WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Reinforcement Learning Based Routing Protocol for Energy Conservation in Wireless Sensor Network

***Chukwuka Chinemelu [a]*, Aniebiet I. Idim [a], Chika A. Egbunugha [b]***

[a] *Department of Electrical and Electronic Engineering, Petroleum Training Institute, Effurun, Nigeria*
[b] *Department of Electrical and Electronic Engineering, Imo State Polytechnic, Omuma, Nigeria*

**A B S T R A C T**

In wireless sensor network (WSN), one of the critical challenges to achieving optimal performance is energy efficiency and network lifetime. Thus, reducing the energy consumed and thereby extending lifetime of the network is the concern of most of the routing techniques. The use of clustering strategy is one of the methods largely employed to offer energy efficiency. Nevertheless, most of the methods using clustering techniques either randomly select the cluster head (CH) without regard to vital parameters or by using centralized scheme that involves the utilization of the base station (BS) which has the potential to influence the scalability of the network. This paper presents reinforcement learning (RL) based routing protocol for energy conservation in WSN. An enhanced clustering-based routing scheme that uses RL algorithm was proposed to improve the energy efficiency and lifetime of a WSN. Since energy consumption is an important factor that significantly affects the computational and routing effectiveness of nodes in WSN, an RL algorithm has been implemented with cluster-based routing scheme to enable the nodes to adjust to changes in the network parameters such as energy level and mobility (for instance, change in BS location), and improvement in routing assessments. In addition, a probability scheme is introduced that allows the nodes to join a cluster based on its energy level rather than assigning fixed energy values to different nodes in heterogeneous network. A WSN field of 100 × 100 square metre was configured in MATLAB to evaluate the performance of the proposed scheme. The simulation analysis conducted in MATLAB has shown that after 5000 rounds, the system routing scheme ensured that a significant number of the nodes remains alive with minimum energy consumption.

Keywords: Energy conservation, Reinforcement learning, Routing protocol, WSN

## 1. Introduction

The design of small, but capable remote wireless sensors that can be inexpensively deployed over large areas have been made possible by advances in technology that are geared towards miniaturization of integrated circuits (IC), transmitters and sensing devices. These wireless sensors have been widely applied in military and commercial applications such as presence or intrusion detection, monitoring of environment, ranging, imaging, and detection of noise (Durišić et al., 2012).

Wireless sensor network (WSN) is achieved by virtue of wireless communication between two or among thousands of sensor nodes fitted to detect or sensed certain type of inputs, which are the physical parameters or factors characterizing the environmental condition or nature of task for which they are deployed. The parameter to be detected can be pressure, temperature, smoke, heat, fire, shadow, water, amongst others depending on the intention the WSN is designed to operate (Muoghalu et al., 2022). Thus, a WSN can regard as a system of specialized nodes or devices that uses wireless medium to communicate collected data from monitored environment to a base station (BS) or sink node. The BS serves as an interface or connection between the WSN and the user.

A typical WSN can have an installed capacity of thousands of sensor nodes, and these nodes are generally resource constrained. This means the nodes are expected to use little power as possible from the available sufficient computational and transmit power in order to accomplish their function. This is because the nodes depend on the energy of a battery for power and are required to perform for long periods without being replaced. In addition, a node in a sensor network is characterised by limited processing speed, communication bandwidth, and storage capacity (Muoghalu et al., 2022).

Routing is responsible for a node to be part of a WSN and it is a high power-consuming process. Hence, each sensor including the routing task in the network place an energy cost or burden for every action carried out, which gradually reduces power of the sensor. When power is lost by a sensor, it cannot sense information, interact with other sensors nodes or route information. The sensor in this case is said to be dead. If single node dies, a major impact may not be felt on the WSN. However, as more and more nodes die out, the impact on WSN becomes obvious and the performance of the network degrades as it may become divided and no longer reliable. Thus, energy preservation is a critical factor to take into account in the design of routing algorithm for WSN. Even though several routing techniques have been developed to solve the problem of energy degradation in WSN, consideration for nodes to adapt to network changes such as mobility and energy level has often not been provided including enhanced routing decision. In this paper, an energy conservation routing based on reinforcement learning (ECR-RL) is proposed for WSN based logistic monitoring.

**Nomenclature**

$N \times N$is the dimensions of target monitoring field with randomly distributed nodes

$n$ number of nodes in the network

$b$ number of intermediate nodes

$c$is factor proportional to number of advanced nodes to total number of nodes $n$

$E_o$ is the residual energy (which is the initial energy level)

$E_{AN}$is the energy of advanced nodes

$E_{IN}$ is the energy of intermediate nodes

$\alpha, \mu$are constants representing the number of times the energy of the advanced nodes and intermediate nodes are greater than the energy of the normal nodes

$p_{opt\ is}$ optimal probability of nodes to be elected as a CH

$p_i$is the probability of ith node being elected as CH

$Q_i$ is the initial Q-value

$N_h$ is the hop count

$E_{\min}$ and $E_{\max}$are the minimum energy and maximum energy dissipated

$D_{link}$ is the distance of a sensor node $s_i$ to the BS through an intermediate sensor node $s_j$

$TX_{range}$ is the transmission range

$D_{ij}$ is the distance between ith sensor node and jth sensor node

$D_{j\text{sink}}$ is the distance of jth sensor node from sink node (or BS)

$x_i, x_j, x_{\text{sink}}$ are the x-coordinate locations of ith sensor node, jth sensor node

$y_i, y_j, y_{\text{sink}}$ are the y-coordinate locations of ith sensor node, jth sensor node, and sink node respectively

$E_{Tx}$ is the energy consumed per *k*-bit packet at transmitter

$E_{Rx}$is the energy consumed per *k*-bit packet at receiver

$E_{amp}$ transmit amplification factor (or amplifier coefficient)

$E_{elect}$is the energy consumed by the transmit electronic or the receive electronic

$k$ is the number of bits transmitted

$d$ is the distance between a sensor node and its cluster head (CH) or distance between a CH and another CH nearer to the base station (BS) or distance between an ith sensor node and a jth intermediate sensor node or distance between a jth intermediate sensor node and BS or simply the distance between CH and BS

$Q_t$is the initial Q-value or Q-value at time t

$Q_{t+1}$ is the updated Q-value,

$\alpha$ is the rate of learning (whose value is usually taken as 1 to speed up the process of learning)$r_{t+1}(s, a)$ is the immediate reward,

$\max_a Q(S', a)$ is the highest course of action used to optimized the reward,

$S'$is a maximum state which the course of action or policy ends up

$\gamma$ is the discount factor (which varies between 0 and 1).

### *1.1 Reinforcement learning in routing*

The formulation of RL problems is based on Markov decision processes (MDP) with a tuple (S, A. P. R), such that S, A, P, R representing the state of an agent at time t, possible action an agent can take, transition probability, and reward obtained by an agent for action carried out (Sutton and Barto, 2020). Though routing protocols are enhanced using RL, there is need to define the main elements of the RL model such as agent and environment, state and action, and reward while applying it (Mutombo et al., 2021). In RL model, the agent makes the decision while the environment is the element being

observed and reacted to by the agent. In this case, each sensor in WSN is regarded an agent, while multi-agent RL is required for the whole network. The state and the action represent any useful information concerning the environment at a specific time and the reaction of the agent at given time. The available routing information from all accessible sensor nodes is considered the state space for an agent (or a node). A set of decision-making factors such as residual energy, hop count, and strength of signal can be a potential state, but this depends on the factors considered during the design of the protocol. Conversely, selecting the next-hop for routing packet towards base station (BS) is referred to an action. The action space is therefore regarded as the set of all available routes through neighbours at a specific time. The reward is the cost of the action carried out by an agent in a given state (Mutombo et al., 2021).

### *1.2 Implemented reinforcement based routing protocol*

There are some studies on routing protocol in WSN that have implemented RL to enhanced energy efficient in the network. In Internet of Things (IoT) based wireless device network, energy consumption has been optimized so as to increase the lifetime of the network by finding the optimal route for data transmission using energy efficient routing protocol based on RL (Mutombo et al., 2021). An approach for scheduling the wake-up cycles of nodes in a WSN based on self-organizing RL was proposed in Mihaylov et al. (2012). With focus on assessing and increasing power consumption efficiency and minimized energy loss of sensor node to extend lifetime of sensor network, an RL based and clustering technique was implemented by Sharma et al. (2022). In order to optimize routing techniques for energy management in WSNs, an energy-efficient control and routing algorithm based on RL was designed by Abadi et al. (2022). Routing protocol utilizing RL for achieving network lifetime optimization was proposed by Guo et al. (2019). Energy efficiency and network lifetime improvement was achieved for WSN using RL based routing protocol by Simon (2022). For improved performance in terms of end-to-end delay, packet delivery ratio, and energy consumption, a classic problem of finding an optimal parent node in a tree topology was addressed using RL based routing algorithm (Kim et al., 2023). The use of RL in WSN was surveyed by Soni and Shivastava (2018). A deep RL (DRL) algorithm was developed for attacking WSN in Paras et al. (2021).

## 2. System model

In developing the routing scheme, the system model is that of sensor nodes operating environment that comprises of $n$ sensor nodes. The nodes are randomly distributed in a target monitoring field of $N \times N$ dimensions, and remain stationary after deployment. The ability to sense or measure activities (or events), collect, process and send data within the network is possessed by each node. Each node in the network is equipped with battery energy that is constrained and mostly depleted during the transmission and reception of data at the radio transceiver of the node. After the deployment of nodes in conventional WSN, they are left unattended. Thus, the replacement of battery or its recharging is unfeasible. When event takes place, data sensing will occur, for instance, temperature rise in an environment where the nodes are deployed. Figure 1(a) shows the structure of the proposed WSN model. Cluster member (CM) node senses data and broadcast it to cluster head (CH). The information sensed is transmitted to the sink via CHs. The communication with sink or base station (BS) is centralized within the network area. It is assumed that the BS has no constraint as regards energy, computation of power, and memory resources. The proposed wireless sensor network set-up uses a centralized routing technique in which the sink node is the centralized item. This means that all nodes can communicate with the sink directly via the CH in the group which they belong. In addition, transmissions from all clusters to sink node are only done via CHs except for sensor nodes with transmission range close to sink node that do not need to join any cluster, but can directly communicate with the base station (or sink node) and thus resulting to energy conservation. Besides, the initial Q-value based on reinforcement learning (RL), which is computed from hop count factor and initial energy, is employed in the selection process of the CH. Figure 1(b) shows the structure of the WSN model in MATLAB.

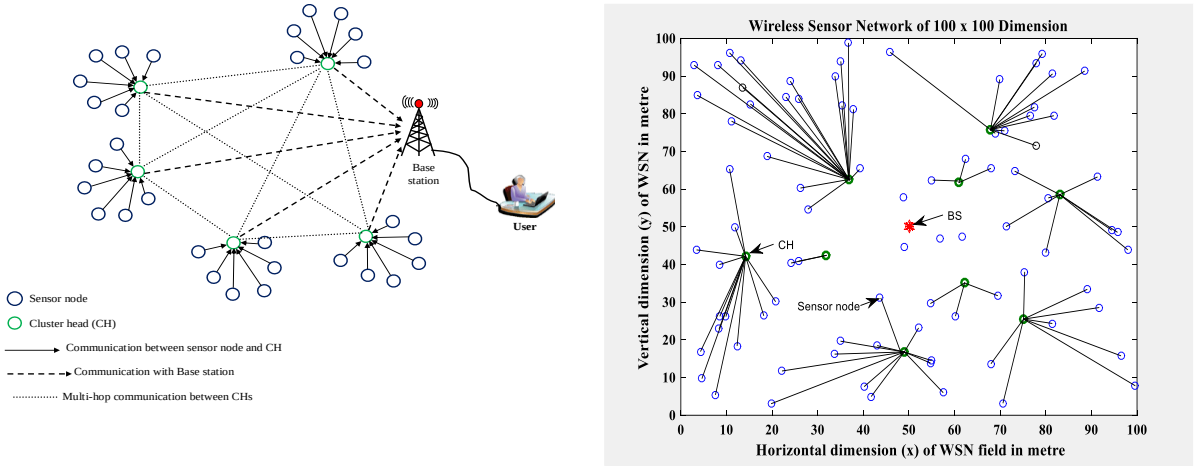


**Fig. 1 - (a) Proposed WSN model (b) WSN set-up in MATLAB**

Since the proposed system is a heterogeneous WSN, the sensor nodes are assumed to have different energy level. Thus, three levels of heterogeneity are considered as: normal nodes, intermediate nodes, and advance nodes. An illustration of the energy levels of sensor nodes in the network is shown in Fig. 2.

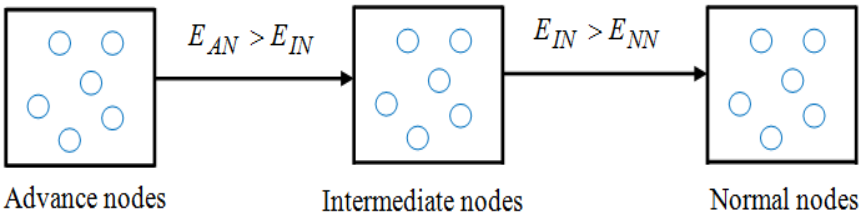


**Fig. 2 – Energy levels of sensor nodes**

The energy distribution of nodes in the network as shown in Fig. 2 is such that the advance nodes have the highest value of energy in the sensor network followed by intermediate nodes and then the normal nodes.

The intermediate nodes are equipped with $\mu$ times energy more than normal nodes $E_o$. The energies of the advance nodes and the intermediate nodes are given by (Kashaf et al., 2012):

$$E_{AN} = E_o(1+\alpha) \quad (1)$$

$$E_{IN} = E_o(1+\mu), \text{where } \mu = \frac{\alpha}{2} \quad (2)$$

Presenting the total energy of each energy level classified in terms of the number of sensors gives:

$$\text{Normal nodes} = n..b(1+\alpha) \quad (3)$$

$$\text{Intermediate nodes} = nE_o.(1-c-bn) \quad (4)$$

$$\text{Advance nodes} = n.c.E_r(1+\alpha) \quad (5)$$

Therefore, the total energy of the nodes is given by:

$$E_T = n..E_o(1+c\alpha+b\mu) \quad (6)$$

Optimal probability of node being elected as CH is given by (Kashaf et al., 2012):

$$p_i = \begin{cases} \frac{p_{opt}}{1+c.\alpha+b.\mu} & \text{if } s_i \text{is the normal node} \\ \frac{p_{opt}.(1+\mu)}{1+c.\alpha+b.\mu} & \text{if } s_i \text{ is the intermediate node} \\ \frac{p_{opt}.(1+\alpha)}{1+c.\alpha+b.\mu} & \text{if } s_i \text{ is the advance node} \end{cases} \quad (7)$$

Now, since the CH election is determined from hop count factor and residual energy used to compute the initial Q-value, the following equations are defined (Mutombo et al., 2021):

$$Q_i = \begin{cases} \frac{1}{N_h}, & \text{if } E_{min} = E_{max} \\ p_i \times \left(\frac{E_o - E_{min}}{E_{max} - E_{min}}\right) + (1-p_i) \times \frac{1}{N_h}, & \text{if } E_{min} \neq E_{max} \end{cases} \quad (8)$$

$$N_h \cong \frac{D_{link}}{TX_{range}} \quad (9)$$

Thus:

$$D_{link} = D_{ij} + D_{j\text{sink}} \quad (10)$$

$$D_{ij} = \left[(x_i - x_j)^2 + (y_i - y_j)^2\right]^{1/2} \quad (11)$$

$$D_{j\text{sink}} = \left[(x_j - x_{\text{sink}})^2 + (y_j - y_{\text{sink}})^2\right]^{1/2} \quad (12)$$

### *2.1 Energy consumption model of WSN*

The proposed scheme in this work uses the energy model adopted from Heinzelman et al. (2000). In the WSN, the sensor nodes are responsible for energy degradation. The energy consumption model of WSN can be described using a radio model. During wireless communication in sensor network, both the sender and receiver consume energy after transmission of packets. However, more energy is consumed by the sender than the receiver over the network (Mutombo et al., 2021). Thus energy consumption model evaluates the dissipated energy at transmission of packet or its reception and then updates the residual energy. The block diagram of a first order radio communication model is shown in Fig. 3.

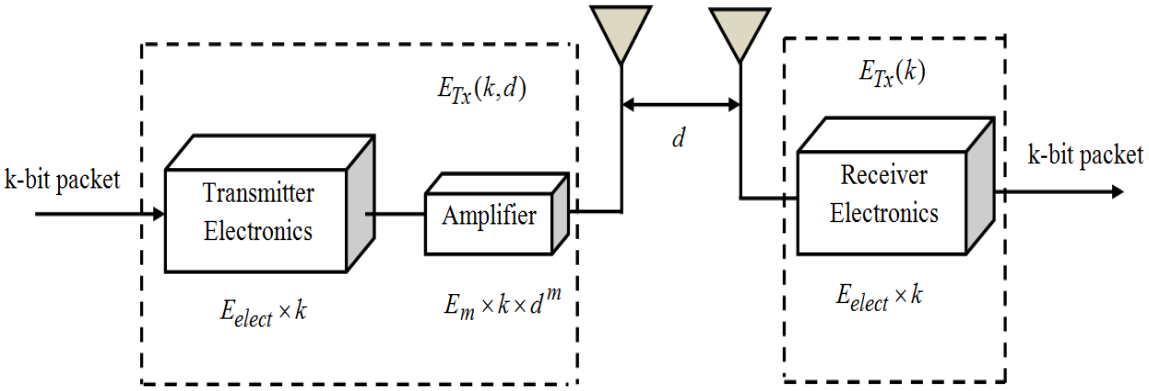


**Fig. 3 – Block diagram of radio energy consumption model**

From Fig. 3, the energy consumed k-bit of packet is transmitted through a distance $d$ is mathematically expressed as given by Heinzelman et al. (2000):

$$E_{Tx} = E_{elect} \times k + E_{amp} \times \left(k \times d^2\right) \quad (13)$$

Equation (13) described the energy dissipated over a shorter transmission distance such as within clusters. However, for transmission over a longer distance, such as from a cluster head to the base station, Eq. (13) can be expressed by Heinzelman et al. (2000):

$$E_{Tx} = E_{elect} \times k + E_{amp} \times \left(k \times d^4\right) \quad (14)$$

It should be noted that the index power of the distance $d$ over which the amplified signal is transmitted has been assigned m = 2 or 4 depending on the distance as in Eq. (13) and (14). Correspondingly, the energy consumed regarding receive of packet is given by:

$$E_{Rx} = E_{elect} \times k \quad (15)$$

### *2.2 Reinforcement learning application*

The application of RL with respect to energy consumption model ensures that the residual energy (or initial energy) is updated by subtracting the energy consumed subsequent to each packet transmission. The updated initial energy, $E_o$ and together with the hop count, $N_h$ are used to compute the reward function. Eventually, the Q-value is then obtained using the reward. The reward function is computed using the expression given by Mutombo et al. (2021):

$$r_{t+1} = \begin{cases} \dfrac{1}{N_h}, & \text{if } \mathrm{E_{min}} = \mathrm{E_{max}} \\ p_i \times \left(\dfrac{E_o - E_{\min}}{E_{\max} - E_{\min}}\right) + (1 - p_i) \times \dfrac{1}{N_h}, & \text{if } \mathrm{E_{min}} \neq \mathrm{E_{max}} \\ -100 & \text{if } \mathrm{E_o} < 0 \end{cases} \quad (16)$$

where $r_{t+1}$ is the reward function.

The updating function for computing Q-value is an action-value function that describes how good it is to carry out an action from a specified state following a course of action $\pi$ (Mutombo et al., 2021; Sutton and Barto, 2020; Mammeri, 2019).The action-value function is given by (Mutombo et al., 2021):

$$Q_\pi(s,a) = \mathrm{E}\left[G_t \mid S_t = s, A_t = a\right] \quad (17)$$

$$G_t = \sum_{k=0}^{\infty} \gamma^k \times R_{t+k+1} \quad (18)$$

An action-value function approximation proposed by Watkins and Dayan (1992) provided an easy to implement and applicable Q-learning in several cases (Mammeri, 2019) and it is given by:

$$Q_{\pi *} = Q*(s,a) \quad (19)$$

Thus the Q-value for learning algorithm is given by (Mutombo et al., 2021):

$$Q_{t+1}(s,a) = (1-\alpha)Q_t(s,a) + \left( r_{t+1}(s,a) + \gamma \max_a Q(S',a) \right) \quad (20)$$

Generally, worthy of note in this work is that the approach to application of RL algorithm implementation in IoT by Mutombo et al. (2021) is being modified for heterogeneous WSN with three (initial) energy levels defined and different probabilities computed according to initial energy level rather than assigning fixed or defined energy level as in previous study. This approach allows the sensor nodes to determine their energy level based on probabilistic model in accordance with their initial energy.

### *2.3 Simulation parameters*

So far this section has presented the approach to implementing the proposed enhanced energy preservation scheme for WSN. In this section, the parameters of a heterogeneous WSN presented in Kashaf et al. (2012) are adopted but modified to accommodate the parameters for implementing the RL algorithm as shown in Table 1.

**Table 1 – Simulation parameters**

| Definition | Symbol | | Value |
|---|---|---|---|
| Sensor network area | A | | $100 \times 100$ m$^2$ |
| Coordinate location of base station (BS) | (x, y) | | (50, 50)m |
| Electronic energy | | $E_{elect}$ | $50 \times 10^{-9}$ J/bit |
| Data aggression energy | $E_{DA}$ | | $5 \times 10^{-9}$ J/bit/message |
| Initial Energy of sensor node | $E_o$ | | 0.5 J |
| Amplifier energy | $E_{amp}$ | | $0.0013 \times 10^{-12}$ J/bit/m$^4$ |
| Number of sensor nodes | n | | [20-100] |
| Packet size | k | | 4000 bits |
| Transmission range | r | | 20 m |
| Speed rate | $\alpha$ | | 1 |
| Discount factor | $\gamma$ | | Varies between 0 and 1 |
| Energy level factor | c | | 0.1 |

## 3. Results

The performance of the developed routing protocol presented in this chapter based on the simulation analysis carried out in MATLAB environment. In the system, 100 sensor nodes, which were randomly distributed, were deployed over a sensing area of 100 by 100 square metres. The sink node or BS was located at the centre of the sensing filed on (x, y) = (50, 50) coordinates. In addition, the WSN is considered to be heterogeneous in which the sensor nodes are categorized into three energy levels defined as normal nodes, intermediate nodes and advanced nodes. The performance of the system was initially varied for different probability level. The performance of the system was evaluated in terms of number of operating nodes per rounds, number of dead nodes per rounds, and energy consumed per rounds. The system was further evaluated in terms of number of nodes, number of packets (or messages), and comparison with previous system.

### *3.1 Simulation results based on varying probability levels*

In this subsection, the simulation results are presented in this case considering different probability (p) regarding energy levels of defined for the heterogeneous. Simulation tests were conducted in this regard for different values of probability so as to optimize the performance of the proposed routing protocol called energy conservation routing protocol based reinforcement learning (ECRP-RL). The results have been further presented in statistical tables and graphs for clearness and perceptive. The simulation curves are shown in Fig.4 to 6.

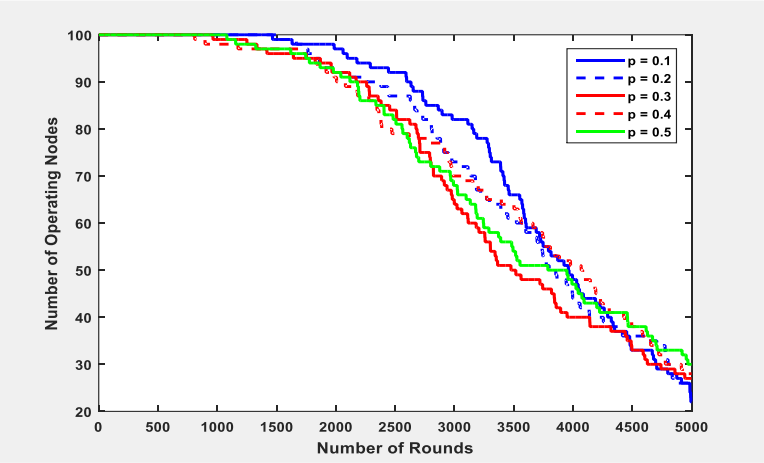


**Fig. 4 – Operational nodes per number of rounds (for different probabilities)**

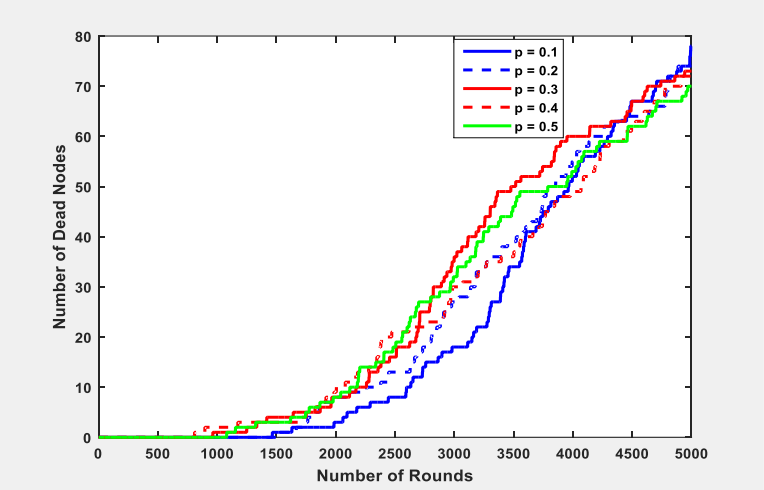


**Fig. 5 – Dead nodes per number of rounds (for different probabilities)**

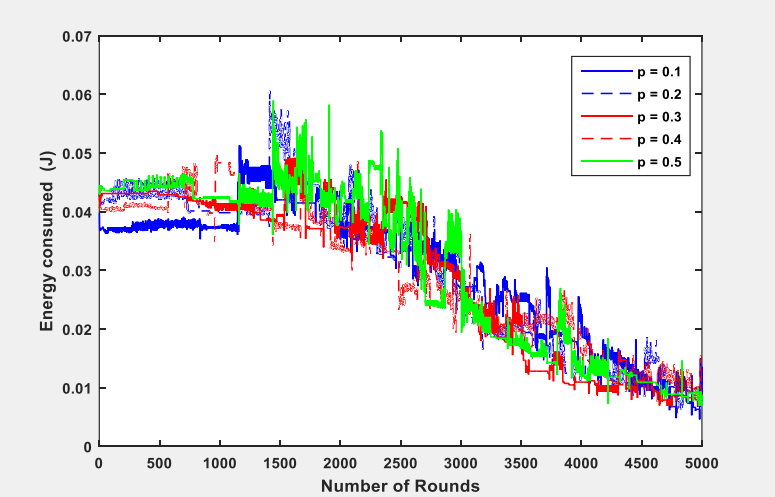


**Fig. 6 –Energy consumed per number of rounds (for different probabilities)**

Looking at Fig.4, it can be seen that as the number of rounds increases, the number of operational (or alive) nodes in the WSN decreases. The numerical analysis revealed that after 5000 rounds, the probability p = 0.5 provided the best performance by prolonging the lifetime of the network since more operational nodes were still available. In Fig. 5, the protocol is evaluated in terms of number of nodes that have used up their energy. It can be seen that as the number of rounds increases, the number of dead nodes (i.e. nodes whose battery no longer has stored energy) in the WSN increases. The analysis revealed that after 5000 rounds, the probability p = 0.5 provided the best performance because at this level of probability, the number of dead nodes in

the network were less compare to other probabilistic factor. Figure 6 shows the simulation curves of the energy consumed by the sensor nodes in the network for each probability factor. Thus, it is the evaluation of the energy consumed in the work with respect to each probability factor considered. it was observed that that for p = 0.5, lesser energy is consumed such that after 5000 rounds, the residual energy was compared to other probability factors. Since, p = 0.5 provided the finest performance, it has been considered the value for optimal performance of the network and used for other subsequent simulations performed. Generally, the number of alive nodes after 5000 rounds was equal to 22, 26, 27, 28, and 32 for p = 0.1, 0.2, 0.3, 0.4, and 0.5.

### *3.3 Simulation results for different number of packets*

This subsection presents the results of the simulation analysis carried out for different volume of traffic (or packets). The MATLAB simulation curves are presented in Fig. 7 to 9.

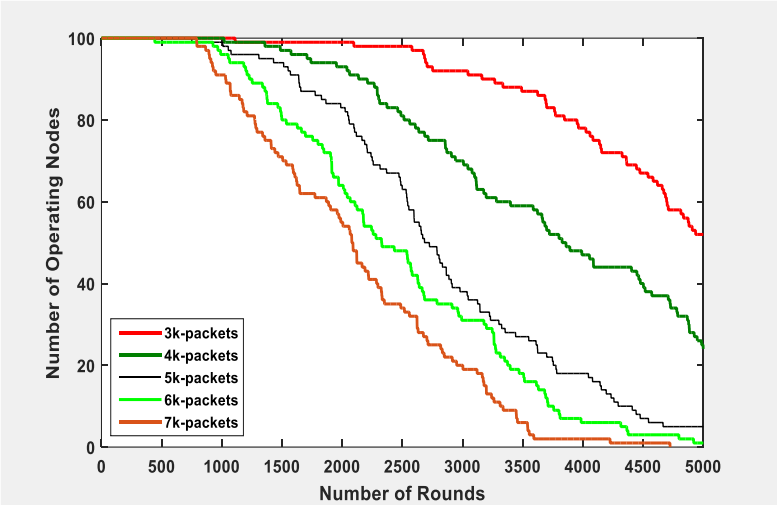


**Fig. 7 –Number of alive nodes per rounds for different packets**

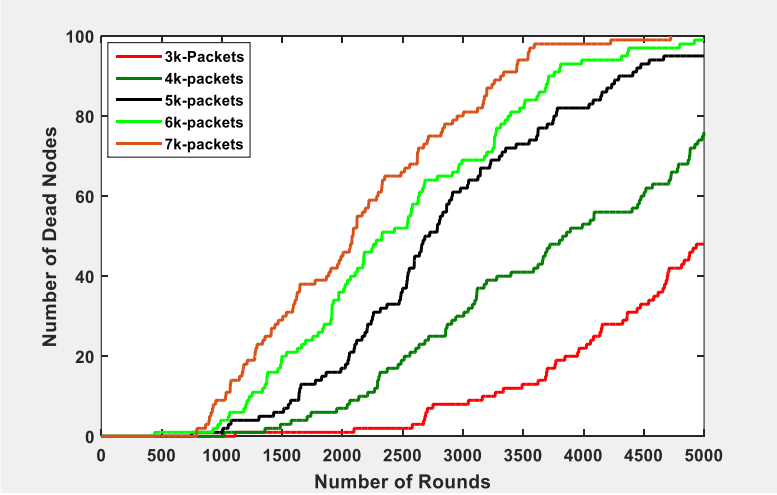


**Fig. 8 –Number of dead nodes per rounds for different packets**

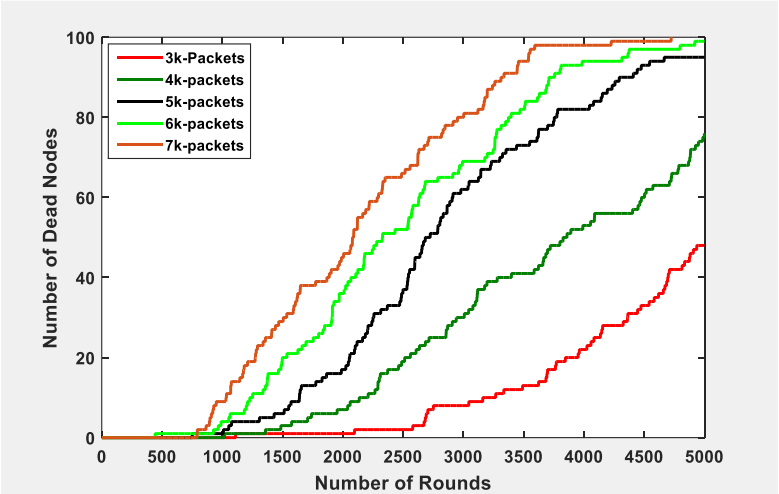


**Fig. 9 –Number of dead nodes per rounds for different packets**

Figure 7 is the evaluation of the routing technique in terms of increasing number of packets. The numerical analysis indicates that as the number of packets increases the number of operating or live nodes decrease. Hence, when the number of packets was 3000, the number of nodes alive was more than the half the number of nodes deployed in the network after 5000 rounds. However, when the number of number of packets was increased to 7000, the number of nodes dropped to zero from 4724 -5000 rounds. Figure 8 is the evaluation of the proposed scheme based on the number of nodes that died considering the volume of traffic in the wireless network. The numerical analysis showed that when the number of packets was 3000, the number of dead nodes were 48, which is less than the number of alive nodes. Whereas, with number of packets increased, it can be seen that the number of dead nodes increases and that reduces the network life and efficiency. Thus as the packets are 7000, at 4724 to 5000, all the nodes were observed to be dead. Figure 9 is the representation of the evaluation of the proposed routing scheme in terms of the effect of increasing number of packets or volume of traffic on the energy consumed per rounds by the sensor nodes in the network. It can be seen that after 5000 rounds, with 3000 packets, 0.01781 J of energy is still available in the network whereas for 7000 packets, as at 4724 rounds the energy left in the network was 0.000423 and drops to 0 at 5000 rounds. Considering that the initial energy of sensor was 0.5 J (as in Table 1), it can be said that with 3000 packets the percentage of energy consumed was by 96.4% and 98.5% for 4000 packets, but it is 99.9% with 7000 packets.

## 4. Conclusion

The foremost objective of this work was designed towards enhancing the energy efficiency and prolonging the lifespan of a wireless sensor network (WSN) deployed for environment-temperature monitoring. The network lifespan has be considered in this work to mean the time duration expressed in terms of rounds, by which there is no possibility for data transmission. This was attributed to the number of nodes still in operation in the network after 5000 rounds. The indices considered in the simulation analysis conducted to evaluate the performance of the energy conservation routing protocol based reinforcement learning (ECRP-RL) are the number of operating nodes per rounds, the number of dead nodes per rounds and the energy consumed per rounds, which is aggregate of energy consumed by the sensor nodes for each number of rounds. The optimal probability factor to balance energy consumption in the network was determined by testing a value from 0.1 to 0.5. From the simulation test conducted in this regard, the outcome revealed slight different for the various probability factors especially for 0.3, 0.4 and 0.5. However, the simulation with 0.5 showed better consistency even though 0.4 and 0.3 yields similar results in some instances. Thus, 0.5 was chosen for simulation analysis conducted in this work. The analysis of the proposed protocol in terms of transmitted packets revealed that as the energy of the network is reduced as the volume of traffic is increased. This means that more energy is consumed in the network with increased number of transmitted packets. This also, resulted in more dead sensor nodes because of the obvious amount of energy that is required to transmit huge volume of packets in the network.

Generally, the primary objective was to minimize the consumption of energy by the nodes in WSN and thereby extending the lifespan of the network. This is achieved using reinforcement learning to find the most favorable route for transmission of data in the network. The performance of the proposed scheme was evaluated in terms of alive nodes, dead nodes and energy consumed per rounds. The simulation analysis conducted in MATLAB has shown that after 5000 rounds, the system routing scheme ensured that a significant number of the nodes remain alive with minimum energy consumption.

**References**

Abadi, A. F. E., Asghari, S. A., Marvasti, M. B., Abaei, C., Nabavi, M., & Savaria, Y. (2022). RLBEEP: reinforcement-learning-based energy efficient control and routing protocol for wireless sensor networks. *IEEE Access,10*, 44123-44135.

Đurišić, M. P., Tafa, Z., Dimić, G., & Milutinović, V. (2012).A survey of military applications of wireless sensor networks. *In Proceeding of Mediterranean Conference on Embedded Computing, Bar, Montenegro*, 196–199.

Guo, W., Yan, C., & Lu, T. (2019). Optimizing the lifetime of wireless sensor networks via reinforcement learning-based routing. *International Journal of Distributed Sensor Networks*, *15*(2), 23-34. DOI: 10.1177/1550147719833541

Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2000). Energy efficient communication protocol for wireless sensor networks. in: *Proceeding of the Hawaii International Conference System Sciences*, Hawaii, January 2000.

Kashaf, A., Javaid, N., Khan, A., & Khan, T. A. (2012). TSEP: Threshold-sensitive Stable Election Protocol for WSNs. arXiv:1212.4092v1 [cs.NI]. http://arxiv.org/abs/1212.4092v1

Kim, B.-S., Suh, B., Seo, I. J., Lee, H. B., Gong, J. S., & Kim, K.-I. (2023). An Enhanced Tree Routing Based on Reinforcement Learning in Wireless Sensor Networks. *Sensors*, *23*, 223. https://doi.org/10.3390/s23010223

Mammeri, Z. (2019). Reinforcement learning based routing in networks: review and classification of approaches. *IEEE Access*, *7*, 55916–55950.

Mihaylov, M., Le Borgne, Y.-A., Tuyls, K., & Nowé, A. (2012). Decentralised reinforcement learning for energy-efficient scheduling in wireless sensor networks. *International Journal Communication Networks and Distributed Systems*, 9(3/4), 207-223.

Muoghalu, C. N., Achebe, P. N., & Aigbodioh, F. A. (2022). Effect of increasing node density on performance of threshold-sensitive stable election protocol. *International Journal of Advanced Networking and Applications*, *13*(6), 5183-5187.

Mutombo, V. K., Lee, S. Lee, L. J., & Hong, J. (2021). EER-RL: energy-efficient routing based on reinforcement learning. *Mobile Information Systems, Volume 2021*, Article ID 5589145, 1-12. https://doi.org/10.1155/2021/5589145

Parras, J., Hüttenrauch, M. Zazo, S., & Neumann, G. (2021). Deep reinforcement learning for attacking wireless sensor networks. *Sensors, 21,*4060, 1-20.https://doi.org/ 10.3390/s21124060

Simon, J. (2022). An energy efficient routing protocol based on reinforcement learning for WSN. *IRO Journal on Sustainable Wireless Systems*, *4*(2), 79-89. https://doi.org/10.36548/jsws.2022.2.002

Soni, S. & Shivastava, M. (2018). Various techniques of reinforcement learning for implementing wireless sensor network. *International Journal of Creative Research Thoughts*, *6*(1), 1206-1209.

Sutton, R. S., & Barto, A. G. (2020). Reinforcement Learning: An Introduction, The MIT Press, Cambridge, UK, 2nd edition