International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

COVERAGE OPTIMIZATION IN UAV-ASSISTED DRL BASED WIRELESS SENSOR NETWORK FOR DISASTER MANAGEMENT

Karunakar Kothapalli[1], Nesarani Abraham[2*]

1 & 2* Research Scholar, Department of Computer Science and Engineering

1 & 2* Associate Professor , Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

1karunakar524@gmail.com

2*nesaraniabraham84@gmail.com

Abstract

Unmanned aerial vehicle integrated with wireless sensor network is efficient in managing the sensor node connectivity as well as data transmission during the existence of disaster, the maximum area coverage and data transmission are ensured especially when nodes are grouped into **clusters**. Wireless Sensor Networks (WSNs) play a vital role in large-scale environmental monitoring, surveillance, and disaster management. However, the inherent energy constraints and limited communication range of sensor nodes pose significant challenges in ensuring efficient coverage and data collection. These limitations are overcome by using this novel proposal, this paper proposes a novel framework integrating Unmanned Aerial Vehicles (UAVs) with Deep Reinforcement Learning (DRL) for coverage optimization in clustered WSNs, each managed by a Cluster Head (CH), which aggregates data and communicates with UAVs. A DRL agent is trained to dynamically plan UAV trajectories and scheduling policies that maximize coverage and data collection efficiency while minimizing energy consumption and communication delays. The state space captures real-time UAV positions, CH locations, and node energy levels, while the reward function encourages full cluster coverage, reduced redundant visits, and balanced load distribution. Simulation results show that the proposed DRL-based UAV-assisted strategy significantly improves sensing coverage, reduces data latency, and extends network lifetime compared to traditional approaches.

Introduction

The area of wireless sensor networks has seen a growth in interest in recent years. There are many wireless sensor nodes in a wireless sensor network. These nodes are distinguished by their extremely tiny size and low energy output, which is typically provided by a battery. They use integrated antennas to communicate over RF signals. These networks are often employed to keep an eye on a field of interest and identify movement, temperature fluctuations, rainfall, and other events. The topic of coverage is one of the most researched areas in wireless sensor networks. Generally, coverage is understood as how well a sensor network would monitor a field of interest. Maintaining connection is essential for a sensor network, as is coverage. The term "connectivity" refers to the capacity of the sensor nodes to communicate with the data sink. The data gathered by a sensor node

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

cannot be processed if there is no route from the sensor node to the data sink. The communication range of each node determines the region where another node may be to receive data. The area that a node can observe is determined by its sensing range, which is distinct from this. Although the two ranges can be the same, they are frequently distinct.

Wireless sensor networks (WSN) have been useful in applications ranging from environmental monitoring to industrial automation, particularly in the field of military surveillance and disaster management. Unmanned aerial vehicle (UAV) are increasingly accessed and act as an aerial base station to extend network coverage and improve the efficiency of data transfer and packet relay for ground WSNs. Deep reinforcement learning can optimize the use of resources within the WSN, which includes energy management and cluster head selection and transmission. The quality of links depends on the reliability between the nodes and UAV, channel model, distance, pathloss and signal quality. The WSN's node sensing data aggregation is a challenging task and inaccurate during the time of natural disaster. UAV-assisted WSN has gained attention due to its channel modelling, the links such as ground-to-air and air-to-ground are blocked by the obstacles and have lack of line-of-sight signals to transfer the real-time data, theoretical models are not sufficient to identify the nature of the links. The authors in [4] evaluated the two-ray

model and propagation characteristics of the channels. The authors have utilized the benefits of UAV and made optimum cluster head (CH) selection to improve the lifetime of sensor nodes. As defined by ITU standard, empirical pathloss are adopted in UAV enabled wireless network. Artificial intelligent studies are widely used to develop pathloss models which are more accurate than the two-ray pathloss model. However, the channel characteristics are varied under drone trajectories. There is a clutter loss is inversely proportional to the altitude of the UAV. The channel model of is more specific and not limited in heights or elevation angle. The movement is assumed to be in the 3-D space and specifically the improper orientation impact radio irregularity. In WSN, cluster selection (CH) is a problem concerned with factors like distance from centroids, residual energy and pathloss, Researchers used various algorithm such as LEACH protocol, PSO algorithm and natural inspired algorithms to involve in the CH selection process. The selection is based on the clustering among the nodes, intra-cluster distance and energy consumption. The authors proposed energy aware clustering techniques to choose CH selection. There are three cases when energy consumption is considered.

Each device of the WSN and the CH selection is responsible for collecting data. The trajectory planning has multiple solutions using machine learning methods, heuristic algorithms applied to design. The authors propose a deep reinforcement learning (DRL) based method in solving the trajectory planning, deep deterministic policy gradient is considered to maximize the energy efficiency jointly considering the coverage and connectivity. UAV's data transmission depends on the flying speed, the authors in [9] formulated a problem as Markov decision process to optimize the energy consumption. Deep neural network is adopted and the actor-critic based algorithm is used to ensure the 3-D trajectory control and user allocation to give connectivity fairness of the ground users. The realization of enabling internet of things (IoT), the ability to handle complex wireless environment is handled by the application of convolutional neural network (CNN) along with deep Q network for dynamic decision optimization. The concept of natural language programming (NLP) is used to design the sequence learning for the path problem in clustered wireless environment. The main advantage of incorporating UAV into WSN are powerful computing and sensing to provide services compared to terrestrial network, UAV has been chosen for some key points, 1) During emergency service, Quick moving UAVs are capable of providing network services in short time, especially during disaster management, 2) For better communication in the midst of obstacles and low signal communication links toward LOS channels, 3) UAVs deployment and management are not expensive.

UAV mounted base stations (BS) are deployed at locations in difficult area where natural disasters occurred, due to the constrained coverage, UAV is good to support communication services with quality of service (QoS) features such as energy conservation, coverage radius and prolonged network lifetime by scheduling and energy harvesting. However, in remote areas, ground BS are not accessible due to reasons, during the insufficient data transmission, coverage radii is adjusted to reduce the battery power and tries to retain the wireless service, sleep scheduling is followed to remove the redundancy of data and consume necessary energy. In DRL, centralized and distributed algorithms are performed to deploy the UAV assisted WSN for computation as well as coverage, but they do suffer with large dimensionality due to high computation and dynamic environment, the global optimization is poor due to the constrained communication range and low coverage, this makes the UAV to act difficult in obtaining the entire data without prior knowledge, In such difficult situation, distributed approach is way behind to achieve a desirable global solution.

The cooperation of each UAVs is improved to achieve a goal as a local optimization and to receive the information from one-hop neighbours to improve the global communication coverage. Some swarm intelligence algorithms are utilized in centralized way. The routing efficiency is enhanced by means of using Q-learning based multi-hop routing methodology, the scalability and dimensionality are the challenges leading to high number of routing overhead and the instability. Q-learning based topology aware routing enhances packet delivery ratio and jitter considering optimal routing. The application of Q-learning is expanded by leveraging neighbour hop information and data transmission alongside changing network conditions. Some researchers used 3-D Q-learning routing combined with greedy and store forward technique to stabilize the spectrum allocation and to improve the network performances. A sector-based clustering is considered the practical wireless energy transfer application to increase the energy head (EH) level. They transmit data to the CH to improve the radio coverage that maximize the transmission rate of the data, the optimal coordinate of the CH is investigated entails a communication overhead and to reduce the dimensionality space.

A DRL based WSN is used to address the communication coverage problem while maintaining energy consumption and data transmission. The summary of the proposed work is as follows,

1) An UAV-assisted wireless sensor network system is developed to provide network life in communication services for users through UAV in the energy-constrained environments such as disaster emergency or environmental monitoring.

2) An UAV and sensor nodes are deployed in a random fashion to show that the system is realistic, where UAV dynamically interacts with CH and base stations via downlink model. The comprehensive design is incorporating critical network parameters including UAV location, outage, flight distance and

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

channel quality.

3) A Deep deterministic Policy Gradient (DDPG)cement Learning (DRL) algorithm is applied for coverage optimization in clustered WSNs. The continuous-action reinforcement learning approach is trained the system to ensure control over UAV trajectory and network coverage. An experience replay buffer to store interaction sample which is used to train actor critic network effectively.

2. System Model

As described in Fig. 1, we deploy a UAV-assisted wireless network with UAVs hovering to provide communication coverage in a region of interest. Let $U = \{u_1, u_2, \dots, u_N\}$ be the set of UAVs on board for coverage, $S =$

$\{s_1, s_2, \dots, S_K\}$ denote the set of ground sensor nodes, The initial position of UAV at time t is denoted as $p_i(t) = [x_i(t), y_i(t), h_i(t)]$ and the position of the sensor node is denoted as $s_k = [x_k, y_k]$. The selected CH node as k where $k \in K$ and the remaining nodes are sensor node denoted as j where $j \in K \setminus \{k\}$. The trajectory of the UAV represented by their hovering position (t) corresponding to $p_i(t)$, and the K target hovering positions $C =$

$\{c_1, c_2, \dots, C_K\}$, which is at fixed altitude for simplicity, the trajectory path is followed as a permutation of its hovering position.

2.1 Channel Model

UAV communications between ground nodes are developed using the pathloss model considering LOS probability during hovering. The LOS probability depends on the dynamic environment and elevation angle, is described as

$$P_{LOS} = \frac{1}{1+\eta\exp(-[\tau-\eta)} \quad (1)$$

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

where η and β are the environmental constants and the elevation angle is denoted as

$$\tau = \frac{180}{\pi} \times \sin^{-1}\left(\frac{}{d_{ik}(t)}\right) \quad (2)$$

The flying position is directly from the ground CH to air UAV. H is the altitude of UAV from ground, for simplicity, it is assumed to be constant, $d(t)$ is the distance between the UAV and CH and is expressed as

$$d(t) = \sqrt{(x_i(t) - x_k)^2 + (y_i(t) - y_k)^2 + h_i(t)^2} \quad (3)$$

Obviously the NLOS probability is computed as

$$P_{NLOS} = 1 - P_{LOS} \quad (4)$$

The average pathloss is developed as

$$PL = P_L(K_0 + \mu_{LOS}) + P_{NLOS}(K_0 + \mu_{NLOS}) \quad (5)$$

where μ_{LOS} and μ_{NLOS} are the link mean values, $K_0 = 10\alpha \log_{10}\left(\frac{4\pi f_c H}{c}\right)$, α is the pathloss exponent, c is the speed of light and f_c is the carrier frequency. The average data rate is expressed as

$$r^t_{data} = B \times \log_2\left(1 + \frac{P^t_{CH}}{PL \times N_0}\right) \quad (6)$$

where B is the available bandwidth and N_0 is the noise power spectral density and P^t_{CH} is the CH transmitted power.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

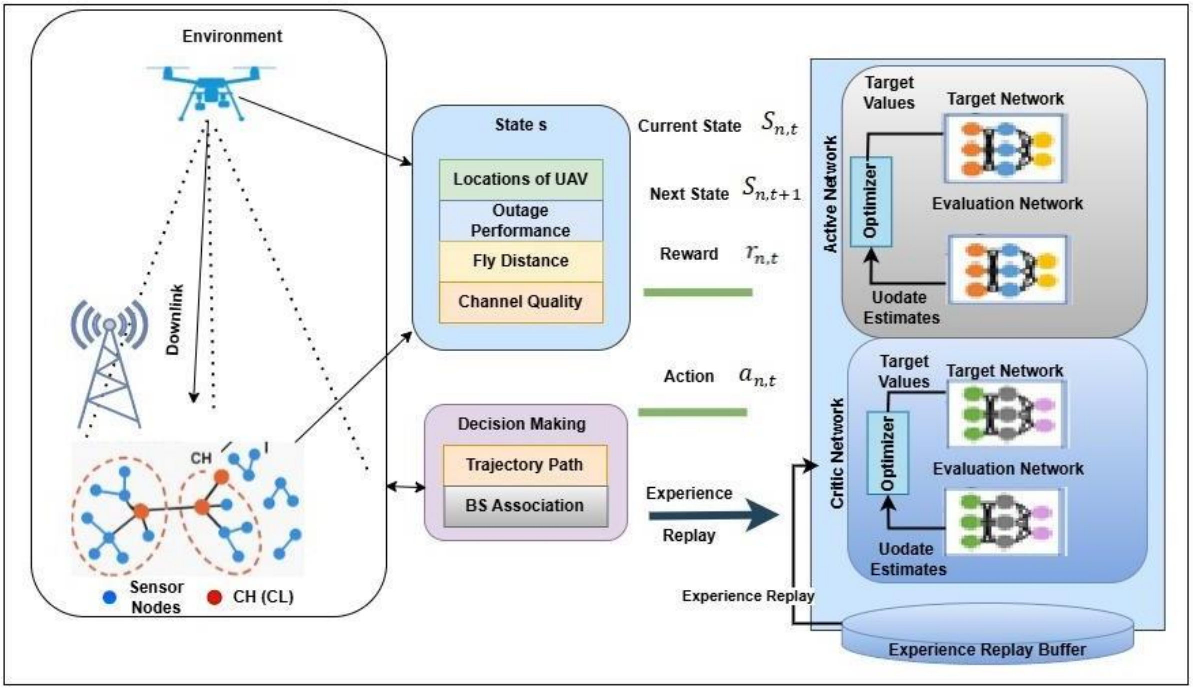

Figure 1 System model

The system model shown in figure 1 comprises of UAVs and cluster of nodes with sensor nodes and CH, the data transmission, coverage and trajectory path information are fed into deep reinforcement learning model, where deep deterministic policy gradient (DDPG) method, DDPG uses target networks to stabilize training by slowly tracking the main (evaluation) networks. The target values for training the critic are computed using the target actor and critic, specifically designed for continuous action space which fits into the coverage and trajectory optimization problem.

2.2 UAV's Energy and Trajectory Model

The hovering of UAV is not possible with acceleration and velocity of the UAV mode, UAV flies with a speed

v_{UAV}, the data collection from the sensor nodes are based on time division multiple access (TDMA) and forwards the data to the UAV, There are communication related energy and hover-related energy. P_{hover}

The power during hovering is expressed as $= \sqrt{\frac{(m_{UAV}g)^3}{2\pi r^2 n_p \rho}}$

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(7)

where g is the gravity, ρ is the air density and (m_{UAV}) is the mass of UAV. The energy consumption during hovering is expressed as

$$E_c = (P_{hover} + P_{com}) \quad (8)$$

$$= \frac{D_k}{r_{data}} (P_{hover} + P_{com}) \quad (9)$$

T_k is the hovering time of an UAV and D_k is the data required to be transferred from CH to UAV. The movement of UAV depends on the

$$P_{move} = \frac{P_{max} - P_{idle}}{v_{max}} v_U + P_{idle} \quad (10)$$

where v_{max} is the maximum speed of the UAV, P_{max} and P_{idle} are the hardware power levels of UAV when it is in maximum speed and being idle.

The velocity is the time derivative of its position vector

$$\vec{v}_U(t) = \frac{d\vec{r}(t)}{dt} \quad (11)$$

The position of UAV in 3D space is

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k} \quad (12)$$

Then its velocity vector is

$$(t) = k_1 v_U(t-1) + (1-k_1)\bar{v}_U + \sqrt{1-k^2}\,\Phi_i \quad (13)$$

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$(t) = k_2\theta_U(t-1) + (1-k_2)\bar{\theta_U} + \sqrt{1-k^2}\ \Phi_i$ (14)

2

Where $0 \leq k_1,\ k_2 \leq 1$ are the adjusted state of UAVs with average velocity $\bar{v_U}$ and average direction $\bar{\theta_U}$ of all UAVs. Φ_i are the Gaussian distributions.

2.3 Coverage Model

As shown in figure, the channel power gain between sensor node and UAV, the energy harvested by sensor node and is expressed as

$E_i = \zeta G_{ik} h_i P_{move} T_e$ (15)

$= \frac{\eta E_i}{T_d} = \frac{\eta \zeta G_{ik} h_j P T}{T_d}\ , \forall j \in \mathcal{N} \setminus \{i\}$ (16)

ik

$SIR_{ik} = \frac{G_{ik} P_{ik}\ (d_{ik}^{-\alpha})}{\sum (d^{-\alpha}) + \sigma^2}$ (17)

P

$j \neq i\ \ j \qquad jk$

The coverage is based on the SIR_{ik} if the signal quality is greater than the threshold value.

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$C_{ik} = \begin{cases} 1 & \text{if } SIR_{ik} \geq \gamma_{ik} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

Where $0 < \eta < 1$ is the ratio of the energy used only for transmission, to the total energy harvested from the base station, T_e is the time duration of energy harvesting T_d is the time duration of data transmission and based on the Shannon capacity, the achievable rate is expressed as

$$R_{ik} = log_2(1 + SIR_{ik})\ \forall i \in \mathcal{N} \setminus \{k\} \quad \text{[b/s/Hz]} \tag{19}$$

$$\mathscr{R} = min\left[\min_{\forall i \in \mathcal{N} \setminus \{k\}} \{R_{ik}\},\ R_i\right] \tag{20}$$

The optimal CH maximizes the sensing data rate followed by the problem objective as,

$$(P1): \quad \arg\max_{\forall i \in \mathcal{N}} \mathcal{R} = \arg\max_{\forall i \in \mathcal{N}} m\left[\min_{\forall i \in \mathcal{N} \setminus \{k\}} \{R_{ik}\},\ R_i\right] \tag{21}$$

$$S.t\ R_{min} \leq R_i \leq R_{max} \tag{21a}$$

$$C_{ik} > 0 \tag{21b}$$

$$E_i > E_{min} \tag{21c}$$

3. Deep reinforcement learning Approach in UAV-enabled WSN

The deep deterministic policy gradient (DDPG) approach is solved using the deep reinforcement learning approach for the following reasons. It is off-policy optimization algorithm, which concurrently

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

learns Q-learning function, and uses the Q-function. An optimization approach only be used for dynamic environments with continuous action spaces, this allows the algorithm to setup efficient and gradient rule based for a policy when the Q-function is presumed to continuous actions. This approach is used to find the optimal action-value function $Q^*(s, a)$, the optimal action $a^*(s)$ is found by using

$$a^*(s) = \arg\max_{a} Q^*(s, a) \tag{22}$$

The optimal value function interleaves learning with the optimal action value $a^*(s)$ for continuous adaption to the changing environment with continuous action space. The computation of maximum over actions in $\max Q^*(s, a)$.

The bellman equation describing optimal action-value function is computed as

$$Q^*(s, a) = \mathop{E}_{s' \sim P} \left[r(s, a) + \gamma \max_{a'} Q^*(s', a') \right] \tag{23}$$

$s' \sim P$ represents the short form for next state where it is sampled by the environment $(.\,|s, a)$. It's a fast gradient optimization that needs to be solved right now. Every time increment may present an optimization problem, and numerical optimization utilizing a sophisticated penalty function may be required. In this section, we will investigate an alternative method by looking at a DRL rather than being difficult. A method for solving optimization problems that doesn't require optimization and makes use of the system's evolution. The execution phase starts after the control policy has been trained. In a case when multiple parties are engaged, this method is known for its straightforward nature and guaranteed convergence throughout runtime decision-making.

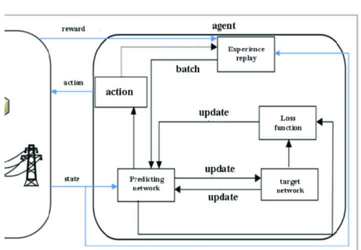

Fig.2 DDPG based optimization Approach

Received: August 05, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In DDPG algorithm as shown in figure 2, the environment space is represented by S_m, A_m,, $\mathfrak{p}_m$, $\Re_m$ involving state space, action space, state transition probability, and reward function, where UAV acts as an agent to decide the control actions based on their current state and reward with state transition probability, the decision is based on the loss function and the experience buffer.

The state space, represented as $S_m = \{1,2,3 \ldots \ldots N\}$, encompasses the limited number of vehicles making requests. The UAV is regarded as an agent referred to as N, where every agent $n \in N$ possesses a selection of UAV that can either transfer their tasks or perform finite actions (t) which involves choosing CH from the available cluster and sending data to the BS. Consequently, the restricted measures are implemented by the group of clusters to prevent excessive queue length.

The action space is $A_m = \{ {}_1(t), A_2(t), \ldots \ldots .\ .\ (t)\}$ determined by observing service rate and transmission duration. The one-step delayed service is included in calculating the aggregate of actions that all agents undertake collectively. A_m represents the collection of all potential actions by each agent within the group, signifying the aggregation of utility functions for all agents.

The state transition probability $S_{t+1} \sim \mathfrak{p}_m$ indicates the movement from one state to another for an agent and seeks to identify the best solution, $S_{t+1} \sim (S_{t+1}|S_t, A_t)$ The aim is to learn the optimal policy $A_m(t) = \pi_n^{BS}(S_m(t)) + \xi$.

the loss function is specified as ξ. Critic update

$$y_t = r_t + \gamma Q'(s_{t+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'} \tag{24}$$

The reward function of the proposed DRL is given as

$$(s, a) = \begin{cases} w_1(1 - P_{NLOS}) + w_2 C_{ik} + w_3 \mathscr{R} + w_4 E_i, & \forall i \in \mathcal{N}\ \{k\} \\ 0, & otherwise \end{cases} \tag{25}$$

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$S.t$ (1) to (3) (25a)

Then, this designed reward function leads to the behavior of the optimized UAV-WSN system.

The primary focus is on utilizing an DRL-based strategy for addressing coverage issue and achieving queue stability is based on taking advantage of the fact that the objective is to DRL is about maximizing the total reward, not just completing tasks. Being greedy and optimizing at each time step leads to achieving the desired outcome by having a carefully crafted reward function.

Algorithm 1 Deep Deterministic Policy Gradient

Input: initial policy parameters θ, Q-function parameters ϕ, empty replay buffer $\mathcal{D}$
Set target parameters equal to main parameters $\theta_{\text{targ}} \leftarrow \theta$, $\phi_{\text{targ}} \leftarrow \phi$
repeat
  Observe state s and select action $a = \text{clip}(\mu_\theta(s) + \epsilon, a_{Low}, a_{High})$, where $\epsilon \sim \mathcal{N}$
  Execute a in the environment
  Observe next state s', reward r, and done signal d to indicate whether s' is terminal
  Store (s, a, r, s', d) in replay buffer $\mathcal{D}$
  If s' is terminal, reset environment state.
  if it's time to update **then**
    for however many updates **do**
      Randomly sample a batch of transitions, $B = \{(s, a, r, s', d)\}$ from $\mathcal{D}$
      Compute targets

$$y(r, s', d) = r + \gamma(1 - d) Q_{\phi_{\text{targ}}}(s', \mu_{\theta_{\text{targ}}}(s'))$$

      Update Q-function by one step of gradient descent using

$$\nabla_\phi \frac{1}{|B|} \sum_{(s,a,r,s',d) \in B} \left(Q_\phi(s, a) - y(r, s', d)\right)^2$$

      Update policy by one step of gradient ascent using

$$\nabla_\theta \frac{1}{|B|} \sum_{s \in B} Q_\phi(s, \mu_\theta(s))$$

      Update target networks with

$$\phi_{\text{targ}} \leftarrow \rho \phi_{\text{targ}} + (1 - \rho)\phi$$
$$\theta_{\text{targ}} \leftarrow \rho \theta_{\text{targ}} + (1 - \rho)\theta$$

    end for
  end if
until convergence

To create a reward function for reinforcement learning that is effective. stability in queues, we have introduced DRL based algorithm

The over-fitting problem is prevented by using the loss function and the predicted network.

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The complexity of a proposed algorithm depends on the convergence and run time of the solution which is scaled as

$\sim(A_m \times S_m \times r(s, a))$, Overall, the total run-time increases linearly with the actions, states and their reward until convergence. It is a centralized approach to reach the global optimum value. The experiences are stored in replay buffer (s_t, a_t, r_t, s_{t+1}). sample the mini batch from buffer. The chosen weights w_1, w_2, w_3, w_4 are adjusted based on the significance of the importance of the reward parameters, The LOS probability is the main parameter for the

working of the entire UAV-WSN as of changing the topology with w_1. $w_2 C_{ik}$ is the second important coverage parameter to improve the coverage of the entire network, Data transmission is based on the rate and the significant importance in the weight w_3 at which data routing from the CH to UAV causing data transmission to avoid loss of

information. Next priority is given to the energy efficiency for the network's lifetime and communication

reliability. A minimum reward value as zero is given to the network for not performing well.

Algorithm 2 DDPG based UAV Assisted WSN coverage Algorithm	
Input:	UAV initial harvested energy E_i, set of sensor nodes $S = \{s_1, s_2, \dots, S_K\}$ and set of UAVs $U =$ $\{u_1, u_2, \dots, u_N\}$
Output:	Energy Efficient UAV assisted WSN coverage for UAV-CH pair
Phase 1: Cluster Node deployment and UAV Location Acquisition	
1:	**if** time to update node location arrives then
2:	**for** each UAV do
3:	Obtain UAV location updates via GPS
4:	**end for**
5:	**end if**
Phase 2: Sensor node and CH Discovery	
1:	Call Algorithm 1 for updating the sensor node and CH information
Phase 3: Computation of parameters	
1:	Compute SIR_{ik} using (17)
2:	Compute C_{ik} using (18)
3:	Compute Achievable data rate $\mathscr{R}$ using (20)
Phase 4: UAV Operational status	

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1:	Initialize UAV-WSN communication
2:	**while** next-hop node$\neq zero$ do
3:	**for** each UAV, do
4:	**if** $C_{ik} > 0$ then
5:	**Start** data transmission and Neighbor Discovery
6:	**else check if** $E_i > E_{min}$
7:	**else** UAV mode=charge
8:	**Update** reward using (25)
9:	**Update** weights
10:	**end if**
11:	**end for**
12	Update Q-matrix
13:	**end while**

4. Experimental Results and Discussion

The actor-critic network is observed with the network loss function and loss gradient, Accordingly, the discount factor and learning factor are adjusted between 0.1 and 0.99 to control the trajectory parameters such as speed and network weights. The simulation parameters are shown in table 1 and 2 for the experimental setup. The loss function is observed to predict the error and convergence.

Table 1 Simulation parameters for DDPG network

Parameter	*Values*
Total Episodes	1000
Time steps	1000
Discount Rate	{0.5, 0.9, 0.999}
Learning Rate	0.001
Size of B (Replay Network capacity)	100000

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Parameter	*Values*
Optimizer	Adam Optimizer
Minimum batch size	100

Table 2 Simulation parameters for a communication link

Parameter	***Values***
Number of Base stations	2
Number of Sensor Nodes	100
Number of UAVs	3
Distance between UAVs	20 m- 100 m
Bandwidth	100 MHz
Pathloss Exponent	-50 dB
Transmitted Power of BS	46 dBm
UAV Bandwidth B_1	[0.5, 1, 1.5, 2] MHz
Noise power	-100 dBm

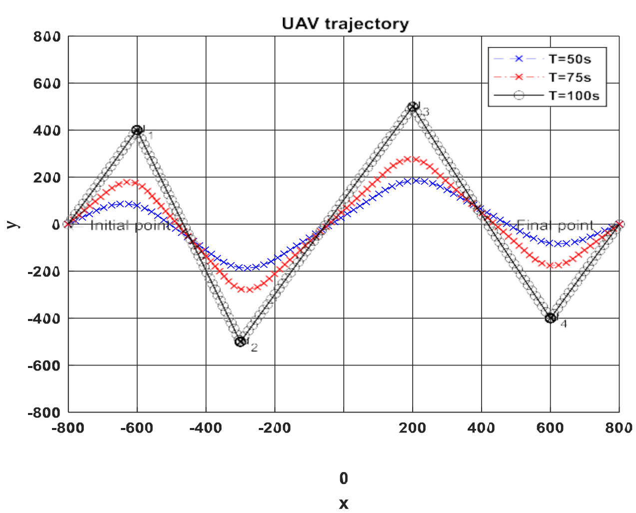

Figure 3 UAV Trajectory Path with varying speed

UAV trajectory optimization over varying time durations is shown in figure 3 at $T = 50\ sec$, 75 sec and 100 sec. The flight path evolves to meet the requirements of CH coverage and the data transmission. For

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

shorter duration $T =$

50 sec, the UAV followed as more energy constrained and smoother path with minimal energy consumption. The waypoints indicated a priority-based coverage and strategic positions during mission. The initial and final points are consistent across all durations implying energy budgeting.

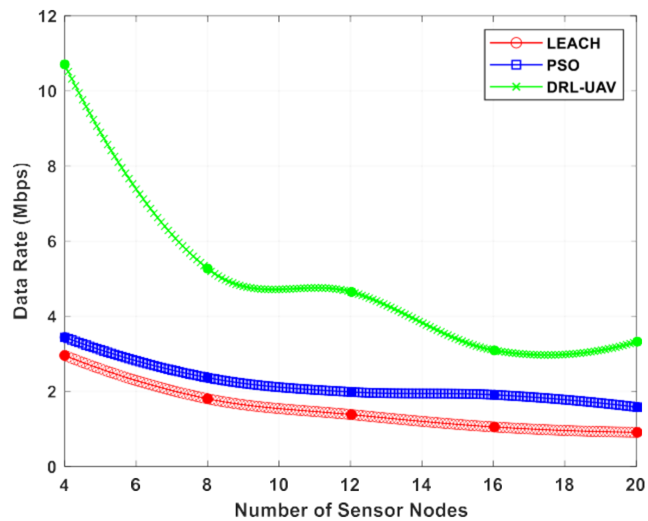

Figure 4 Comparative analysis of data rate performance

Figure 4 presents a comparative analysis of data rate performance as a function of the number of sensor nodes for different approaches. LEACH (Low-Energy Adaptive Clustering Hierarchy) , PSO (Particle swarm optimization) and with proposed method. The data rate is decreased when the sensor nodes increased, this is due to increased network contention, greater communication overhead and limited bandwidth shared among the nodes.The comparison highlights the importance of intelligence and adaptive strategy especially during the dynamic movement of UAV enabled architecture.

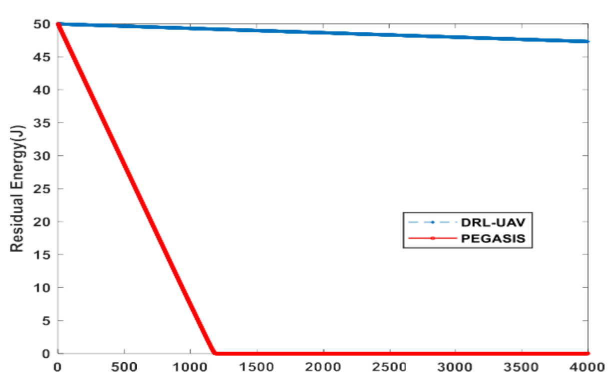

Figure 5 Residual Energy over time

Received: August 05, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The plot shown in figure 5 describes the residual energy over time for two different approaches such as PEGASIS (Power-Efficient Gathering Sensor Information System) in as WSN, The proposed DRL-UAV appraoch maintains a consistently high level of residual energy throughput rounds, The energy decreases steadily and optimized UAV

trajectory and allocation. The DRL-UAV approach prologs network lifetime where UAVs are expected to operate over long periods without maintanence. The tracking of alive nodes and dead nodes are observed during the operational mode over time, The network alive nodes per round is shown in figure 6 to be analyzed for netwrok longevity under different management strategies, the communication reliability is ensured over extended operations.

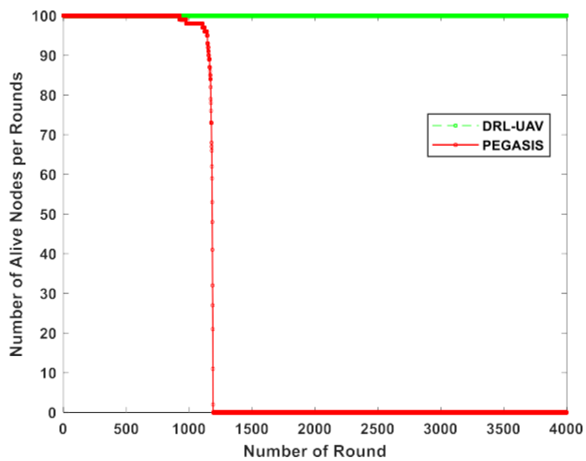

Figure 6 Number of Alive nodes over time

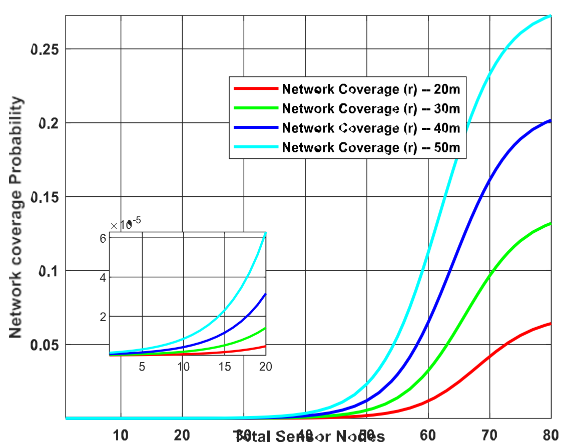

Figure 7 Network coverage analysis over distance

Received: August 05, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DRL-UAV network coverage approach is shown in figure 7, marginally better performance over different distance, maintaining network connectivity for slightly more time before the inevitable energy depletion causes widespread node failure. The energy efficient data collection is WSN offering a more sustainable alternative to traditional methodologies.

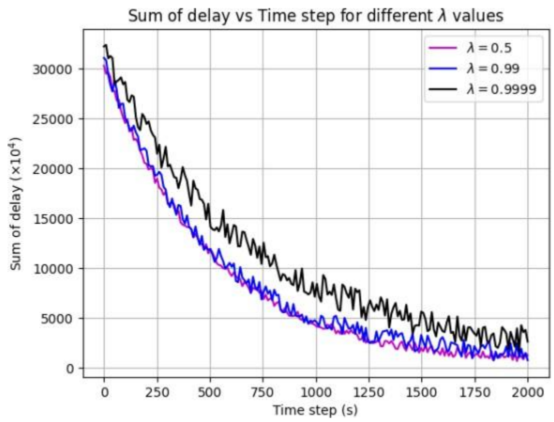

Figure 8 sum of delay for different discount factor values

The DDPG learning agent is trained through different learning factor started from 0.5 to 0.9999 to show how the agent responds through vigorous training in figure 8, the sum of delay versus time step is shown, we can observe that the increase in learning rate increases to perform better slightly more rounds. Figure 8 is the obtained solution for the sum of delay and cost in terms of time step. It is taken for increasing number of nodes requesting UAVs and the finding the global solution

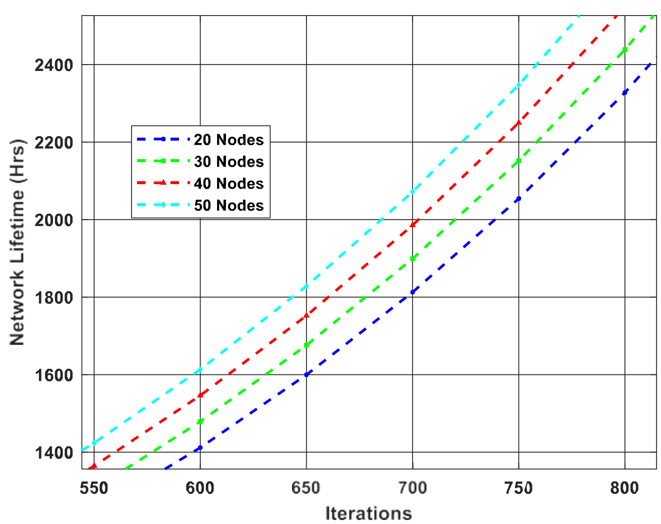

Figure 9 Network Lifetime over nodes

Figure 9 The network lifetime is given for increasing sensor nodes where the rate experienced by individual UAVs varied from 0.5 to 2 Gbps for the DRL approach without cost optimization, respectively. The proposed algorithm increases the total rate by 30% in terms of corroboration with the DRL compared to these benchmarks. The performance is shown in figure 9 for the steady state data rate in the use of DRL-UAV algorithm.

5. conclusion

The proposed work an UAV assisted WSN is solved using the DDPG approach, the problem of coverage and trajectory optimization through DRL A DDPG algorithm is utilized to derive the optimal joint problem of coverage and trajectory control and effectively assigned WSN for efficient data transmission, The evaluation results are

shown in the results and analysis part to produce the successful learning of the agents and outperforms in terms of sum-rate maximization and minimum delay. The results of the evaluation show that our method for efficient coverage and achieves superior performance when compared to other optimization methods. The performance including trajectory waypoint optimization with different time duration, residual energy of WSN, network lifetime and data rate are realistically simulated and shown.

REFERENCES

A. Alwarafy, B. S. Ciftler, M. Abdallah, M. Hamdi, and N. Al-Dhahir, “Hierarchical multi-agent DRL-based framework for joint multi-RAT assignment and dynamic resource allocation in next-generation het nets,” IEEE Trans. Netw. Sci. Eng., early access, Apr. 5, 2022, doi: 10.1109/TNSE.2022.3164648.

L. Qian, Y. Wu, N. Yu, F. Jiang, H. Zhou, and T. Q. Quek, “Learning driven NOMA assisted vehicular edge computing via underlay spectrum sharing,” IEEETrans.Veh.Technol.,vol.70,no.1,pp. 977–992,Jan.2021.

Q. Yuan, J. Li, H. Zhou, T. Lin, G. Luo, and X. Shen, “A joint service migration and mobility optimization approach for vehicular edge comput ing,” IEEE Trans. Veh. Technol., vol. 69, no. 8, pp. 9041–9052, Aug. 2020.

D. D. Nguyen, H. X. Nguyen, and L. B. White, “Reinforcement learning with network-assisted feedback for

heterogeneous RAT selection,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6062–6076, Sep. 2017.

Samuel Amalorpava Mary Rajee and Arulraj Merline. Machine intelligence technique for blockage effects in next- generation heterogeneous networks. Radioengineering, 29(3), 2020

Samuel, A.M.R., Arulraj, M.: Performance analysis of flexible indoor and outdoor user distribution in urban multi- tier heterogeneous network. Int. J. Mob. Commun. **21**(1), 119–133 (2023)

Zhai, H., Zhao, R., Li, H., Wu, L., Chen, X. (2023). A Decision-Making Method Based on Dynamic Programming Algorithm for Distribution Network Scheduling. “Cyber Security Intelligence and Analytics. CSIA 2023. Lecture Notes on Data Engineering and Communications Technologies, vol 172. Springer, Cham. https://doi.org/10.1007/978-3-031-31860-3_21

J.Bi,H.Yuan,S.Duanmu,M.Zhou,andA.Abusorrah,"Energy-optimized partial computationoffloading inmobile- edgecomputingwithgenetic simulated-annealing-basedparticleswarmoptimization,"IEEEInternet ThingsJ.,vol.8,no.5,pp.3774–3785,Mar.2021.

Longxin Zhang, Runti Tan, Yanfen Zhang, Jiwu Peng, Jing Liu, Keqin Li, UAV-assisted dependency-aware computation offloading in device–edge–cloud collaborative computing based on improved actor–critic DRL, Journal of Systems Architecture, Volume 154, 2024, 103215 https://doi.org/10.1016/j.sysarc.2024.103215

. Zhou H., Wang Z., Zheng H., He S., Dong M. Cost minimization-oriented computation offloading and service caching in mobile cloud-edge computing: An A3C-based approach IEEE Trans. Netw. Sci. Eng., 10 (3) (2023),

pp. 1326-1338

"He X., Jin R., Dai H. Multi-hop task offloading with on-the-fly computation for multi-UAV remote edge computing IEEE Trans. Commun., 70 (2) (2022), pp. 1332-1344

Chen Y., Ai B., Niu Y., Zhang H., Han Z. Energy-constrained computation offloading in space-air-ground integrated networks using distributionally robust optimization IEEE Trans. Veh. Technol., 70 (11) (2021),

pp. 12113-12125

F. Zeng, Q. Chen, L. Meng, and J. Wu, "Volunteer assisted collaborative offloading and resource allocation in vehicular edge computing," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 6, pp. 3247–3257, Jun. 2021.

X. Bai, W. Yan, and S. S. Ge, "Efficient task assignment for multiple vehicles with partially unreachable target locations," IEEE Internet Things J., vol. 8, no. 5, pp. 3730–3742, Mar. 2021.

A. Alwarafy, B. S. Ciftler, M. Abdallah, M. Hamdi and N. Al-Dhahir, "DRL-Based Joint RAT Association, Power and Bandwidth Optimization for Future HetNets," in *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1503-1507, July 2022, doi: 10.1109/LWC.2022.3177250.

F. Jiang, Y. Peng, K. Wang, L. Dong and K. Yang, "MARS: A DRL-Based Multi-Task Resource Scheduling Framework for UAV With IRS-Assisted Mobile Edge Computing System," in *IEEE Transactions on Cloud Computing*, vol. 11, no. 4, pp. 3700-3712, Oct.-Dec. 2023, doi: 10.1109/TCC.2023.3307582.