a bicycle using reinforcement learning and shaping," in *Proceeding of the fifteenth international conference on machine learning, ACM DL*, pp. 463–471, 1998.

[12] A. Förster and A. L. Murphy, "CLIQUE: Role-free clustering with Q-learning for wireless sensor networks," in *Proceedings of the 2009 29th IEEE International Conference on Distributed Computing Systems Workshops, ICDCS, 09*, pp. 441–449, Canada, June 2009.

[13] H. Luo, F. Ye, J. Cheng, S. Lu, and L. Zhang, "TTDD: two-tier data dissemination in large-scale wireless sensor networks," *Wireless Networks*, vol. 11, no. 1-2, pp. 161–175, 2005.

[14] G. Chen, J.-S. Cheuh, M.-T. Sun, T.-C. Chiang, and A. A.-K. Jeng, "Energy-efficient mobile targets detection in the presence of mobile sinks," *Computer Communications*, vol. 78, pp. 97–114, 2016.

[15] W. J. Guo, C. R. Yan, Y. L. Gan, and T. Lu, "An intelligent routing algorithm in wireless sensor networks based on reinforcement learning," *Applied Mechanics and Materials*, vol. 678, pp. 487–493, 2014.

[16] T. T. T. Le and S. Moh, "Reinforcement-Learning-Based Topology Control for Wireless Sensor Networks," in *Proceedings of the Grid and Distributed Computing 2016*, pp. 22–27.

[17] M. Akbar, N. Javaid, W. Abdul et al., "Balanced Transmissions Based Trajectories of Mobile Sink in Homogeneous Wireless Sensor Networks," *Journal of Sensors*, vol. 2017, 2017.

[18] N. Ghosh and I. Banerjee, "An energy-efficient path determination strategy for mobile data collectors in wireless sensor network," *Computers and Electrical Engineering*, vol. 48, pp. 417–435, 2015.

[19] A. Wichmann and T. Korkmaz, "Smooth path construction and adjustment for multiple mobile sinks in wireless sensor networks," *Computer Communications*, vol. 72, pp. 93–106, 2015.

[20] H. Ahmadi, F. Viani, and R. Bouallegue, "An accurate prediction method for moving target localization and tracking in wireless sensor networks," *Ad Hoc Networks*, vol. 70, pp. 14–22, 2018.

[21] M. Asif, S. Khan, R. Ahmad, M. Sohail, and D. Singh, "Quality of service of routing protocols in wireless sensor networks: A review," *IEEE Access*, vol. 5, pp. 1846–1871, 2017.

[22] G. Dangelo, D. Diodati, A. Navarra, and C. M. Pinotti, "The Minimum κ-Storage Problem: Complexity, Approximation, and Experimental Analysis," *IEEE Transactions on Mobile Computing*, vol. 15, no. 7, pp. 1797–1811, 2016.

[23] W. Heinzelman B, *Application Specific c Protocol Architectures for Wireless Networks [Ph.D. thesis]*, 2000.

[24] K. A. Yau, H. G. Goh, D. Chieng, and K. H. Kwong, "Application of reinforcement learning to wireless sensor networks: models and algorithms," *Computing: Archives for Scientific Computing*, vol. 97, no. 11, pp. 1045–1075, 2015.

[25] W. Wen, S. Zhao, C. Shang, and C.-Y. Chang, "EAPC: Energy-Aware Path Construction for Data Collection using Mobile Sink in Wireless Sensor Networks," *IEEE Sensors Journal*, 2017.

[26] H. Huang, A. V. Savkin, and C. Huang, "I-UMDPC: The Improved-Unusual Message Delivery Path Construction for Wireless Sensor Networks with Mobile Sinks," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1528–1536, 2017.

[27] I. Mustapha, B. M. Ali, A. Sali, M. F. A. Rasid, and H. Mohamad, "An energy efficient Reinforcement Learning based Cooperative Channel Sensing for Cognitive Radio Sensor Networks," *Pervasive and Mobile Computing*, vol. 35, pp. 165–184, 2017.

[28] W. Osamy and A. M. khedr, "An algorithm for enhancing coverage and network lifetime in Cluster Based wireless sensor networks," *International Journal of Communication Networks and Information Security*, vol. 10, pp. 1–9, 2018.

[29] M. Zayoud, H. M. Abdulsalam, A. Al. Yatama, and S. Kadry, "Split and merge leach based Routing algorithm for wireless sensor networks," *International Journal of Communication Networks and Information Security*, vol. 10, pp. 155–162, 2018.

[30] P. Zhong and F. Ruan, "An energy efficient multiple mobile sink based routing algorithm for Wireless sensor networks," *ICFMCE*, vol. 323, pp. 01–04, 2018.

[31] S. Guo, X. Wang, and Y. Yang, "Topology control for maximizing network lifetime in wireless sensor networks with mobile sink," in *Proceedings of the 10th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, MASS 2013*, pp. 240–248, China, October 2013.

[32] H. Zhang, X. Wang, P. Memarmoshrefi, and D. Hogrefe, "A Survey of Ant Colony Optimization Based Routing Protocols for Mobile Ad Hoc Networks," *IEEE Access*, vol. 5, pp. 24139–24161, 2017.

[33] C. Wu, Y. Liu, F. Wu, W. Fan, and B. Tang, "Graph-Based Data Gathering Scheme in WSNs with a Mobility-Constrained Mobile Sink," *IEEE Access*, vol. 5, pp. 19463–19477, 2017.

[34] Y. Gu, F. Ren, Y. Ji, and J. Li, "The evolution of sink mobility management in wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 507–524, 2016.

[35] S. Yu, B. Zhang, C. Li, and H. T. Mouftah, "Routing protocols for wireless sensor networks with mobile sinks: a survey," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 150–157, 2014.

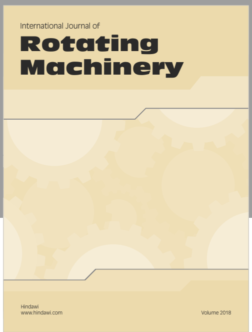
[36] I. Mustapha, B. M. Ali, M. F. A. Rasid, A. Sali, and H. Mohamad, "An energy-efficient spectrum-aware reinforcement learning-based clustering algorithm for cognitive radio sensor networks," *Sensors*, vol. 15, no. 8, pp. 19783–19818, 2015.

[37] M. Wiering and M. van Otterlo, "Adaptation learning and optimization," *Reinforcement Learning state of the art, Book*, vol. 12, pp. 3–326, 2012, ISSN 1867-4534.

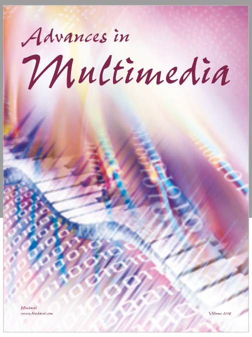
[38] https://in.mathworks.com/solutions/deep-learning.html?s_tid=srchtitle.

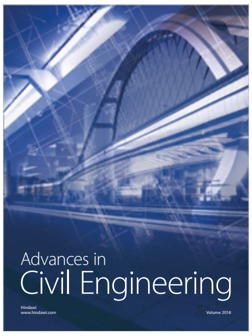
Journal of
Engineering

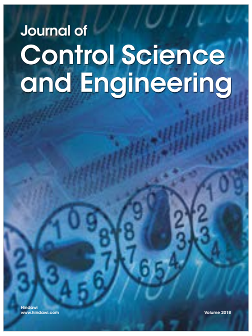
The Scientific
World Journal


International Journal of
Rotating
Machinery

Journal of
Sensors

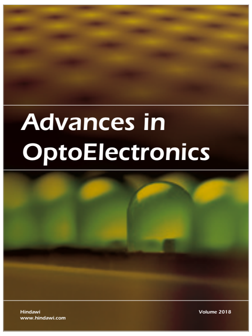

Advances in
Multimedia


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

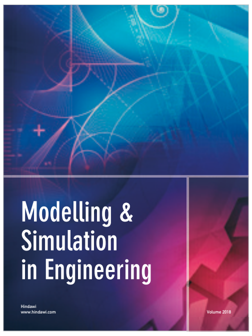
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

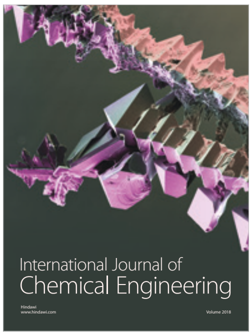

Advances in
OptoElectronics

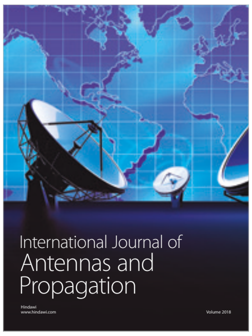
VLSI Design

International Journal of
Navigation and
Observation

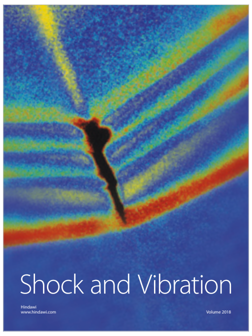

Modelling &
Simulation
in Engineering

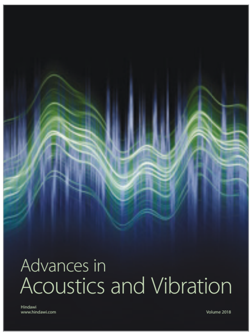
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration