



Multi-objective intelligent clustering routing schema for internet of things enabled wireless sensor networks using deep reinforcement learning

Walid K. Ghamry^{1,2} · Suzan Shukry³

Received: 24 July 2023 / Revised: 20 October 2023 / Accepted: 24 November 2023 / Published online: 9 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The Internet of Things (*IoT*) is built on a foundation of wireless sensor devices that connect humans and physical objects to the Internet and enable them to interact with one another to improve the living conditions of citizens. Wireless Sensor Networks (*WSNs*) are widely utilized in systems based on *IoT* to collect the data required by intelligent environments. However, *IoT* – enabled *WSNs* encounter a variety of difficulties such as poor network lifespan, limited throughput, and long communication delays, due to the massive non-homogenous data streaming from numerous sensor devices. Therefore, a multi-objective intelligent clustering routing schema for *IoT* – enabled *WSNs* utilizing deep reinforcement learning is proposed in this paper to overcome these shortcomings. The proposed schema partitions the entire network into various unequal clusters based on the present data load existing in sensor nodes, effectively preventing the network from dying prematurely. In addition, an unequal clustering mechanism is utilized to balance inter-cluster and intra-cluster energy consumption among cluster heads. The simulation findings demonstrate the effectiveness of the proposed schema in terms of energy efficiency, delivered packets, end-to-end delay, alive nodes, energy balancing, and network lifespan compared with the other two state-of-the-art existing schemes.

Keywords Deep reinforcement learning (*DRL*) · Multi-objective · Wireless sensor networks (*WSNs*) · Intelligent routing · Internet of things (*IoT*)

1 Introduction

Over recent years, *IoT* device technologies have evolved significantly, leading to the development of paradigms for dynamic wireless sensing technology to provide seamless communications over the Internet [1]. Wireless sensor networks are a crucial element in the *IoT* that plays a basic and vital role in collecting data and communications

through the fifth generation and beyond from the perception of sixth-generation Internet networks [2–4].

The *IoT* – enabled *WSNs* have many possibilities in various smart applications. In the military, smart applications often involve sensitive data, privacy, and security concerns, as well as the monitoring of critical military zones to enhance national defense [5, 6]. In a home automation network, valuable data is gathered from monitoring sensors placed within the environment using smart consumer sensor nodes. Subsequently, this collected data is transmitted to the central base station with no direct human intervention [7, 8].

The *IoT* – enabled *WSNs* play a crucial role in monitoring and providing real-time predictions for various environmental events such as floods and tsunamis in oceanic regions, monitoring rainfall patterns, detecting seismic activities related to earthquakes, and monitoring volcanic eruptions [9]. An innovative emergency evacuation system is designed to identify potential hazards, such

✉ Walid K. Ghamry
wkamal@bu.edu.sa

Suzan Shukry
suzan.shouckrey@hti.edu.eg

¹ Computer Science - Faculty of Computing and Information, Al-Baha University, Al Bahah, Saudi Arabia

² Information Systems Engineering Department, National Research Center, Cairo, Egypt

³ Higher Technological Institute, Cairo, Egypt

as fires, noxious gases, and the presence of individuals, within an indoor monitoring environment [10]. The system aims to provide a safe and unobstructed path for evacuation, prioritizing the shortest and safest route for individuals during emergencies.

The *IoT* – enabled *WSNs* have been introduced to periodically record the internal conditions of the patient in healthcare monitoring [11] and have found a wide array of industrial applications, ranging from enhancing product quality to promptly monitoring machine efficiency [12]. In Smart Transportation, traffic surveillance application leverages *IoT* technology to monitor and manage traffic flow and conditions in real-time, enabling more efficient and informed decision-making in urban mobility and transportation management [13].

Accurate and timely data acquisition is paramount for these real-time smart applications, ensuring that correct decisions are made within their respective environments. Delays in data collection can often result in heightened consequences, particularly in critical domains such as healthcare and forest fire management. Therefore, the precise and prompt collection of data has evolved into an essential requirement for these smart applications. Mobile sink-based data acquisition stands out as a highly significant technique for achieving accurate, delay-free data collection, executed efficiently with commendable performance [14, 15].

The sensors deployed in *IoT* – enabled *WSNs* continuously monitor the surrounding physical environment and transmit the sensing information directly to the Base Station (*BS*). However, the repercussions of unbalanced power consumption and reduced lifespan of battery-powered sensor nodes limit the seamless connectivity of smart devices over the *IoT* network, and thus the sensing data will not be continually transmitted to the *BS* [16].

These shortcomings produce several problems within the network, such as higher communication delay and an imbalance in energy consumption among all the deployed sensor nodes, which are unacceptable in particular applications. To avoid these repercussions, *WSNs* are often designed in a hierarchical structure partitioned into small different clusters [17, 18]. Each cluster has two categories of sensor nodes: Cluster Heads (*CHs*) and Cluster Members (*CMs*).

The communication between the sensor nodes in the clustering approach is classified into two communication modes: intra-cluster and inter-cluster. In intra-cluster communication, non-*CH* nodes (*CMs*) transmit their data to the respective *CH*, while in inter-cluster communication, the respective *CH* fuses the aggregated data to the *BS* either directly or through multi-hop routing [19, 20]. Clustering is considered the robust approach for increasing network lifetime and achieving higher energy-efficient data transmission [21].

Nevertheless, the existing clustering routing approaches in the literature suffer from severe issues, including an

increase in communication delays, ineffective performance as evidenced by lower throughput, and a hot spot problem [22–24]. In addition, heavy traffic loads are introduced within the networks due to the massive messaging overhead in partitioning the networks into various clusters, particularly when the size of the network becomes larger, which causes an unbalance in energy consumption between all the deployed sensor nodes. Routing utilizing experienced-based Reinforcement Learning (*RL*) is a promising technique to solve the aforementioned issues [25].

RL represents a branch of the Machine Learning (*ML*) approach that explores the interaction with the local environment to acquire knowledge [26]. In *RL* technique, the Q-Learning method is usually employed to choose a routing path, where the reward represents the routing metric in the learning operation. However, the state-action pairs of *RL* are often small such that existing *RL* routing techniques cannot exploit the most historical information of all the dynamic network traffic changes to choose the optimal routing path, due to the renowned “curse of dimensionality”, and thus the space complexity of the state-action pairs becomes a major obstacle to the proliferation of *RL* routing methods [27].

The “curse of dimensionality” has recently been overcome and averted to a great extent through applying *DRL*, which relies on a Deep Neural Network (*DNN*) to realize the logical relationship between the states and actions, ensuring that all state-action pairs in *DRL* do not need to be traversed as in Q-Learning method [28]. *DRL* technique has become popular in designing many successful complex *IoT* systems such as resource optimization [29], cellular scheduling [30], video streaming [31], and routing policy against hard traffic patterns predictable [32].

Most of the prior works applying reinforcement learning in network routing problems focus on addressing single objective parameters such as communication delay or message overhead. Although, in many real-life problems, network routing methods often deal with multi-objective parameters such as network latency, energy saving, and channel bandwidth. The objectives can be directly related, independent, and conflicting. In most routing problems, some of the objectives are often conflicting with others, so that maximizing one object leads to minimizing another. Therefore, a trade-off between objectives is considered the challenge issue to be addressed and overcome.

Inspired by the potential of *DRL* and given the aforementioned limitations, this paper proposes a multi-objective intelligent clustering routing schema for *IoT* – enabled *WSNs* to avoid hot spot problem, reduce latency and message overhead as well as prolong network lifetime. An unequal clustering mechanism is proposed to balance the intra-cluster and inter-cluster energy consumption that prolongs network lifespan and maximizes network throughput as well as avoid hot spot issue.

Moreover, a Multi-Objective *DRL(MODRL)* intelligent routing technique is proposed to minimize network latency and network messaging overhead. Thus, an enhanced network quality of service is obtained and the problem of network partition can be avoided and overcome, through intelligent clustering routing in *IoT* – enabled *WSNs*.

The following are the main contributions of this study.

- (1) The study introduces a new mechanism based on unequal clustering to effectively prevent the hot spot problem in *IoT* – enabled *WSNs*. This mechanism alleviates uneven energy consumption among nodes, ultimately enhancing network reliability and longevity.
- (2) The study presents an innovative load-balancing schema, both intra-cluster and inter-cluster, to optimize energy consumption in *IoT* – enabled *WSNs*. This schema aims to prolong the network lifespan and maximize overall network throughput, providing a more sustainable and efficient network infrastructure.
- (3) Furthermore, the study introduces an intelligent routing technique based on *MODRL* for reducing network latency and minimizing messaging overhead significantly. By adopting *MODRL*, the study contributes to more efficient and responsive communication within the network.
- (4) Finally, comprehensive simulations illustrate the efficiency and effectiveness of the introduced schema. The findings highlight that the introduced schema outperforms existing schemes, signifying a substantial improvement in system performance and contributing valuable insights into *IoT* – enabled *WSNs*.

This paper is organized as follows: Section two discusses the related work. Section three covers the preliminaries through which introduced principles are presented. The detailed design of the *MODRL*-based clustering routing schema is given in section four. Complexity analysis of the proposed schema is discussed in section five. Section six presents simulation experiments results and discussion. Section seven concludes the study.

2 Related work

A brief overview of the existing literature review that concentrates on routing methods in *IoT* – enabled *WSNs* using experienced-based reinforcement learning is introduced in this section. The existing literature works can be categorized as follows.

2.1 RL-based routing protocols

The first attempt to apply a reinforcement learning approach to the routing problem is proposed in [33]. A Q-

routing algorithm for packet routing, based on the Q-learning model, is proposed to choose the best route that achieves a single objective parameter, the smaller mean delivery delay. However, the limited lifespan of battery-powered sensor nodes is not considered in this algorithm, resulting in a shorter network lifespan.

An Adaptive Spanning Tree Routing Protocol (*ASTRP*) is proposed in [34] based on reinforcement learning to achieve two objectives, load balancing and congestion evasion. The simulations demonstrate that the proposed routing protocol is robust for unexpected failures. However, the protocol suffers from significant communication delays resulting in low throughput, particularly in larger-scale networks of high traffic loads.

An Adaptive Routing (*AdaR*) strategy is proposed for *WSNs* in [35] based on Q-Learning and Least Squares Policy Iteration (*LSPI*). The *AdaR* considers multi-objective parameters such as residue energy, hop count, and aggregated proportion to evaluate an optimal Q-value for a given policy. The results demonstrate that *AdaR* obtains a high convergence speed. However, it has a poor throughput.

A Feedback Routing for Optimizing Multiple Sinks (*FROMS*) method based on Q-learning for multicast routing is proposed in [36] for *WSNs*. *FROMS* considers multi-objective parameters such as communication delay, battery energy, and hop count to choose the optimal path, which delivers packets from a source node to multiple sinks. *FROMS* has a drawbacks of low network lifespan and high messaging overhead. An extension to *FROMS*, *E – FROMS*, is introduced in [37] to address energy consumption in *WSNs*.

A routing protocol is presented in [38] for underwater *WSNs*. The remaining node energy and the node group's average energy are considered to choose a forward node and balance energy consuming. The proposed protocol prolongs the network lifespan over other protocols. However, it suffers from a poor ratio of delivery packets. A Distributed Adaptive Cooperative Routing (*DACR*) protocol is proposed in [39] considering reliability, communication delay, and residual energy to find the optimal path that consumes the lowest amount of energy to prolong network lifespan.

Multi-agent Reinforcement Learning Based Self-Configuration and Self-Optimization protocol (*MRL – SCSO*) is proposed in [40] for unattended *WSNs*. It considers both remaining energy and buffer length for effective routing, as well as utilizing sleep scheduling schema to conserve energy. This protocol provides a longer network lifespan and higher throughput, however, it has drawbacks of increasing communication delay and poor delivery of packets.

A Reinforcement-Learning Based Routing (*RLBR*) protocol is proposed in [41] to improve the network lifespan of

WSNs. The protocol considers three parameters such as remaining energy, hop count, and link distance to find the next forwarder node. This protocol provides a gain to decrease the total energy consumed and increase the delivery of packets. However, it has drawbacks of high communication delay and energy imbalance.

In [42], a Q-learning-based Data Aggregation-aware Energy-Efficient Routing (*Q-DAEER*) protocol is proposed. The protocol considers link distance, energy node, hop count, and dynamics of node data aggregation to find the optimal path that prolongs network lifespan and decreases energy consuming. An *RL* – based routing protocol is presented in [43] to achieve effective energy consumption and improve network lifespan. It considers the current state of the network to find an optimal route that minimizes the delay and increases the reliability.

Another work for underwater *WSNs* is presented in [44]. An *RL* – based routing approach is proposed to set up the optimal path to a destination. It considers residual energy and the underwater environment to select the forwarder node on the optimal routing path. A Q-learning-based transmission routing scheme is proposed in [45] to decrease and balance the energy consumption of the sensor nodes and prolong the network lifespan. The routing scheme considers four factors, distance, transmission direction, residual energy, and energy consumption to find a suitable forwarder node that obtains effective energy transmission in a distributed manner.

An *RL* – based tree routing algorithm is proposed in [46] to achieve multi-objective in *WSNs* such as minimizing link breaking and congestion avoidance. The algorithm formulates three types of cognitive metrics to find the best parent node in the tree routing. The algorithm provides a gain to reduce the delay, increasing the packet delivery ratio, and reducing energy consumption.

2.2 DRL-based routing protocols

Numerous routing protocols employ *DRL*, and the majority of them use it to select a data routing relay node. The study in [47] develops a deep-Q-network-based cooperative and adaptive approach to identify the optimum relay node. In *WSNs*, compared to Q-learning-based methods, it enhances the Quality of Service (*QoS*) for networks. In essence, the approach disregards communication delay and just concentrates on node relaying.

In wireless ad-hoc networks, [48] develops a multi-hop routing strategy utilizing the *DDQN* paradigm in *DRL* to find the best-relaying node. Additionally, it is a routing protocol for selecting relay nodes that ignores communication delay and message overhead. The study in [49] proposes a *DRL* – based routing protocol to find the optimum shortest path for network control and management. This method just takes distance into account when routing data. Hence, this strategy results in poor *QoS*.

The study in [50] investigates the viability of the *DRL* method to solve a problem with two objectives: maximizing throughput and energy-effective routing. The study introduces a multi-objective actor-critic model-based Proximal Policy algorithm (*PPO*) to find near-optimal solutions.

A decentralized collaborative *DRL* – based routing protocol is introduced in [51] to efficiently enhance and manage *P2P* wireless sensor network routing. It learns *WSN* routing policies using extended parameters for state space and a neural network.

The work in [52] investigates the utilization of routing technology and *DRL* together to provide an effective routing technique for adapting to changes in network topology. The nodes can decide on routing based on energy consumption level and network traffic load to find the optimal path. A *DRL* technique is adopted in [53] to optimize routing in dynamic Internet of Things networks. The routing strategy is implemented in both distributed and centralized modes.

The study in [54] introduces a fault diagnosis model referred to as Multi Fault Detector (MFD) for sensor nodes, which is based on a Neural Network (NN) approach. The model utilizes historical data encompassing instances of both faults and fault-free conditions within the network. The MFD model is engineered to handle a diverse range of fault types, including hard permanent, soft permanent, intermittent, and transient faults. Notably, the proposed MFD model goes beyond mere fault detection; it is also capable of categorizing the faulty nodes and identifying problematic links associated with the sensor nodes in the network, thus providing a comprehensive fault diagnosis solution.

2.3 Cluster-based learning routing protocols

Clustering means partitioning nodes into several groups, with each group belonging to its cluster header. The use of reinforcement learning in cluster-based routing protocols has been extensively studied. The authors of *FROMS* extend their work in [55] and propose a Q-learning-based cluster routing technique to cope with energy conservation. The algorithm takes two objectives into account such as battery power and hop count to determine the efficient *CHs*. This algorithm provides lower clustering overhead, however, it has a problem with energy holes.

The study in [56] proposes a Q-learning-based hierarchical routing scheme. The scheme takes three objectives into account such as residual node energy, link distance, and hop count to perform routing and clustering within a network. However, it performs poorly in large-scale *WSNs* in terms of delay and throughput.

The work in [57] proposes an *RL* – based clustering routing algorithm to effectively conserve energy and prolong network lifespan in *IoT* – enabled *WSNs*. The

algorithm considers four different objectives such as distance, traffic intensity, delay, and energy level for efficient *CH* selection. In addition, the algorithm utilizes *DRL* to identify the shortest path for data transmission. The study in [58] presents an *RL* – based clustering routing algorithm for effective energy control in *WSN*. This algorithm aims to maximize each node’s long-term reward through optimizing routing policies. Additionally, the algorithm proposes three energy management strategies to improve network lifetime.

The research in [59] proposes an *RL* – based enhanced clustering routing algorithm to manage energy efficiently in *WSNs*. The algorithm takes two different objectives into account such as hop count, and initial energy to determine the effective *CH*. Moreover, three stages are introduced to look for the most efficient data transmission routing path. The work in [60] proposes an *RL* – based clustering routing strategy to reduce energy consumption and extend network lifespan. The strategy considers two factors such as initial energy and hop count to determine the preliminary Q-value for *CH* selection. In addition, hop count and remaining energy are considered to select the optimum routing path for transmitting data.

A novel method called floating node-assisted cluster-based routing has been presented in [61] for effective data collecting in underwater acoustic sensor networks, utilizing the unique characteristics of underwater communication. In this method, clusters are formed by dividing the network space into cubes. Each *CH* in the cubes is wired to a floating node and source nodes are in charge of transmitting the sensed data to the nearest *CH* or floating node. The floating nodes receive the data collected by the *CHs* and transmit it across a radio frequency link to the on-shore monitoring center.

In [62], the authors introduce an intelligent fault-tolerance technique in order to enhance the resilience of *IoT* – enabled *WSNs*. The proposed key solutions encompass a range of techniques, including the utilization of a Maximum Coverage Location Problem (MCLP) method for identifying optimal locations for *CH* placement. Additionally, the study introduces a *MODRL* method, which serves a dual purpose: fault detection with minimal energy consumption and the selection of optimal data routing paths under fault-free conditions. The study also presents a mobile sink-based data-gathering scheme designed to further enhance the network’s overall reliability.

3 Preliminaries

This section discusses the fundamental principles employed in the proposed work.

3.1 Energy model

To assess the energy consumed by a sensor node when transmitting and receiving a data bit, E_{Tx} and E_{Rx} , the energy model adopted in [63] is taken into consideration. The required energy to receive and transmit data of size l bits over a distance d is expressed as

$$E_{Tx}(l, d) = \begin{cases} l \times E_{elec}^{tx} + l \times \varepsilon_{fs} \times d^2, & d < d_0 \\ l \times E_{elec}^{tx} + l \times \varepsilon_{amp} \times d^4, & d \geq d_0 \end{cases} \quad (1)$$

$$E_{Rx}(l) = l \times E_{elec}^{rx}, \quad (2)$$

where E_{elec} denotes the energy dissipation per bit in the receiver or transmitter circuits, ε_{amp} and ε_{fs} represent the energy consumed by the power amplifier per data bit for multi-path radio channel and free space models, respectively. In addition, the radio channel model is specified by the threshold distance $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{amp}}$, and the physical distance between sender and receiver nodes is denoted by d .

3.2 Wireless sensor network model

We depict our *WSN* model as a directed graph $G(V, E)$, where V stands for the set of vertices and E for the set of directed edges that each connects an ordered pair of vertices. The vertices represent sensor (non-*CH* or *CH*) nodes and the edges represent wireless links between them. The cluster head nodes are distinguished as advanced nodes in comparison to other sensor nodes. The sensor nodes are connected to *CH* nodes within the communication radio *CH* range. It is assumed the following to evolve the proposed routing algorithm.

1. Advanced nodes and sensor nodes are distributed at random over a square area.
2. Each sensor node has the same limited energy capacity, processing power, and memory storage.
3. The sensor node may adjust the level of its transmitter power based on the receiver’s distance.
4. Sensor nodes and advanced nodes remain stationary.
5. Advanced nodes have more effective energy compared to sensor nodes.
6. Contrasted with sensor nodes, the number of advanced nodes is extremely low.
7. The sink node (base station) has unlimited energy.

The *WSN* model initially assumes that each sensor node has the same maximum residual energy.

4 MODRL-based clustering routing schema

This section introduces the proposed schema’s comprehensive design process in more detail. Four stages make up the proposed schema: (1) initialization stage, (2) unequal

cluster construction stage, (3) *MODRL*-based clustering routing stage, and (4) Energy consumption stage.

4.1 Initialization

In the initialization stage, the sink node broadcasts an advertisement message *BS_ADV* to all sensor nodes within its coverage area. Each sensor node that receives a *BS_ADV* replies to *BS* with an *RPL MSG* including sensor location, ID, and residual energy. Then, each sensor node in turn sends an *SN_ADV* message to neighbor nodes, containing sensor ID and residual energy to hold.

4.2 Unequal cluster construction

Once the *WSN* is initialized, *BS* collects and stores all network-entire information, such as sensor ID, distance, and residual energy. Following that, *BS* maintains a list of all sensor node's information in decreasing order of remaining energy and picks the top 10% of them as advanced nodes (*CHs*). The proposed schema adopts an unequal cluster mechanism to balance network-entire energy consumption (load) and avoids hot spot problem. As the forwarding load increases with node proximity to the base station, a cluster nearer to the *BS* should be smaller in size than a cluster further away.

To produce an unequal cluster that balances the load between the clusters nearest to the *BS* and clusters further away from it, each advanced node AD_k should compute its cluster radius using the equation below.

$$R_{AD_k} = \left[1 - c \cdot \frac{d_{\max} - d_{AD_k-BS}}{d_{\max} - d_{\min}} \right] \cdot R_{\max}, \quad (3)$$

$$\forall k, k \in \{1, \dots, n\},$$

where d_{AD_k-BS} is the distance between *BS* and any advanced node AD_k , d_{\min} and d_{\max} stand for the minimum and maximum distance from the selected *CHs* to *BS*, respectively, c is a weighted factor with a value between 0 and 1, and R_{\max} is the maximum transmission range of advanced nodes. The distance is calculated as $d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$, where (x_i, y_i) and (x_j, y_j) are the coordinates of the two nodes.

The size of the nearest *BS* cluster in the proposed schema is small in comparison to the farthest cluster to spend less energy on intra-cluster communication traffic and conserve more energy for inter-cluster relay communication traffic. In other words, it balances the load generated by the aggregation of data from both inter-cluster heads and intra-cluster members. In addition, when the *CH* distance to *BS* rises, the corresponding cluster radius gradually rises to maintain the *CH* node's and its cluster member nodes' dissipation of energy in balance.

Figure 1 depicts our proposed unequal clustering-based *WSN* schema architecture.

Once *CHs* are elected and their cluster radii are determined, the next challenge is the cluster formation. Each advanced node broadcasts cluster-forming message *CLFM* within its coverage cluster radius area to form the members of the cluster (non-*CHs*). The *CLFM* includes information about advanced node residual energy, location, distance to *BS*, and ID. In this context, there are four potential cases for replying to the message, as follows:

Case 1: If a sensor (non-*CH*) node overhears and receives the *CLFM*, it responds to the corresponding advanced node with cluster member joining message *CMJ* containing its residual energy, location, and ID.

Case 2: A sensor (non-*CH*) node may overhear and receive the *CLFM* from multiple advanced nodes. In such a case, the sensor node will select the advanced node with maximum residual energy as its corresponding *CH*.

If there are more than one *CH* has the same maximum residual energy, the *CH* with the smallest ID is picked.

Case 3: Sensor (non-*CH*) nodes may be located in an intersecting area of the cluster radius of neighboring advanced nodes. In such cases, these sensor nodes are referred to as autonomous non-*CH* nodes. The autonomous sensor nodes have the option of sending *CMJ* to whichever of the neighboring clusters at random.

Case 4: If a sensor (non-*CH*) node does not overhear and receive the *CLFM*. In such a case, this sensor node is referred to as a lone node and broadcasts an assistance message *ASSIST* to neighboring nodes within its communication range. Each neighboring node replies with a respond message *RPM* containing its ID, location, and corresponding advanced node information (ID and residual energy). Then, the lone node sends a *CMJ* to the corresponding *CH* which has the maximum residual energy and closest distance to *BS*.

The details of unequal cluster construction are provided in Algorithm 1.

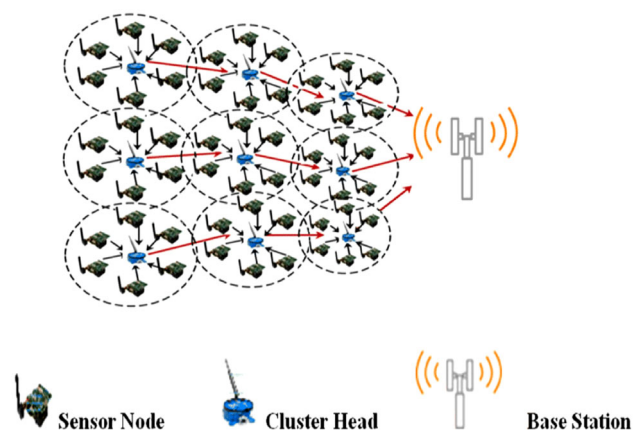


Fig. 1 Unequal clustering-based *WSN* schema architecture

Algorithm 1 Unequal cluster construction

Input: initialization of a parameter

a: $SN = \{SN_1, SN_2, \dots, SN_l\}$; // Set of IoT sensor nodes

b: BS ; // Base station

c: n ; // Number of CHs , $n \ll l$

d: $CH = \phi$; $CMLIST = \phi$;

Output:

a: Optimum set of $CHs = \{CH_1, CH_2, \dots, CH_n\}$;

b: Cluster member list $CMLISTs = \{CMLIST_1, CMLIST_2, \dots, CMLIST_n\}$;

1: BS broadcasts BS_ADV message to all l sensor nodes under its coverage area

2: **for** each sensor node SN_i under BS coverage region **do**

3: SN_i reply with RPL MSG to BS ;

4: **end for**

5: BS stores all sensor nodes SNs information

6: BS arranges SNs list information in descending order of residual energy and picks the top 10% of them ($n = \lfloor 0.1 \cdot l \rfloor$) as advanced nodes (CHs)

7: **for** ($k = 1, j \leq n, k++$) **do** // CHs list formation

8: $CH = CH \cup SN_k$;

9: **end for**

10: **for** each advanced node CH_k **do**

11: Calculate R_{AD_k} ; // Cluster radius according to (3)

12: CH_k broadcasts CLFM;

13: **end for**

14: **for** each sensor node SN_i **do** // non-CH nodes

15: **switch (state)**

16: **begin**

17: **case 1:** // SN_i in its CH_k cluster coverage radius area

18: **if** SN_i receives CLFM from CH_k **then** SN_i replies with CMJ; **end if, break;**

19: **case 2:** // SN_i in coverage radius area of multiple CHs nodes

20: **if** SN_i receives CLFM from multiple CHs **then**

```

21:  $SN_i$  sends CMJ to  $CH_k$  which has the maximum residual energy and smallest ID;
22: end if
23: break;
24: case 3: //  $SN_i$  in an intersecting area of the cluster radius of neighboring  $CHs$ 
nodes (autonomous nodes)
25: if  $SN_i$  in an intersecting area of multiple  $CHs$  then
26:  $SN_i$  sends CMJ to  $CH_k$  (either of adjacent  $CHs$ ) at random;
27: end if
28: break;
29: case 4: //  $SN_i$  does not receive CLFM from any  $CH_k$  (lone nodes)
30: if  $SN_i$  does not receive CLFM then
31:  $SN_i$  sends ASSIST message to neighboring nodes within its communication range;
32: Each neighboring node replies  $SN_i$  with RPM;
33:  $SN_i$  sends CMJ to the corresponding  $CH_k$  which has the maximum residual energy
and closest distance to  $BS$ .
34: end if
35: break;
36: end switch
37:  $CMLIST_k = CMLIST_k \cup SN_i$ ; //Add sensor node (non-CH)  $SN_i$  as cluster
member into  $CH_k$  member list
38: end for

```

4.3 MODRL-based clustering routing algorithm

The problem with routing is considered a multi-objective problem in which the optimal routing path should be determined based on several parameters in *IoT* – enabled *WSNs*. There are two phases in the proposed clustering routing algorithm: intra-cluster routing and inter-cluster routing.

4.3.1 Intra-cluster MODRL-based routing

Utilizing a *MODRL*-based framework, the sensor nodes (non-*CHs*) and advanced nodes (*CHs*) collaborate in order to optimize intra-cluster routing. The sensor node cluster members act as multi-agents for routing data packets to advanced nodes (*CHs*). Three objectives are considered

carefully to optimize intra-cluster routing, where maximizing network throughput is the first; reducing the network latency is the second; and extending the limited sensor battery lifespan is the third.

However, the *MODRL*-based are often conflicting with one another, thus maximizing one usually results in minimizing another. Hence, trade-offs among objectives must be taken into account in this challenging scenario. A Pareto optimality [64] frequently served as the basis for providing compromise options between the objectives and evaluating *MODRL* algorithms.

Intra-cluster routing aims to transfer data packets from multiple source nodes (sensors) to the destination (corresponding *CH*). If the sources are within the *CH* transmission range, the data is transferred directly to *CH*; otherwise, it is transferred indirectly through relaying of

multiple nearby nodes. A multi-objective Markov Decision Processes (MDP) optimization is used to represent intra-cluster routing.

An MDP model is a tuple (st_t, ac_t, pr, rw_t) , where $st_t \in ST$ is a finite set of states, $ac_t \in AC$ is a finite set of actions, $pr(st_{t+1}|st_t, ac_t) \in PR$ is the transition probability, and $rw_t(st_t, ac_t) \in RW$ is a reward function. In intra-cluster routing, the tuples of MDP are defined as follows.

State space: At any time t , the state space of agent i (sensor i) is denoted as $st_t^i = \{d^{AD}, ci^i, is^i\}$, where d^{AD} is the destination of the current generated packet of agent i towards its corresponding advanced node (CH), ci^i is the current information of agent i , and is^i is the information of agent i 's neighboring sensor nodes.

Action space: The action space of agent i at time t is denoted as $ac_t^i = \{AD^i, NA^i\}$, where AD^i is the corresponding advanced node to which the agent i belongs, and NA^i represents the set of neighbor nodes of agent i within its associated cluster.

Reward function: At any time t , the agent i receives a vector of three rewards for each conflicting objective. The three reward functions of each agent i are introduced under the given constraints as follows:

4.3.1.1 Throughput maximization

$$rw_{1,t}^i = (np_t^i \times ps) / td_t^i, \tag{4}$$

where np_t^i represents the number of successfully delivered packets to its corresponding advanced node AD^i , ps is the packet size, and td_t^i is the time it takes a sensor node to deliver a packet.

4.3.1.2 Delay minimization

$$rw_{2,t}^i = qt_t^i + tt_t^i, \tag{5}$$

where qt_t^i and tt_t^i stand for sensor node queuing time and sensor node transmission time, respectively.

4.3.1.3 Lifespan maximization

$$rw_{3,t}^i = re_t^i / (tr_t^i \times tp_t^i), \tag{6}$$

where re_t^i is the sensor node i 's remaining energy, tr_t^i is the sensor node i 's transmission rate, and tp_t^i is the sensor node i 's transmission power.

Thus, at any time t , the reward vector for agent i can be represented as

$$\vec{R}_t^i = [\max(rw_{1,t}^i), \min(rw_{2,t}^i), \max(rw_{3,t}^i)] \tag{7}$$

$$\begin{aligned} s.t. \text{ Constrain 1 : } & re_t^i \leq re_{th}, \\ \text{ Constrain 2 : } & tp_t^i \leq tp_{th}, \\ \text{ Constrain 3 : } & qt_t^i \leq qt_{th}, \end{aligned}$$

where re_{th} represents sensor node threshold energy, tp_{th} represents maximum sensor node transmitting power, and qt_{th} is the sensor node queuing time threshold.

Furthermore, there is a separate state-action value function (Q-value) for each objective $Q_j(st, ac)$, $j = 1$ to 3, and the vector of Q-values that includes $Q_j(st, ac)$ for each objective j may be defined as

$$\vec{Q}(st, ac) = [Q_1(st, ac), Q_2(st, ac), Q_3(st, ac)] \tag{8}$$

A policy in multi-objective MDP, denoted by Ψ , is the probability of choosing action $ac_t \in AC$ in state $st_t \in ST$. The policy Ψ can be improved by Q-value. So, knowing $Q(st, ac)$ enables to acquire the optimal policy via choosing the action having the highest Q-value. The estimation of $Q_\Psi(st, ac)$ function employing the Bellman equation [26] can be defined as

$$Q_\Psi(st, ac) = E \{R_t + \gamma Q_\Psi(st_{t+1}, ac_{t+1})\}, \tag{9}$$

where γ indicates the learning rate, R_t is the instant reward, and $E\{\cdot\}$ is the expectation.

Substituting for all objectives in the reward vector in (9) using the distribution function [65]

$$E_\Psi\{Q_\Psi(st, ac)\} = \sum_{t=0}^{\infty} D_\Psi(st_t) \Psi(ac_t|st_t) Q_\Psi(st_t, ac_t), \tag{10}$$

where D_Ψ indicates the distribution function. A Deep Q-network (DQN) is utilized to approximate $Q(st, ac)$ values. Thus, a separate DQN is used as an approximator for each $Q_j(st, ac)$, and multiple of DQNs operating in parallel would control such an agent. Figure 2 depicts the three DQNs multi-objective parallel architecture of our proposed model.

A DQN offers the approximation of the function $Q(st, ac; \theta)$ by the state-of-the-art in this field, where θ are the neural network's learnable parameters. There is a $Q_j(st, ac; \theta_j)$ function of DQN_j that is related to the objective j , as demonstrated in our proposed model utilizing multiple DQNs. Each DQN_j is optimized utilizing the following loss function:

$$Loss(\theta_j) = E\&[(Tr_j - Q_j(st_t, ac_t|\theta_j))^2], \tag{11}$$

where Tr_j is the target value and can be expressed as

$$Tr_j = \begin{cases} rw_j, & \text{if } ac_{t+1} = \varphi, \\ rw_j + \delta \max_{ac_{t+1}} Q_j(st_{t+1}, ac_{t+1}|\theta_j^-), & \text{otherwise,} \end{cases} \tag{12}$$

where δ represents the discount factor, $Q_j(st_t, ac_t | \theta_j)$ and $Q_j(st_t, ac_t | \theta_j^-)$ represent on-line network and target network, respectively.

In addition, the learning process is improved by utilizing experience replay, and for each DQN_j , actions, experienced rewards, and states are stored in a replay memory G_j . Then, a sample of prior experiences chosen evenly at random from the relevant replay memory G_j is used to train each DQN_j , during iterations. These obtained samples act as mini-batches to optimize gradient descent. The Rectifier Linear Unit (ReLU) is used as the activation function and adaptive moment as an optimizer (Adam) to minimize the loss function.

One of the most significant aspects of multi-objective optimization is the selection of actions based on a variety of objectives, which may be independent, conflicting, or complimentary. A popular approach to dealing with this important aspect is the transformation of multi-objective problems into a single objective using scalarization functions, which are utilized as a scoring technique for action choice strategies to acquire a combined score for action ac for various objectives j . The typical action selection methods of single-objective reinforcement learning, such as Boltzmann and ϵ -greedy, can therefore be employed in deciding which action to select given these scores.

A scalarization of $\vec{Q}(st, ac)$, considering $\vec{Q}(st, ac, j)$ -values, and a weight vector are applied for selecting the particular single action ac . The typical approach is to apply a linear scalarization function [66], so that, the scalarized Q-values can be obtained as

$$SQ(st, ac) = \sum_{j=1}^3 w_j \cdot \vec{Q}(st, ac, j), \quad (13)$$

where $w_j \in [0, 1]$ is the weighted coefficient of each objective j , $\sum_{j=1}^3 w_j = 1$, and $\vec{Q}(st, ac, j)$ denotes the DQN_j function of each objective j .

The Q-values are normalized (re-scaled) using the min-max scaling function to guarantee that the values with various scales have the same impact and accurately represent votes for certain actions. The scaling function is as follows.

$$scale(\vec{q}_j) = \frac{\vec{q}_j - \min(\vec{q}_j)}{\max(\vec{q}_j) - \min(\vec{q}_j)} \quad (14)$$

After normalization, Eq. (13) can be defined as:

$$SQ(st, ac) = \sum_{j=1}^3 w_j \cdot scale(\vec{q}_j) \quad (15)$$

Then, the action ac' correspondent to the maximal value of scaled $SQ(st, ac)$ is regarded the greedy action in state st , and evaluated as

$$greedy_{ac'}(st) = \max_{ac'} SQ(st, ac') \quad (16)$$

Algorithm 2 describes the learning process of the proposed DQN architecture for multiple objectives optimization.

Algorithm 2 Multi-objective *DQN* architecture

```

1: Input  $\delta$  : discount factor , replay memory:  $G$ 
2: Initialize  $\bar{Q}(st, ac, j)$  with random parameters weights  $\theta$ 
3: for  $t = 1$  to  $T$  do
4:  $SQlist = \{\}$ 
5: for each action  $ac_i \in AC$  do
6:  $\bar{Q} \leftarrow \{Q(st, ac_i, j_1), Q(st, ac_i, j_2), Q(st, ac_i, j_3)\}$ 
7:  $SQ(st, ac) \leftarrow \text{scalarize}(\bar{Q})$  // Q-values scalarization
8: Append  $SQ(st, ac)$  to  $SQlist$ 
9: end for
10: Choose action  $ac_t$  using  $\varepsilon$  – greedy( $SQlist$ )
11: Take the action  $ac_t$ 
12: Observe the reward vector  $\bar{R}_t$  and the new next state  $st_{t+1}$ 
13: Save tuples  $(st_t, ac_t, rw_t, st_t)$  in  $G$ 
14: Sample random minibatch of tuples  $(st_{t+1}, ac_{t+1}, rw_{t+1}, st_{t+1})$  from  $G$ 
15: Calculate  $Tr$ 
16: Calculate  $Loss(\theta)$ 
17: In Adam Optimizer, train Q-networks
18: Update  $\theta^-$ 
19: end for
20: end process

```

4.3.2 Inter-cluster MODRL-based routing

In inter-cluster routing, advanced nodes (*CHs*) act as multi-agents for routing data packets toward the base station (*BS*) using a *MODRL* framework. Three objectives are considered carefully to optimize inter-cluster routing, where maximizing network throughput is the first; reducing the network latency is the second; and minimizing the traffic load upon advanced nodes (*CHs*) is the third. A multi-objective *MDP* is used to resolve these conflicting objectives.

Inter-cluster routing aims to transfer data packets aggregated at *CHs* from their sensor node members to the base station (*BS*) while maintaining a balanced traffic load upon them. If a *CH* is connected to the *BS*, the aggregated data is transferred there directly; otherwise, it is transferred indirectly through the relaying of other *CH* nodes. The inter-

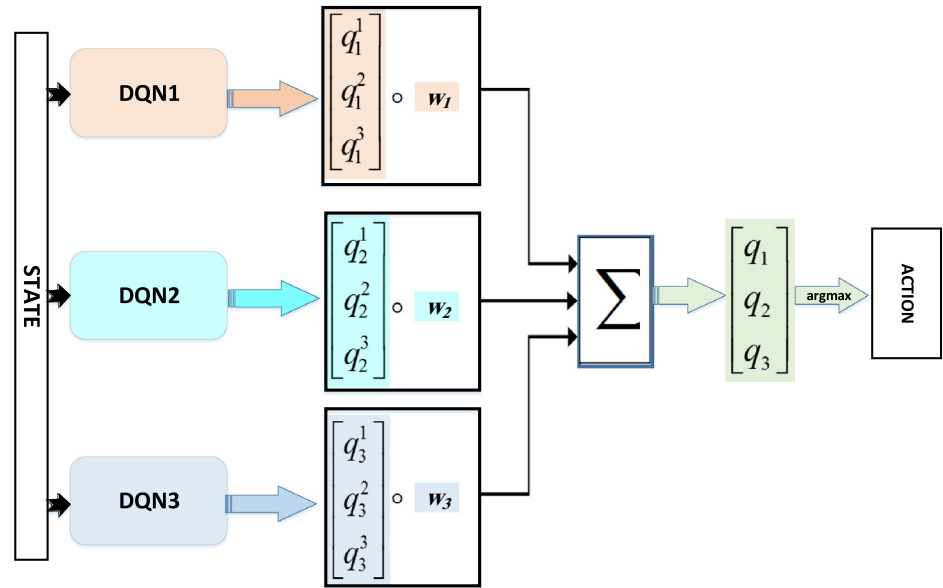
cluster routing is represented as a multi-objective Markov Decision Processes (*MDP*) optimization model. The following defines the tuples of *MDP* for inter-cluster routing.

State space: At any time t , the state space of agent k (CH_k) is denoted as $st_t^k = \{d^{BS}, ci^k, is^k\}$, where d^{BS} denotes the base station's destination of aggregated packets from an agent k (CH_k), ci^k is the current information of agent k , and is^k is the information of agent k 's neighboring *CHs* nodes.

Action space: The action space of agent k at time t is denoted as $ac_t^k = \{BS, NA^k\}$, where *BS* is the base station's destination, and NA^k represents the set of *CHs* neighbor nodes of agent k .

Reward function: At any time t , the agent k receives a vector of three rewards for each conflicting objective. The three reward functions of each agent k are introduced under a given constraint as follows.

Fig. 2 Three *DQNs* parallel architecture



4.3.2.1 Throughput maximization

$$rw_{1,t}^k = (np_t^k \times ps) / td_t^k, \tag{17}$$

where np_t^k represents the number of successfully delivered packets to *BS*, ps is the packet size, and td_t^k is the time it takes CH_k to deliver a packet.

4.3.2.2 Delay minimization

$$rw_{2,t}^k = qt_t^k + tt_t^k, \tag{18}$$

where qt_t^k and tt_t^k stand for CH_k queuing time and CH_k transmission time, respectively.

4.3.2.3 Traffic load minimization

$$rw_{3,t}^k = re_t^k \times ps \times sp_t^k, \tag{19}$$

where re_t^k is the advanced node k 's remaining energy, ps is the packet size, and sp_t^k represents the number of successfully serviced packets by CH_k .

Thus, at any time t , the reward vector for agent k can be represented as

$$\vec{R}_t^k = [\max(rw_{1,t}^k), \min(rw_{2,t}^k), \min(rw_{3,t}^k)] \tag{20}$$

s.t. *Constrain 1* : $re_t^k \leq re_{th}^{AD}$,

Constrain 2 : $qt_t^k \leq qt_{th}^{AD}$,

where re_{th}^{AD} represents advanced node threshold energy, and qt_{th}^{AD} is the advanced node queuing time threshold.

Our proposed model for inter-cluster *MODRL*-based routing is shown in Fig. 2 as a three *DQNs* multi-objective parallel architecture, and algorithm 2 describes the learning

process of the proposed *DQN* architecture for multiple objectives optimization.

4.4 Total energy consumption

The *BS* calculates the maximum energy consumed by *CHs* based on their inter-cluster and intra-cluster traffic loads after unequal clustering formation and *MODRL*-based clustering routing construction. The energy consumed by any non-*CH* node (cluster member) is represented as

$$E_{cm}(j) = l \times E_{elec}^{tx} + l \times \epsilon_{fs} \times d_{ch}^2(j), \tag{22}$$

where $d_{ch}(j)$ is the distance between cluster member j and its corresponding CH_j . The total energy consumption of CH_k owing to intra-cluster activity is represented as

$$E_{CH-intra}(k) = N_{CH}(k) \times E_{Rx} + N_{CH}(k) \times E_{DA} + E_{Tx}(k), \tag{23}$$

where $N_{CH}(k)$ is the number of cluster members of CH_k , and $E_{Tx}(k)$ is the energy dissipated by CH_k to transmit the aggregated data toward other *CH* or *BS*. In addition, E_{DA} and E_{Rx} represent the energy dissipated by CH_k due to data aggregation and data reception, respectively.

Additionally, for inter-cluster traffic load, CH_k serves as a relay node. Therefore, the total energy consumed by CH_k as a result of inter-cluster activity can be presented by

$$E_{CH-inter}(k) = RL_{CH}(k) \times E_{Rx} + RL_{CH}(k) \times E_{Tx}(k), \tag{24}$$

where $RL_{CH}(k)$ is the number of packets incoming from other *CHs*.

5 Complexity

In this section, the complexity analysis is investigated in terms of message complexity and computational time complexity, as well as network lifespan is estimated, in order to demonstrate the effectiveness of the proposed routing schema.

5.1 Complexity of clustering mechanism

Message complexity: It takes initial l messages to broadcast BS_ADV to all sensor nodes. Then, advanced nodes reply with n messages. Cluster radius allocation requires n messages by CHs. Then, it takes $l - n$ non-CH messages for joining these n CHs. Thus, the message complexity can be expressed as $O(l + n + n + (l - n))$. As $n \ll l$, then the message complexity $\approx O(l)$.

Time complexity: $O(l)$ time is taken to broadcast BS_ADV message and $O(n)$ time is taken by advanced nodes to reply. Cluster radius allocation takes $O(n)$ time by CHs and $O(l - n)$ time is taken by non-CHs to join these n CHs. Therefore, the time complexity $\approx O(l)$.

5.2 Complexity of DRL-based intelligent clustering data routing schema

5.2.1 Message complexity

The n CHs receive $l - n$ non-CH messages in intra-cluster routing, while BS receives n messages from n CHs in inter-cluster routing. Thus, the message complexity is expressed as $O(l - n + n) \approx O(l)$.

5.2.2 Time complexity

The routing schema proposes three DQNs having the same architecture and working in parallel. We consider the DQN architecture which takes up less memory and faster routing schema execution. The routing schema utilizes Convolutional Neural Network (CNN) layers. The CNN consists of two Depthwise Separable Convolution (DSC) layers and three Fully Connected (FC) layers. The complexity is computed in terms of multiply-accumulate operations (MACCs).

The total MACCs for DSC layers is given by:

$$C_{1in} \times H_{1out} \times W_{1out} \times (K_1 \times K_1 + C_{1out}) + C_{2in} \times H_{2out} \times W_{2out} \times (K_2 \times K_2 + C_{2out}), \quad (25)$$

where $C_{in} \times H_{out} \times W_{out}$ is the feature map size, $K \times K$ represents kernel size, and C_{out} denotes the number of

convolution kernels. The computation performed by the FC layer is given by:

$$y = \text{matmul}(x, W) + b, \quad (26)$$

where W is $I \times J$ matrix holding the weights of the layer, x is a vector of I input values, b represents a vector of J bias values that are also included, and y is also a vector of size J containing the output values computed by the FC layer. Then, the total MACCs for FC layers can be represented as

$$(I_1 \times J_1) + (I_2 \times J_2) + (I_3 \times J_3) \quad (27)$$

The activation function's computational time is so brief that it can be disregarded. Therefore, the time complexity of the proposed DRL-based intelligent clustering data routing schema is given by:

$$O \left(\begin{array}{l} C_{1in} \times H_{1out} \times W_{1out} \times (K_1 \times K_1 + C_{1out}) \\ + C_{2in} \times H_{2out} \times W_{2out} \times (K_2 \times K_2 + C_{2out}) \\ + ((I_1 \times J_1) + (I_2 \times J_2) + (I_3 \times J_3)) \end{array} \right) (l - n + n) \approx O \left(\begin{array}{l} C_{1in} \times H_{1out} \times W_{1out} \times (K_1 \times K_1 + C_{1out}) \\ + C_{2in} \times H_{2out} \times W_{2out} \times (K_2 \times K_2 + C_{2out}) \end{array} \right) l \quad (28)$$

5.3 Estimated network lifespan of the proposed intelligent clustering routing schema

Network Lifespan (NLS) is defined as the amount of time a network is alive for data collecting up until the last node dies within the network due to energy consumption. Let TE_{in} represent the network's total initial energy. Additionally, let TE_{ex} represent the total energy expended by all sensor and advanced nodes during the data processing, clustering routing process, clustering formation, and other activity of the network.

Therefore, NLS can be defined as the lowest ratio of the network's overall initial energy to its overall energy consumption. Therefore, NLS can be represented as

$$NLS = \min (TE_{in}/TE_{ex}) \quad (29)$$

6 Performance evaluation

In this section, we evaluate the performance of the proposed multi-objective clustering routing schema through simulations under various system parameters. The performance of the proposed routing schema is compared with

RLBEEP [58] and EER-RL [60]. The simulation is carried out within an area of $100 \times 100 \text{ m}^2$ network. The sensor nodes are distributed randomly over this area.

Two simulation scenarios are carried out to investigate the performance evaluation of our proposed schema. The scenarios for the simulation are described in detail below.

- *Simulation scenario 1:* In this scenario, there are 200 sensor nodes deployed randomly in the network size area, and *BS* is located at the center of the monitoring area, i.e., (50, 50).
- *Simulation scenario 2:* In this scenario, there are 300 sensor nodes deployed randomly in the network size area, and *BS* is located outside the monitoring area, i.e., (250, 200).

Tables 1 and 2 show the parameters of the simulation. The results for each simulation are the mean of 20 runs with various seed values. The effectiveness of the three routing protocols is compared uniformly.

6.1 Performance evaluation metrics

The effectiveness of the proposed routing schema is evaluated in terms of the following metrics:

6.1.1 Energy efficiency ,(EE)

The number of packets delivered per unit energy consumed, which is expressed as

$$EE = N_{PD}/TE_{ex}, \quad (30)$$

where TE_{ex} is the total energy consumption, and N_{PD} represents the number of delivery packets.

$$TE_{ex} = \sum_{j=1}^l (E_{in}^j - E_{re}^j), \quad (31)$$

where E_{in}^j represents the initial energy of node j , and this value is consistent for all nodes in the network, and l stands for the total number of nodes within the network. E_{re}^j , on the other hand, corresponds to the residual energy of a specific node j .

6.1.2 Delivered packets over time

The number of delivered data packets to the *BS* over time.

$$Delivery\ packets(t) = \sum_{j=1}^l p_j(t), \quad (32)$$

where l stands for the total number of nodes within the network, and $p_j(t)$ is the number of packets successfully delivered by a node j to the *BS* during time t .

Table 1 System simulation parameters

Parameter	Value
Network size	$100 \times 100 \text{ m}^2$
Number of sensor nodes	200, 300
d_o	87 m
Data packet size	512 bits
E_{DA}	5 nJ/bit/signal
E_{elec}	50 nJ/bit
ϵ_{fs}	10 pJ/bit/m ²
ϵ_{amp}	0.0013 pJ/bit/m ⁴
Initial energy	2 J

Table 2 Simulation parameters for the DQN model

Parameter	Value
Replay memory capacity	2000
The size of Mini-batch	64
Episodes EP	200
Steps T	400
Iterations Z	300
Discount factor	0.9
Learning rate	0.01
Maximum ϵ -greedy	0.9
Activation function	ReLU
Optimizer	Adam

6.1.3 End-to-end delay

The time it takes for data packets to arrive at the *BS*.

$$End - to - end\ delay = \sum_{j=1}^l (qt_j + tt_j + pt_j + dt_j), \quad (33)$$

where l stands for the total number of nodes within the network, qt_j , tt_j , pt_j , and dt_j stand for sensor node queuing delay, sensor node transmission delay, sensor node processing delay, and sensor node propagation delay, respectively.

6.1.4 Alive nodes over time

Number of alive nodes in the network over time.

$$Alive\ nodes(t) = \sum_{j=1}^l \delta(E_{re}^j(t) > E_{TH}), \quad (34)$$

where l stands for the total number of nodes within the network, $E_{re}^j(t)$ represents the residual energy of node j at

time t , E_{TH} is the threshold energy that determines whether a node is considered alive or dead, and $\delta(\cdot)$ is a mathematical function that returns 1 if the condition inside the parentheses is true and 0 if it's false.

Therefore, as described in Eq. (34), the summation involves nodes ranging from 1 to l , and it evaluates whether each node's residual energy $E_{re}^j(t)$ at time t exceeds the threshold energy E_{TH} . If the condition is satisfied (i.e., the node is alive), it contributes a value of 1 to the cumulative sum; otherwise, it contributes 0.

6.1.5 Network lifespan

The time until the First Node Exhausted (*FNE*), or until Half of Nodes Exhausted (*HNE*), or until the Last Node Exhausted (*LNE*).

The calculation of *FNE* involves identifying the minimum time t at which the energy of a node drops below or equals the threshold energy (E_{TH}), i.e., the point in time when the first sensor node (j) within the network exhausts its energy reserves.

$$FNE = \min_t \left(\sum_{j=1}^l \delta(E_{re}^j(t) \leq E_{TH}) \right), \quad (35)$$

where \sum denotes the loop counter involves nodes ranging from 1 to l .

The calculation of *LNE* involves identifying the maximum time t at which the energy of a node drops below or equals the threshold energy (E_{TH}), i.e., the point in time when the last sensor node (j) within the network exhausts its energy reserves.

$$LNE = \max_t \left(\sum_{j=1}^l \delta(E_{re}^j(t) \leq E_{TH}) \right), \quad (36)$$

where \sum denotes the loop counter involves nodes ranging from 1 to l .

HNE is identified by locating the time t when the number of nodes with energy less than or equals to the threshold energy (E_{TH}) is equal to half of the total nodes, i.e., the point in time when half of the nodes in the network are exhausted.

$$HNE = \min_t \left(\sum_{j=1}^l \delta(E_{re}^j(t) \leq E_{TH}) = \frac{l}{2} \right). \quad (37)$$

Therefore, as described in Eq. (37), the summation involves nodes ranging from 1 to l , and it evaluates whether each node's residual energy $E_{re}^j(t)$ at time t drops below or equals the threshold energy E_{TH} . If the condition is satisfied (i.e., the node is exhausted), it contributes a value of 1 to the cumulative sum; otherwise, it contributes 0.

6.1.6 Energy balancing

The amount of average energy consumption of *CHs*.

$$\text{Energy balancing } (t) = \frac{\sum_{k=1}^n E_{CH}^k(t)}{n}, \quad (38)$$

where n stands for the total number of cluster heads within the network, and $E_{CH}^k(t) = E_{CH-intra}^k(t) + E_{CH-inter}^k(t)$ is the energy consumed by the cluster head k during time t .

6.2 Performance evaluation results

In this section, the performance evaluation results for the proposed schema compared to the other two existing schemes are presented by the metrics for performance evaluation.

6.2.1 Energy efficiency ,(EE)

Figure 3a and b show the number of delivered packets to the *BS* versus energy consumption in scenario 1 and scenario 2 respectively. It is shown from Fig. 3a and b that the number of delivered packets in the three routing schemes increases as the energy consumption increases, and which in the proposed schema is more than that in RLBEED and EER-RL.

In scenario 1, the proposed schema improves *EE* by 37 and 84% as compared to EER-RL and RLBEED schemes, respectively. In addition, it is demonstrated that the proposed schema outperforms EER-RL and RLBEED, and improves 39.2 and 86.60% more *EE* than both of them respectively, in scenario 2.

The reasoning is that advanced nodes with high energy serve as a *CHs* and an intelligent *DRL* – based algorithm determines the optimal path route for data routing. Moreover, employing a multi-objective intelligent strategy gives more opportunities for increasing the delivery of packets to the base station, in which the multi-hop data routing path is used to carry out the inter-cluster routing process among the *CHs*.

6.2.2 Delivered packets over time

Figure 4a and b show the number of delivered packets to the *BS* versus time in scenario 1 and scenario 2 respectively. It is shown from Fig. 4a and b that the number of delivered packets in the three routing schemes increases as the time increases, and which in the proposed schema is higher than that in RLBEED and EER-RL.

According to Fig. 4a for scenario 1, the proposed schema delivers 43% more packets than EER-RL and 89% more packets than RLBEED. Additionally, it is shown from

Fig. 4b that the proposed schema outperforms EER-RL and RLBEET, and increases 41.6 and 87.3% more packet delivery than both of them respectively, in scenario 2.

The reasoning is due to identifying the optimal routing path utilizing an intelligent *DRL* – based algorithm based on multiple objectives such as packet size, remaining energy, traffic rate, and queuing time, and thus minimizes traffic congestion, which in turn reduces packet delivery loss.

6.2.3 End-to-end delay

Figure 5a and b depict the end-to-end delay of delivered packets versus node density in scenario 1 and scenario 2 respectively. It is shown from Fig. 5a and b that the end-to-end delay in the three routing schemes increases as the node density increases, and which in the proposed schema is lower than that in RLBEET and EER-RL.

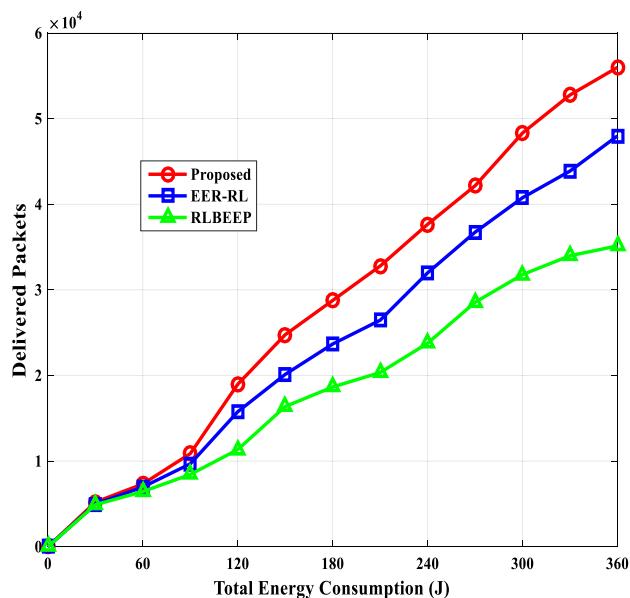
In scenario 1, it is observed that the proposed schema can decrease the end-to-end delay by approximately 41.46 and 51.23% compared to EER-RL and RLBEET, respectively. Additionally, it is demonstrated that the proposed schema reduces end-to-end delay by up to 44.5% compared to EER-RL and up to 53.6% compared to RLBEET in scenario 2.

The reasoning is due to the intelligent *DRL* – based algorithm that assigns an appropriate inter-cluster relay traffic load to a *CH*. As a result, there is less traffic congestion, which decreases queuing time and lessens the end-to-end delay for delivered packets.

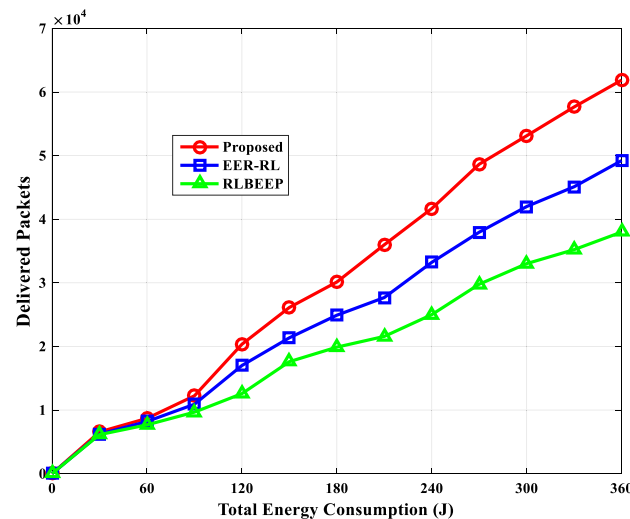
6.2.4 Alive nodes over time

Figure 6a and b show the percentage of alive nodes within the network over time in scenario 1 and scenario 2 respectively. It is shown from Fig. 6a and b that the three routing schemes experience a decline in the percentage of alive nodes as time goes on, which in the proposed schema is higher than that in RLBEET and EER-RL.

In scenario 1, the proposed schema enhances the number of alive nodes by 68.1 and 81.2% as compared to EER-RL and RLBEET schemes, respectively. In addition, it is demonstrated that the proposed schema outperforms EER-RL and RLBEET, and improves 71.2 and 83.4% more alive nodes than both of them respectively, in scenario 2. The reason is that the proposed schema addresses the hot spot problem by adopting an unequal cluster mechanism that balances the entire network’s load, which in turn promotes network stability.



(a) Scenario 1



(b) Scenario 2

Fig. 3 Energy efficiency (*EE*)

6.2.5 Network lifespan

Figure 7a and b depict the lifespan of the network as represented by the time until *FNE*, or until *HNE*, or until *LNE*, in scenario 1 and scenario 2 respectively. It is shown from Fig. 6a and b that the proposed schema outperforms the other routing schemes across all lifespan metrics (*FNE*, *HNE* and *LNE*).

In scenario 1, under *FNE* criterion the proposed schema improves the network lifespan by 26.4 and 58.8% as compared to EER-RL and RLBEET schemes, respectively. On the other hand, the proposed schema outperforms EER-

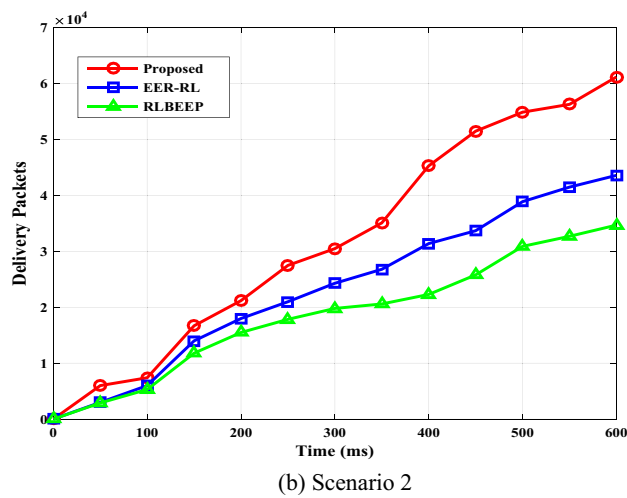
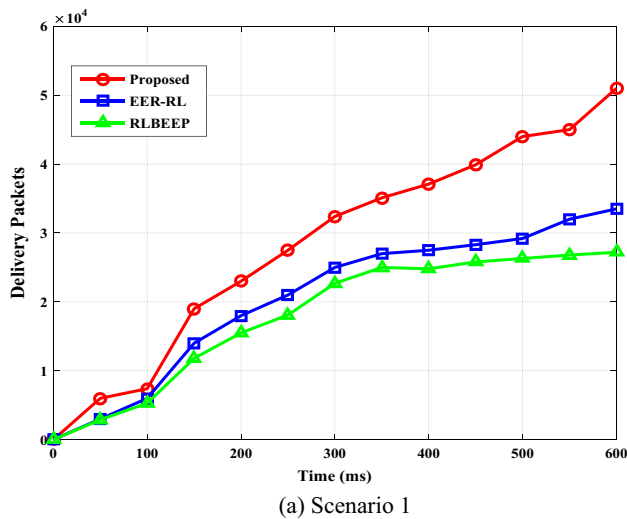


Fig. 4 Delivered packets over time

RL and RLBEEP schemes in terms of *HNE* by 23.8 and 50%, respectively.

Similarly, under the *LNE* criterion, the proposed schema outperforms the EER-RL and RLBEEP schemes in terms of network lifespan by 15.24 and 42.37%, respectively. The reason is the effective load balancing among the *CHs*, as well as *non-CH* nodes.

Additionally, the proposed schema outperforms EER-RL and RLBEEP schemes across all lifespan metrics in scenario 2. The reason is due to the intelligent *DRL* – based algorithm substantially decreases message overhead throughout the data routing stage which makes networks less energy-intensive. Thus, all sensor nodes’ energy consumption is decreased in both inter-cluster and intra-cluster environments, which greatly improves their lifespan in dense scenarios.

6.2.6 Energy balancing

This section examines the energy balance (uniform energy consumption) amongst *CHs*, as shown by estimating the amount of average energy consumption of *CHs*. Figure 8a and b show the amount of average energy consumption of *CHs* within the network over time in scenario 1 and scenario 2 respectively. Compared with EER-RL and RLBEEP schemes, the proposed schema maintains a roughly equal amounts of average energy consumption of *CHs* in both scenarios.

The reason is due to the intelligent *DRL* – based algorithm elects the optimal relay *CH* nodes to reduce and equally balance the intra-cluster and inter-cluster traffic load amongst the *CHs*.

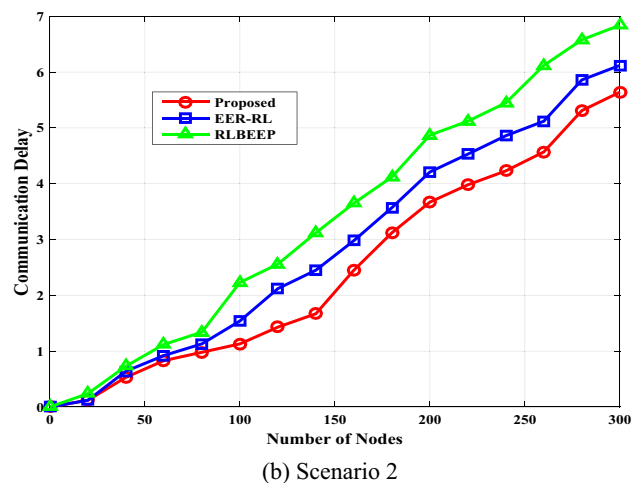
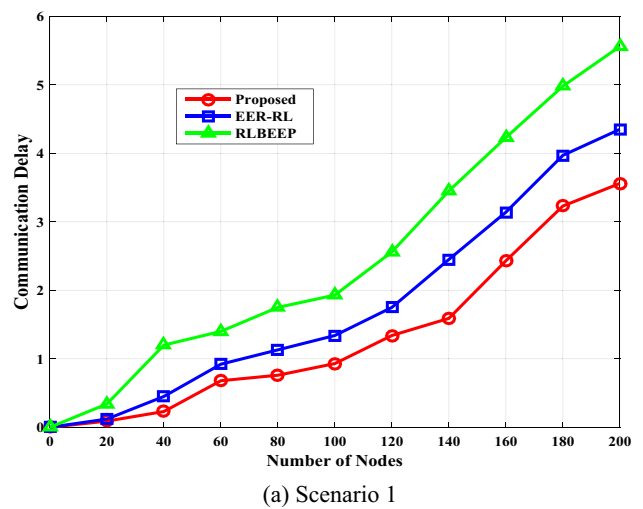


Fig. 5 End-to-end delay

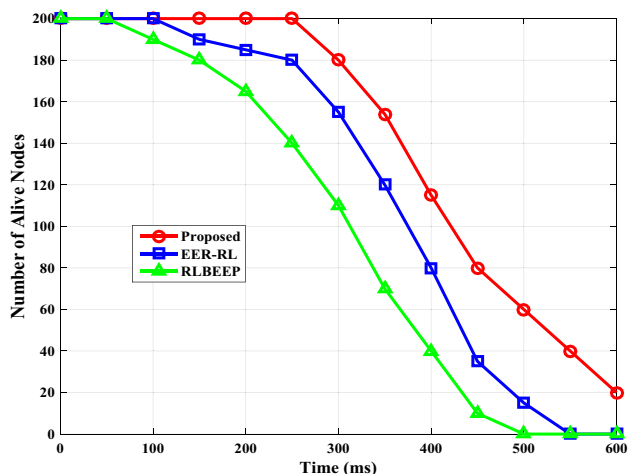
7 Conclusions

A multi-objective intelligent clustering routing schema is proposed for *IoT* – enabled *WSNs* utilizing Deep Reinforcement Learning, in this paper. The proposed schema involves an innovative unequal clustering mechanism in which an advanced node serves as a cluster head and keeps track of the deployment and management of sensor nodes to prohibit the network from dying prematurely. Energy consumption balancing is achieved to prevent network partition and hot spot problems.

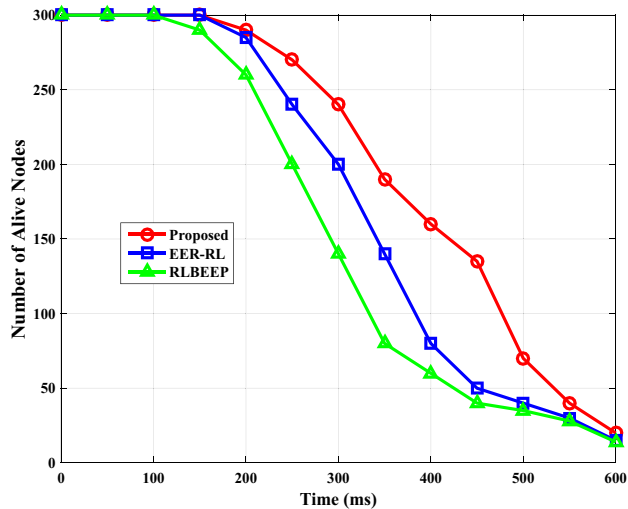
The proposed schema considers various objective parameters for inter-cluster routing and intra-cluster

routing that dramatically improve both network performance and network lifespan. Furthermore, this study analyzes the proposed schema’s message and time complexity as well.

In addition, comprehensive simulations under different system parameters have been carried out to demonstrate the superior performance of our proposed intelligent routing schema in terms of energy efficiency, delivered packets, end-to-end delay, alive nodes, energy balancing, and network lifespan compared with the other two existing approaches. As a future work, a fault tolerance mechanism will be involved in our proposed schema to improve its reliability.

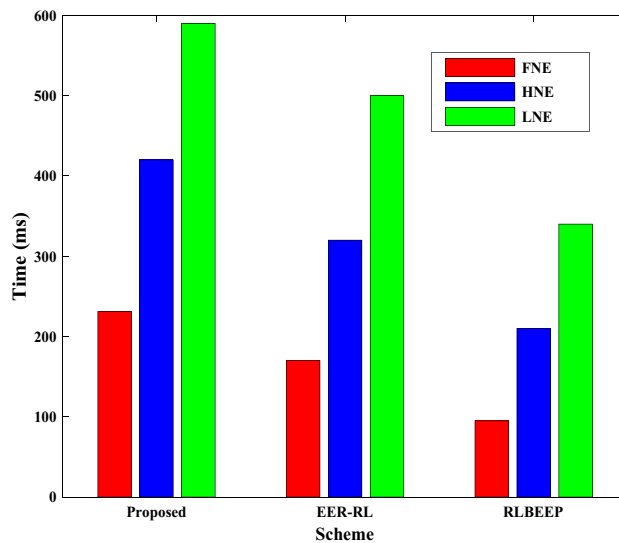


(a) Scenario 1

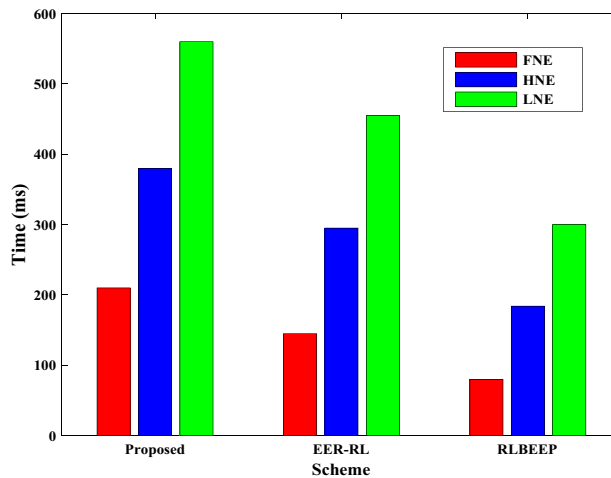


(b) Scenario 2

Fig. 6 Alive nodes over time



(a) Scenario 1



(b) Scenario 2

Fig. 7 Network lifespan

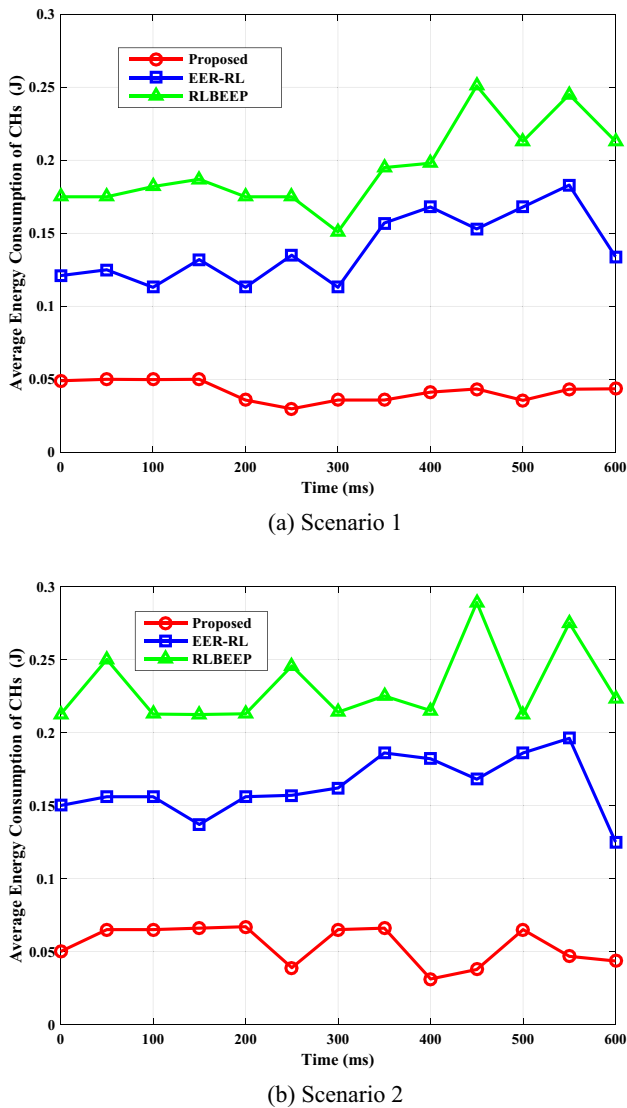


Fig. 8 Energy balancing

Author contributions W. K. proposed the protocol and made all the analysis. S. S. conducted and performed the simulations. W. K. read and approved the final manuscript.

Declarations

Competing interests The authors declare no competing interests.

References

1. Aanchal, S.K., Omprakash, K., Abdul, H.A.: Green computing for wireless sensor networks: optimization and Huffman coding approach. *Peer-to-Peer Netw. Appl.* **10**(3), 592–609 (2017)
2. Najm, A., Ismail, M., Rahm, T., Al Razak, A.: Wireless implementation selection in higher institution learning environment. *J. Theor. Appl. Inform. Technol.* **67**, 477–484 (2014)
3. Rahem, A.A.T., Ismail, M., Najm, I.A., Balfaqih, M.: Topology sense and graph-based TSG: efficient wireless ad hoc routing protocol for WANET. *Telecommun. Syst.* **65**(4), 739–754 (2017)
4. Aalsalem, M.Y., Khan, W.Z., Gharibi, W., Khan, M.K., Arshad, Q.: Wireless sensor networks in oil and gas industry: recent advances, taxonomy, requirements, and open challenges. *J. Netw. Comput. Appl.* **113**, 87–97 (2018)
5. Ball, M.G., Qela, B., Wesolkowski, S.: A review of the use of computational intelligence in the design of military surveillance networks. *Stud. Comput. Intell.* **621**, 663–693 (2016)
6. Kumar, V., Kumar, S., AlShboul, R., Aggarwal, G., Kaiwartya, O., Khasawneh, A., Lloret, J., Al-Khasawneh, M.: Grouping and sponsoring centric green coverage model for internet of things. *Sensors* **21**(12), 3948 (2021)
7. Zualkernan, I.A., Al-ali, A.R., Jabbar, M.A., Zabalawi, I., Wasfy, A.: InfoPods: ZigBee-based remote information monitoring devices for smart homes. *IEEE Trans. Consumer Electron.* **55**(3), 1221–1226 (2009)
8. Rehman, A., Haseeb, K., Saba, T., Lloret, J., Sendra, S.: An optimization model with network edges for multimedia sensors using artificial intelligence of things. *Sensors* **21**(21), 7103 (2021)
9. Lazarescu, M.T.: Design of a WSN platform for long-term environmental monitoring for IoT applications. *IEEE J. Emerg. Select. Topics Circ. Syst.* **3**(1), 45–54 (2013)
10. Jindal, A., Agarwal, V., Chanak, P.: Emergency evacuation system for clogging-free and shortest-safe path navigation with IoT-enabled WSNs. *IEEE Internet Things J.* **9**(13), 10424–10433 (2022). <https://doi.org/10.1109/JIOT.2021.3123189>
11. Al Ameen, M., Liu, J., Kwak, K.: Security and privacy issues in wireless sensor networks for healthcare applications. *J. Med. Syst.* **36**(1), 93–101 (2012)
12. Mikhaylov, K., Tervonen, J., Heikkila, J., Kansakoski, J.: Wireless sensor networks in an industrial environment: real-life evaluation results. In: *The 2nd baltic congress on future internet communications (BCFIC)*, pp. 1–7 (2012)
13. Oladimeji, D., Khushi, G., Nuri, A.K., Kubra, G., Linqiang, G., Fan, L.: Smart transportation: an overview of technologies and applications. *Sensors* **23**(8), 3880 (2023). <https://doi.org/10.3390/s23083880>
14. Gutam, G., Donta, P.K., Annavarapu, C.S.R., Hu, Y.-C.: Optimal rendezvous points selection and mobile sink trajectory construction for data collection in WSNs. *J. Ambient. Intell. Humaniz. Comput.* **14**, 7147–7158 (2023)
15. Mehto, A., Tapaswi, S., Pattanaik, K.K.: Optimal rendezvous points selection to reliably acquire data from wireless sensor networks using mobile sink. *Computing* **103**(4), 707–733 (2021)
16. Rani, R., Kumar, S., Kaiwartya, O., Khasawneh, A., Lloret, J., Al-Khasawneh, M., Mahmoud, M., Alarood, A.: Towards green computing oriented security: a lightweight postquantum signature for IoE. *Sensors* **21**(5), 1883 (2021)
17. Shanthi, M., RamaDevi, E.: A cluster based routing protocol in wireless sensor network for energy consumption. *Int. J. Adv. Netw. Appl.* **5**(4), 2015–2020 (2014)
18. Guo, W., Zhu, W., Yu, Z., Wang, J., Guo, B.: A survey of task allocation: contrastive perspectives from wireless sensor networks and mobile crowdsensing. *IEEE Access* **7**, 78406–78420 (2019). <https://doi.org/10.1109/ACCESS.2019.2896226>
19. Tyagi, S., Kumar, N.: A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks. *J. Netw. Comput. Appl.* **36**, 623–645 (2013)
20. Tanwara, S., Kumarb, N., Rodrigues, J.: A systematic review on heterogeneous routing protocols for wireless sensor network. *J. Netw. Comput. Appl.* **53**, 39–56 (2015)
21. Zeb, A., Islam, A.M., Zareei, M., Al Mamooun, I., Mansoor, N., Baharun, S., Katayama, Y., Komaki, S.: Clustering analysis in wireless sensor networks: the ambit of performance metrics and

- schemes taxonomy. *Int. J. Distrib. Sensor Netw.* **12**(7), 4979142 (2016)
22. Liu, A.-F., Wu, X.-Y., Chen, Z.-G., Gui, W.-H.: Research on the energy hole problem based on unequal cluster-radius for wireless sensor networks. *Comput. Commun.* **33**(3), 302–321 (2010)
 23. Altamimi, A.B., Ramadan, R.A.: Towards internet of things modeling: a gateway approach. *Complex Adapt. Syst. Model.* **4**(1), 1–11 (2016)
 24. Fraile, F., Tagawa, T., Poler, R., Ortiz, A.: Trustworthy industrial iot gateways for interoperability platforms and ecosystems. *IEEE Internet Things J.* **5**(6), 4506–4514 (2018)
 25. Alarifi, A., Tolba, A.: Optimizing the network energy of cloud assisted internet of things by using the adaptive neural learning approach in wireless sensor networks. *Comput. Ind.* **106**, 133–141 (2019)
 26. Sutton, R.S., Barto, A.G.: *Reinforcement learning: an introduction*. MIT Press, Cambridge (2018)
 27. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013)
 28. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. *Nature* **518**(7540), 529–533 (2015)
 29. Li, M., Yu, F.R., Si, P., Wu, W., Zhang, Y.: Resource optimization for delay-tolerant data in blockchain-enabled iot with edge computing: a deep reinforcement learning approach. *IEEE Internet of Things J.* **7**, 9399 (2020)
 30. Xu, Z., Wang, Y., Tang, J., Wang, J., Gursoy, M.C.: A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs. In: 2017 IEEE International Conference on Communications (ICC). pp. 1–6. Paris, France, (2017)
 31. Mao, H., Netravali, R., Alizadeh, M.: Neural adaptive video streaming with pensieve. In: Proceedings of the Conference of the ACM Special Interest Group on Data Communication, pp. 197–210 (2017)
 32. Stampa, G., Arias, M., Sánchez-Charles, D., Muntés-Mulero, V., Cabellos, A.: A deep-reinforcement learning approach for software-defined networking routing optimization. (2017)
 33. Boyan, J.A., Littman, M.L.: Packet routing in dynamically changing networks: a reinforcement learning approach. In: Smith, J. (ed.) *Advances in Neural Information Processing Systems*, pp. 671–678. Morgan Kaufmann, San Francisco (1994)
 34. Zhang, Y., Huang, Q.: A learning-based adaptive routing tree for wireless sensor networks. *J. Commun.* **1**(2), 12–21 (2006)
 35. Wang, P., Wang, T.: Adaptive routing for sensor networks using reinforcement learning. In: Proceedings of the IEEE International Conference on Computer & Information Technology, Seoul, South Korea, 20–22, pp. 219–224. IEEE, New York (2006)
 36. Forster, A., Murphy, A.L.: FROMS: Feedback routing for optimizing multiple sinks in WSN with reinforcement learning. In: 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, (pp. 371–376). IEEE (2007)
 37. Förster, A., Murphy, A.L.: Balancing energy expenditure in WSNs through reinforcement learning: a study. In: Proceedings of the 1st International Workshop on Energy in Wireless Sensor Networks (WEWSN), Santorini Island, Greece, (p. 7). (2008)
 38. Hu, T., Fei, Y.: QELAR: a machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks. *IEEE Trans. Mobile Comput.* **9**(6), 796–809 (2010)
 39. Razzaque, M.A., Ahmed, M.H.U., Hong, C.S., Lee, S.: QoS-aware distributed adaptive cooperative routing in wireless sensor networks. *Ad Hoc Netw.* **19**, 28–42 (2014)
 40. Renold, A., Chandrakala, S.: MRL-SCSO: multi-agent reinforcement learning-based self-configuration and self-optimization protocol for unattended wireless sensor networks. *Wireless Pers. Commun.* **96**, 5061–5079 (2016)
 41. Guo, W., Yan, C., Lu, T.: Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing. *Int. J. Distrib. Sensor Netw.* (2019). <https://doi.org/10.1177/1550147719833541>
 42. Yun, W.-K., Yoo, S.-J.: Q-learning-based data-aggregation-aware energy-efficient routing protocol for wireless sensor networks. *IEEE Access* **9**, 10737–10750 (2021). <https://doi.org/10.1109/ACCESS.2021.3051360>
 43. Prabhu, D., Alageswaran, R., Miruna Joe Amali, S.: Multiple agent based reinforcement learning for energy efficient routing in WSN. *Wireless Netw.* (2023). <https://doi.org/10.1007/s11276-022-03198-0>
 44. Shruthi, K.R., Kavitha, C.: Reinforcement learning-based approach for establishing energy-efficient routes in underwater sensor networks. In: 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, pp. 1–6, doi: <https://doi.org/10.1109/CONECCT55679.2022.9865724>. (2022)
 45. Su, X., Ren, Y., Cai, Z., Liang, Y., Guo, L.: A Q-learning based routing approach for energy efficient information transmission in wireless sensor network. *IEEE Trans. Network Serv. Manag.* (2022). <https://doi.org/10.1109/TNSM.2022.3218017>
 46. Kim, B.-S., Suh, B., Seo, I.J., Lee, H.B., Gong, J.S., Kim, K.-I.: An enhanced tree routing based on reinforcement learning in wireless sensor networks. *Sensors* **23**(1), 223 (2023)
 47. Su, Y., Lu, X., Zhao, Y., Huang, L., Du, X.: Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks. *IEEE Sens. J.* **19**(20), 9561–9569 (2019)
 48. Kwon, M., Lee, J., Park, H.: Intelligent IoT connectivity: deep reinforcement learning approach. *IEEE Sens. J.* **20**(5), 2782–2791 (2019)
 49. Shao, Y., Rezaee, A., Liew, S.C., Chan, V.: Significant sampling for shortest path routing: a deep reinforcement learning solution. *IEEE J. Select. Areas Commun.* **38**, 2234 (2020)
 50. Ryu, K., Kim, W.: Multi-objective optimization of energy saving and throughput in heterogeneous networks using deep reinforcement learning. *Sensors (Basel)* **21**(23), 7925 (2021)
 51. Barker, A., Swamy, M.: Distributed cooperative reinforcement learning for wireless sensor network routing. In: 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, pp. 2565–2570, (2022)
 52. Zhang, A., Sun, M., Wang, J., Li, Z., Cheng, Y., Wang, C.: Deep reinforcement learning-based multi-hop state-aware routing strategy for wireless sensor networks. *Appl. Sci.* **11**, 4436 (2021). <https://doi.org/10.3390/app11104436>
 53. Cong, P., Zhang, Y., Liu, Z., Baker, T., Tawfik, H., Wang, W., Ke, Xu., Li, R., Li, F.: A deep reinforcement learning-based multi-optimality routing scheme for dynamic IoT networks. *Comput. Netw.* **192**, 108057 (2021)
 54. Swain, R.R., Khilar, P.M., Dash, T.: Multifault diagnosis in WSN using a hybrid metaheuristic trained neural network. *Digit. Commun. Netw.* **6**(1), 86–100 (2020)
 55. Forster, A., Murphy, A.L.: CLIQUE: role-free clustering with Q-learning for wireless sensor networks. In: 2009 29th IEEE

- International Conference on Distributed Computing Systems, (pp. 441–449), IEEE (2009)
56. Kiani, F., Amiri, E., Zamani, M., Khodadadi, T., Abdul Manaf, A.: Efficient intelligent energy routing protocol in wireless sensor networks. *Int. J. Distrib. Sensor Netw.* **11**(3), 18072 (2015)
 57. Arya, G., Bagwari, A., Chauhan, D.S.: Performance analysis of deep learning-based routing protocol for an efficient data transmission in 5G WSN communication. *IEEE Access* **10**, 9340–9356 (2022)
 58. Abadi, A.F.E., Asghari, S.A., Marvasti, M.B., Abaei, G., Nabavi, M., Savaria, Y.: RLBEED: reinforcement-learning-based energy efficient control and routing protocol for wireless sensor networks. *IEEE Access* **10**, 44123–44135 (2022)
 59. Sharma, T., Balyan, A., Nair, R., Jain, P., Arora, S., Ahmadi, F.: ReLeC: a reinforcement learning-based clustering-enhanced protocol for efficient energy optimization in wireless sensor networks. *Wireless Commun. Mobile Comput.* **2022**, 3337831 (2022). <https://doi.org/10.1155/2022/3337831>
 60. Mutombo, V.K., Lee, S., Lee, J., Hong, J.: EER-RL: energy-efficient routing based on reinforcement learning. *Mobile Inform. Syst.* **2021**, 5589145 (2021). <https://doi.org/10.1155/2021/5589145>
 61. Jatoi, G.M., Das, B., Karim, S., Pabani, J.K., Krichen, M., Alroobaea, R., Kumar, M.: Floating nodes assisted cluster-based routing for efficient data collection in underwater acoustic sensor networks. *Comput. Commun.* **195**, 137–147 (2022)
 62. Agarwal, V., Tapaswi, S., Chanak, P.: Intelligent fault-tolerance data routing scheme for IoT-enabled WSNs. *IEEE Internet Things J.* **9**(17), 16332–16342 (2022)
 63. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wireless Commun.* **1**(4), 660–670 (2002)
 64. Tajmajer, T.: Modular multi-objective deep reinforcement learning with decision values. In: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS). pp. 85–93, IEEE (2018)
 65. Hayes, C.F., Verstraeten, T., Roijers, D.M., Howley, E., Mannion, P.: Expected scalarised returns dominance: a new solution concept for multi-objective decision making. *Neural Comput. Appl.* (2022). <https://doi.org/10.1007/s00521-022-07334-x>
 66. Moffaert, K.V., Drugan, M.M., Nowé, A.: Scalarized multi-objective reinforcement learning: novel design techniques. In: 2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL) (pp. 191–199). IEEE (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Walid K. Ghamry received his B.Sc. in 1988, the M.Sc. in 2003, and Ph.D. in 2010. All these degrees in Electronics and Computer Communications Engineering from faculty of Engineering at Cairo University. His research interests include mobile-wireless communications, network communications design, network routing, and Internet topology.



Suzan Shukry received her B.Sc. with cumulative grade of appreciation very Good with honor from Cairo University in 1996, the M.Sc. in 2003, and the Ph.D. in 2010. All these degrees in telecommunications and networks from the faculty of Engineering at Cairo University. Her research interests include telecommunications and networks.