Reinforcement Learning Based Energy Management (Rl-EM) Algorithm For Green Wireless Sensor Network

[*]V Mahima[1], A. Chitra[2]

{[*]mahi.90may@gmail.com , ctr.mca@psgtech.ac.in }

[*]Research scholar, PSG College of Technology, Coimbatore-641004, India[1]
Professor and Head, Department of Computer Applications, PSG College of Technology, Coimbatore-641004, India[2]

Abstract-Green network is basically capable of producing its own energy through renewable energy resources. The sensor nodes serving to collect the physical parameters are equipped with small scale solar panels and tiny wind mills to harvest their own energy. The limitations on battery and availability of renewable energy resource create challenges in continuous monitoring without missing any events. The energy harvesting sensor network requires dynamic adaptation to the time varying behavior of the environment. The reinforcement learning based energy management algorithm named RL-EM is proposed which observes and learns from the environment. The voltage level of the sensor node is considered for designing the algorithm. The RL-EM outperforms existing algorithms with 40% reduced sleep nodes and 60% reduced energy overflow with LEACH protocol. The proposed RL-EM algorithm outperforms the LEACH, SEP-M and ACO algorithms in terms of throughput, sleep nodes, energy overflow, and load distributions.

Keywords – Markov Decision Process, Reinforcement Learning, Wireless networks, Internet of Things and Energy management.

1   Introduction

The Wireless sensor network is the primary data collection module of Internet of Things (IoT) applications. The sensor nodes play a vital role in creating IoT environment. The co-operative functioning of sensor nodes ensures reliable communication within the network. This effective communication prolongs till the node's energy is drained. The energy of the nodes is dependent on the capacity of the battery. The rechargeable batteries are provided with certain lifetime and rechargeable cycles. Though, the deployment of these sensor nodes in remote area reduces the chances of recharging the battery. In such situations, the effective utilization of energy from the battery is required. Many works were discussed on increasing the lifetime of the sensor nodes as well as the network. The efficient clustering and routing techniques, battery modeling, sleep scheduling techniques and many node scheduling techniques were proposed to enhance the lifetime of the network[1]. However, many critical applications like disaster monitoring, military and health care applications require continuous operation of the sensor nodes. For such applications, energy harvesting proves to be a promising solution. The energy harvesting module along with an energy buffer is provided with the sensor node[2]–[4]. This ensures that the energy harvested can be used directly or can be stored in buffer for future use. Though there are various ambient sources like solar, wind, vibration and RF are available for harvesting, the solar

energy is widely preferred due to its abundance[5]–[8]. The type of application decides the source of energy to be harvested. The solar energy availability is random and usually unpredictable. There may be energy overflow where, excessive energy is available but there is no space in storage buffer to store the energy. The other case is energy outage, where there is meager or no energy available to the sensor node for storing in the buffer. This uncertainty causes major issue in energy harvesting sensor network.
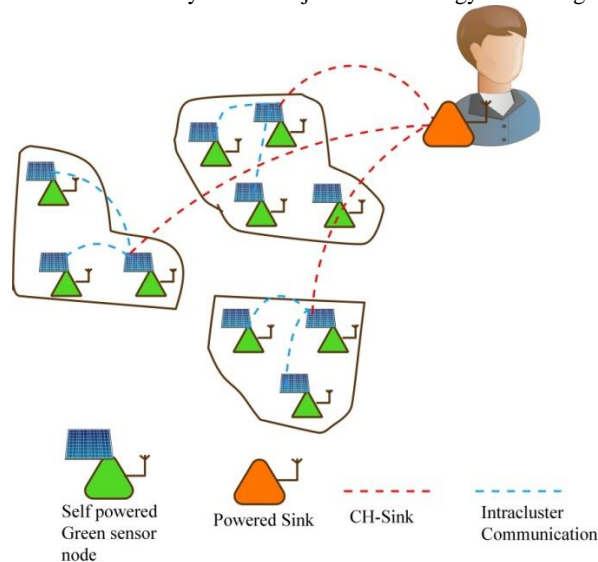

Figure 1 Architecture of Green network for IoT applications

Many effective scheduling techniques are available which relies on the energy arrival rate. The prediction based energy harvesting helps in determining the energy availability. Due to the significant overhead caused due to prediction techniques, the prediction free methods are used. Many IoT applications involves heterogeneous environment with heterogeneous sensor nodes where the knowledge about the environment is not known in prior [9]–[11]. In such cases, the system should learn from the environment and dynamically adapt to the changes. To ensure the dynamic adaptation of sensor nodes to the time varying behavior of energy sources a prediction free technique is required. To resolve the problem of uncertainty, the Reinforcement Learning technique is used, where the system learns from the environment based on rewards. The RL-EM continuously explores the environment to attain the optimal solution. In this paper a Reinforcement Learning based energy management algorithm is designed, which requires the voltage status of the battery for optimal scheduling and energy management.

The rest of the paper is organized as follows. The section 2 discusses related works. The section 3 discusses about the formulation of energy harvesting problemand proposed RL-EM algorithm with Markov Decision process. In Section 5, the evaluation of RL-EM and comparison of RL-EM with related algorithms are presented. The results show significant improvement in lifetime, throughput and energy efficiency. The following section discusses about the works supporting our proposed algorithm.

2 Literature Survey

An energy manager based on reinforcement learningRLMan is discussed which requires only the state of charge of the battery in [12]. The RLMan algorithm focuses on maximizing the packet rate while maintaining the node sustainability. An MDP is formulated as tuple with State space, Action space, Transition Function and Rewards. The simulation of RLMan is performed in Pow Wow platform with indoor light and outdoor wind. As the SoC of the battery is required for operation of this algorithm, it is hard to determine for heterogeneous environment. In [13], the RL is used to find the shortest path. The Q-Learning and SARSA is used and the path finding procedure is modeled as a MDP. In [14], the

Reinforcement based sleep scheduling algorithm is discussed for time slotted operation. Here, the re-grouping frequency is reduced such that the process is done when one node dies. This algorithm concentrates on desired coverage area. In [15], an energy preserving MAC protocol is designed with the aim of increasing network lifetime. The Q-Learning technique is used for solving minimization problem as well as to make the network self-adaptive. This is achieved by concentrating on the radio module's sleep and active schedule. In [16], the optimal routing path is created with the help of RL algorithm. The reward function is formed based on hop count, residual energy and link distance. This algorithm concentrates on improving the packet delivery along with balancing the energy consumption. In [17], a reinforcement learning-based throughput on demand (ToD) method is discussed which mainly focuses on throughput improvement. The learning process makes this method to adjust the duty cycle autonomously. The residual energy of the battery is improved along with the throughput.In [18], the RL is used for point to point communication. Here an assumption is made that there is data available in the transmitter always to be transmitted. The Markov Decision process is modeled with the aim of improving throughput. Also, function approximation is used to improve the performance [19] also concentrates on throughput maximization as but with the two hop communication. Here, each node is capable of solving power allocation problem independently. In [20], the review of Reinforcement Learning algorithm is done. It clearly pictures the applications of RL for WSN. It also discusses performance features and issues associated with RL algorithms. In [21], an intelligent algorithm named RLLO is proposed which concentrates on Intelligent routing by uniformly distributing the energy consumption. The packet delivery rate is also improved with this algorithm. The [22], discusses about the applications of RL algorithms in distributed routing. It also discusses about the routing challenges along with three different RL models for routing. The RL based hardware implementation for routing is discussed. In [23], the PSO based routing is proposed which concentrates on equally distributing the load to the Cluster Heads. Here the fault tolerance is considered as a factor to avoid the sudden failure of the node. In [24], the Enhanced PSO based optimization is proposed which again concentrates on clustering Cluster head selection. In [25], a dynamic power management technique is employed with RL algorithm to improve the energy efficiency and lifetime. The [26], discusses about the automatic configuration of sensors with Reinforcement Learning technique. It is concentrated on energy harvesting sensors for periodic and event driven indoor sensing. Here the quality of the sensor is improved by making it continually adapt to the changing environmental patterns. The above discussed works concentrates on improving lifetime, throughput and efficiency of the network with homogeneous nodes through RL algorithm. However, only few works were concentrated on heterogeneous environment which is a very important constraint of IoT applications. Many real world applications for which the exact environmental knowledge is not available, the RL algorithms are used to provide optimal solution. In our work, the RL based energy management algorithm is proposed which concentrates on efficient energy management for heterogeneous network.

3 Reinforcement Learning Based Energy Management Approach

Reinforcement approach is self-adaptive and capable to configure based on the environmental factors. The nodes inside the heterogeneous environment are subjected to diversified conditions like different energy states, dynamic battery level etc. The reinforcement learning approach is better suitable for real time deployment where prior knowledge about the region of interest is unknown.

Formulation Of Energy Harvesting Problem:

RL algorithms are used to solve optimization problems formulated as Markov Decision Process (MDP). A MDP is a tuple <S, A, T, R> where S is the state space, A is the Action space, T is the probability transition function and R is the reward function. It is assumed that the energy stored in the storage device is of finite capacity. The voltage level of the node is assumed to be in maximum level initially V_p. The minimum voltage threshold level is V_l below which the node dies due to lack of power. The transition function gives the probability of transition to next state, when the action is performed in current state. The reward is a function of packet rate and voltage level of the battery.

1)   The set of states' **S**: Current Voltage state of the battery V_c

State space S ϵ [V_L, V_r, V_p]

where V_p is the peak Voltage lying in high voltage state &V_r is the recovery voltage lying in medium Voltage state, V_l is the Voltage below which node goes to sleep state.

2)   The set of actions **A** : Setting the Packet rate P_r

Action space A ϵ [P_{min}, P_{max}]

where P_{min}, $P_{max}$is the minimum and maximum packet rate of the node.

3)   **The Transition FunctionT** : Probability of transition to V_c[n+1] when action P_r is performed in State V_c[n].

4)   **The Reward FunctionR** : Computed as a function of P_r and V_c as given in equation 1, in which V_c is the current voltage of the   battery at time t.

R = Φ. P_r (1)

where Φ = (V_c- V_r) / ($V_{p-}V_r$)

The probability transition matrix indicating the action taken is given by the equation 2

$$P_{xx'}^{a} = P[S_{t+1} = s'|S_t = s, A_t = a]$$ (2)

The reward function given to the node is given by the equation 3.

$$R_x^a = E[R_{tsec+1}|S_t = s, A_t = a]$$ (3)

The 'a' in the equation 3 indicates that the reward given to the sensor node mainly depends on the action role played by it. The optimal criterion followed in immediate reward and average reward function in the network.

$$R_{total} = \sum_{i=0}^{t} r_i \text{where r is finite t} \epsilon \{1, N\}$$ (4)

$$R_{total} = \frac{1}{t}\sum_{i=0}^{t} r_i \text{where t} \epsilon \infty$$ (5)

The Markov Decision Process with finite reward option and average reward option is used in this research to solve the energy overflow and energy outage problem in the network. The Q function average for the network is computed with the equation 6.

$$\pi^*(s) = argmax_a Q^*(s, a) \qquad\qquad (6)$$
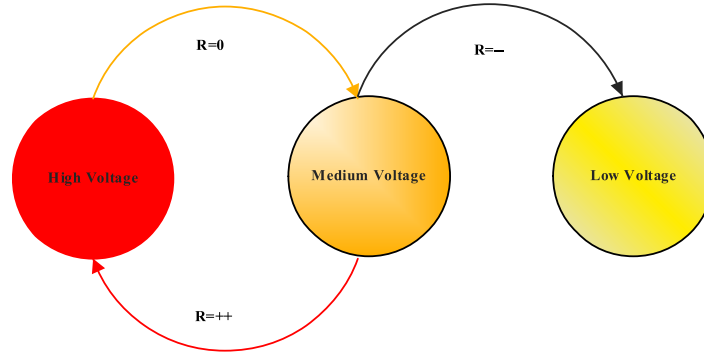

Figure 2   Battery reward status based on the role of action played by   the node as CH

Figure 2 illustrates the state machine model of the battery where the node with high voltage level receives positive reward if it acts as CH/router in the network.

Table 1 illustrates the reward provided to the corresponding node with respect to role and battery voltage aspects.

Table 1 Role/Voltage reward aspects table

Role/Voltage	High	Medium	Low
CH	R++	R=0	R--
CM	R=0	R++	R=0
Idle	R--	R=0	R++

Algorithm 1 describes the proposed RL-EM (RL based Energy Management Algorithm) for green network.

It is assumed that the time is divided into equal slots of duration τ. The Energy manager is executed at the beginning of each time slot. The MDP is computed for Arrival rate λ using energy model. The maximum desired packet rate P_{dd} for an array is obtained by computing the function of radio model, arrival rate λ and the MDP. The P_{dd} is assigned to all the nodes in the cluster. The P_{dd} is checked after time slot τ+1. If the desired packet rate delivery is greater than the packets actually transmitted P_{td} then the reward is given to individual node, else the reward is declined. The reward for a particular node is given as per the table 1. The node's battery voltage and role played by the node is considered for rewarding a particular node. The process is repeated at every time slot and the reward is assigned to the nodes based on the packet transmission rate

ALGORITHM -1   RL-EM ALGORITHM FOR GREEN WSN

Algorithm:

maximum desired packet rate P_{dd}

packets actually transmitted P_{td}

Voltage of the node (V)

Arrival rate λ

Input : Current Voltage of the Node (V_c), Distance (d), Radio model(RM)

Output: Optimal packet rate, CH scheduling

      Begin process for all clusters in network:

            for all nodes in cluster i at time slot τ

                compute

                    f(MDP <S, A, T, R>) for λ using energy model

                    f(P_{dd}) = f (Radio model (RM), Arrival rate (λ), MDP

            <S,A,T,R>, V)

                    obtain max f(P_{dd}) for array

                    assign P_{dd} to all nodes

                    check after $\tau+1$

                    if ($P_{dd} > P_{td}$)

                        grant reward;

                    else

                        null reward;

      end process

4 Results And Discussion

        The proposed RL-EM algorithm is simulated in Matlab based environment with 500 m length and width as region of Interest. Totally 100 nodes were considered for simulating and validating the algorithm. The energy of the nodes are differentiated with a) 80% of nodes are with 0.5 J, b) 20% nodes are with 0.7 J and c) 10% of the nodes are with 1J energy. The energy arrival rate is taken as $\lambda = 1$ throughout the simulation. Figure 3illustrates the node deployment inside the region of interest, they are deployed randomly.
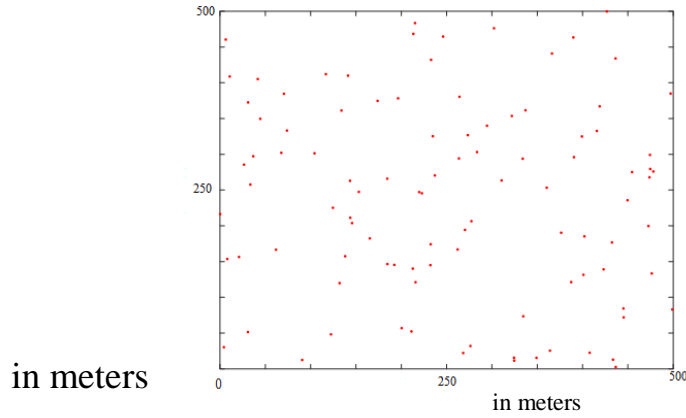
in meters

in meters

Figure 3 Node deployments inside the Region of Interest

Figure 4 illustrates the number of sleep nodes for 500 and 1000 rounds. The RL-EM shows reduced number of sleep nodes when compared to LEACH, SEP-M and Ant Colony Optimization algorithms. The proposed RL-EM algorithm provides approx. 0.5 times reduced sleep nodes when compared with standard LEACH protocol.
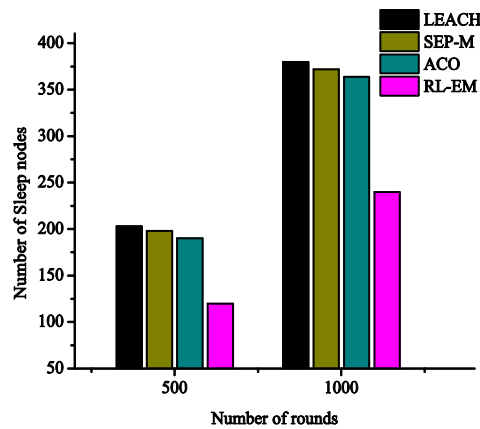

Figure 4 Sleep nodes in the network

Figure 5 shows the average energy overflow in the network for 500 and 1000 rounds. The energy overflow is reduced by 0.47 and 0.6 times in case of RL-EM when compared to LEACH, algorithm after 500 and 1000 rounds respectively. The RL-EM algorithm also outperforms recent protocols such as SEP-M and ACO approaches. The harvested energy is efficiently used in RL-EM such that the energy overflow is reduced.
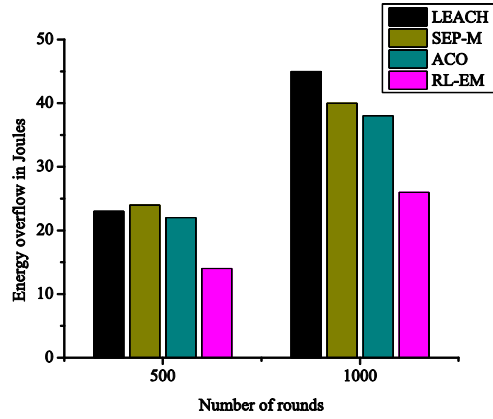
Figure 5 Average Energy over flow of the nodes

Table 2 elucidates the average reward point obtained by the network for 500 rounds and 1000 rounds. The network provides reduced sleep nodes in case of immediate reward function and reduced energy overflow rate for average reward function as given in MDP design section.

The results discussed throughout the section are on average reward function based RL-EM algorithm.

Table 2 Reward nature with network parameters

Parameter	500 rounds	1000 rounds
Reward - Average	200	280
Reward- Immediate	197	285
Sleep nodes(average reward)	180	240
Sleep Nodes(Immediate Reward)	169	232
Energy overflow (average reward) (J)	12	23
Energy overflow (immediate reward) (J)	14	26

Figure 6 represents the network throughput for 1000 rounds. The results show that the RL-EM provides increased throughput as maximum number of packets are sent to the sink.
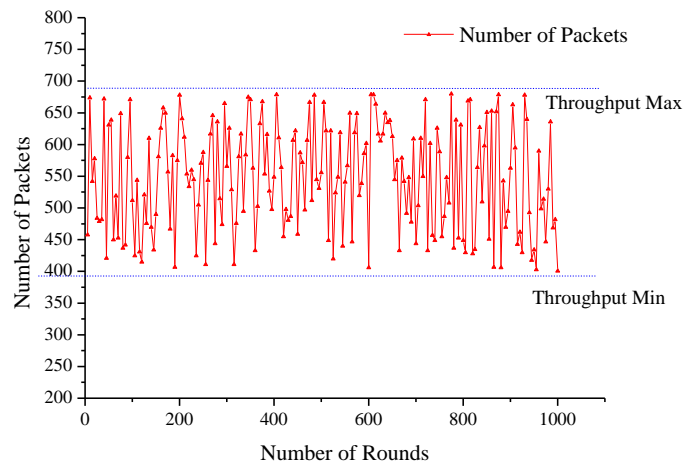
Figure 6 Network throughput

Figure 7 illustrates the energy map of the network. The present energy level of the nodes in the region of interest is shown. The node away from the sink shows high availability of energy.

The color legend (top to bottom): 1.000, 0.8875, 0.7750, 0.6625, 0.5500, 0.4375, 0.3250, 0.2125, 0.1000

X axis label: * 50 in meters

Y axis label: * 50 in meters

Annotation: **High energy zone far away from sink**
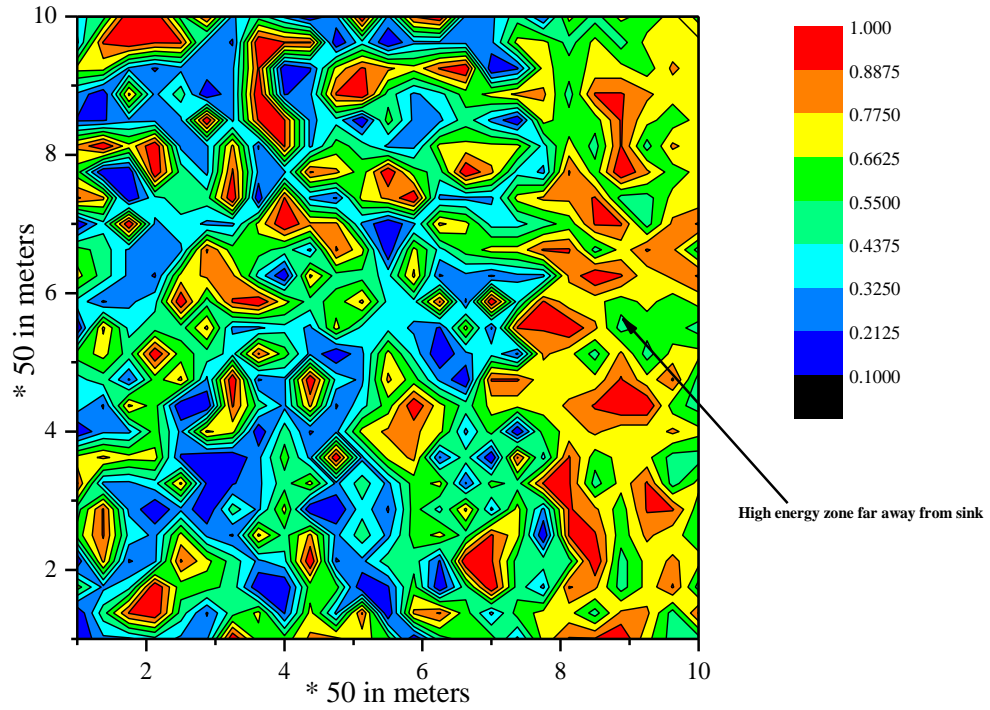
Figure 7   Total energy map in the Region of Interest in case of LEACH algorithm

Figure 8 illustrates the energy map of the proposed RL-EM algorithm, amount of energy overflow and energy outage is very low in the energy map. The energy inside the network is between 0 and 1.0. The nodes energy are equally distributed throughout the region of interest, the colour intensity determines the same. The nodes near to sink and far away from sink are treated well with respect to their load conditions and available energy levels.
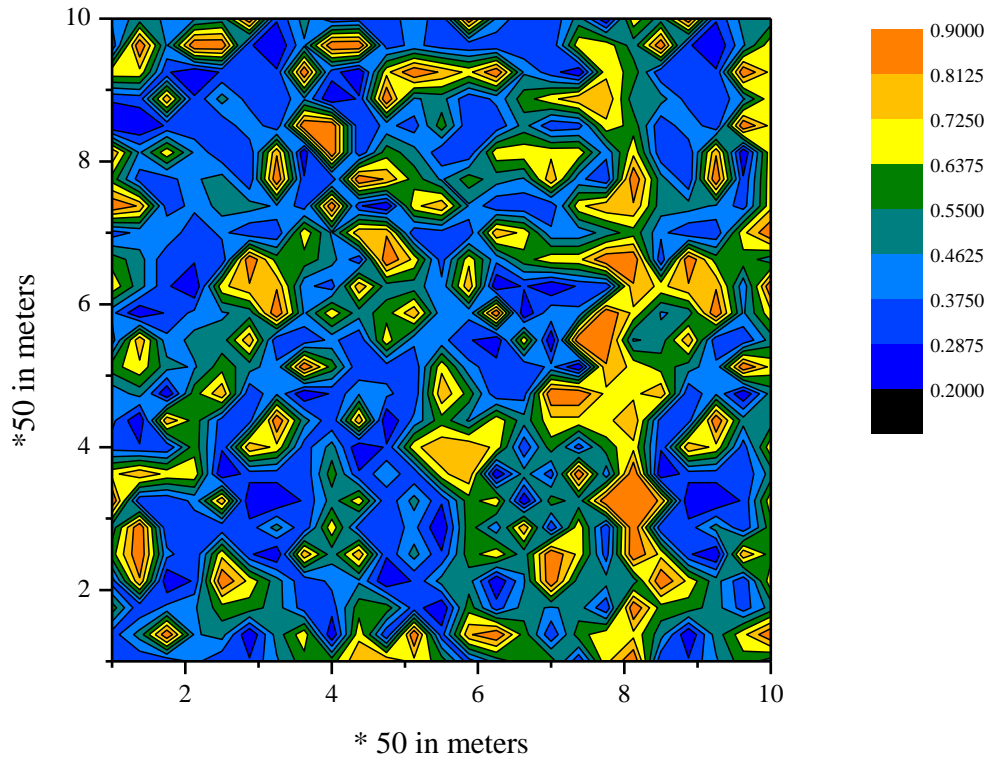
Figure 8   Energy map of the RL-EM algorithm

Figure 9 illustrates overall throughput map of the proposed algorithm within the region of Interest. All nodes inside the region of interest transmit packets within the assigned threshold and maximum level. This methodology reduces the energy overflow and energy outage problem in the network. The proposed RL-EM algorithm provides 1.07 times throughput when compared with the classical LEACH approach
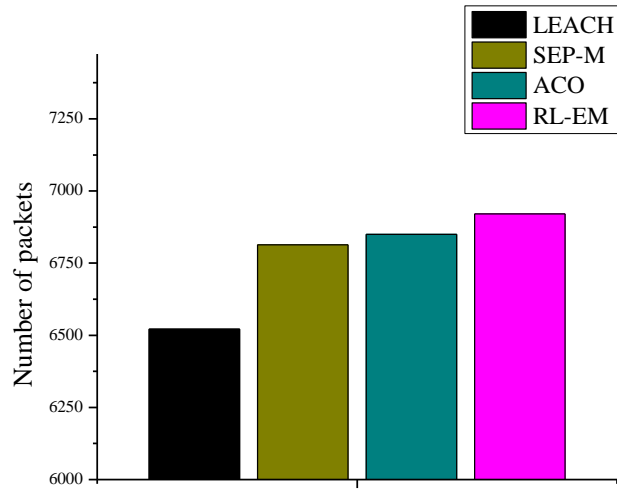
Figure 9   Over all throughput after 1000 rounds

Figure 10shows the drop packet of the network after 1000 rounds. The proposed RL-EM algorithm provides low drop packets when compared with all the other algorithms.
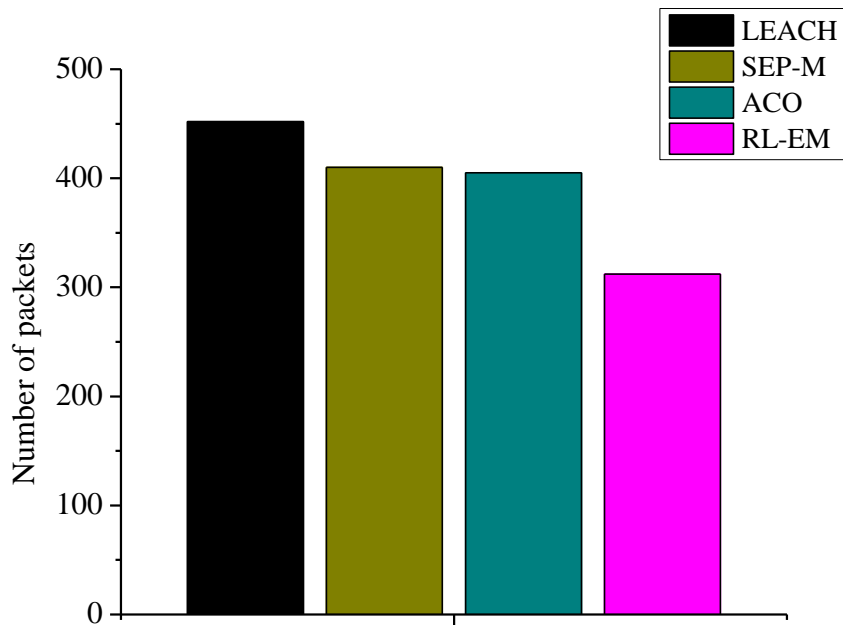

Figure 10   Drop packets after 1000 rounds

Conclusion

Node scheduling in this research work is done through RL approach and the same is compared with bio inspired algorithms such as ACO and with classical probability based algorithms like LEACH and SEP-M approach. The proposed work provides better energy management with 40% reduced sleep nodes and 60% reduced energy overflow in the network when compared with the LEACH benchmark protocol. The results provided by the proposed RL-EM algorithm prove equal load distribution. The research work also utilizes maximum resources to achieve its goal of successfully transmitting data to the sink. The RL-EM avoids energy hole and HOTSPOT problem in the network, which other compared algorithm does not solve. The algorithm seems to be a novel solution for energy management problem in the Wireless Sensor Network.

REFERENCES

[1] A. Sharma and S. Chauhan, "A distributed reinforcement learning based sensor node scheduling algorithm for coverage and connectivity maintenance in wireless sensor network," *Wirel.Netw.*, vol. 26, no. 6, pp. 4411–4429, Aug. 2020, doi: 10.1007/s11276-020-02350-y.

[2] D. K. Sharma, J. J. P. C. Rodrigues, V. Vashishth, A. Khanna, and A. Chhabra, "RLProph: a dynamic programming based reinforcement learning approach for optimal routing in opportunistic IoT networks," *Wirel. Netw.*, vol. 26, no. 6, pp. 4319–4338, Aug. 2020, doi: 10.1007/s11276-020-02331-1.

[3] T. Kaur and D. Kumar, "A survey on QoS mechanisms in WSN for computational intelligence based routing protocols," *Wirel.Netw.*, vol. 26, no. 4, pp. 2465–2486, May 2020, doi: 10.1007/s11276-019-01978-9.

[4] B. Post, S. Borst, and H. van den Berg, "A self-organizing base station sleeping and user association strategy for dense cellular networks," *Wirel.Netw.*, Sep. 2020, doi: 10.1007/s11276-020-02383-3.

[5] V. Mahima and A. Chitra, "Battery Recovery Based Lifetime Enhancement (BRLE) Algorithm for Wireless Sensor Network," *Wirel. Pers. Commun.*, vol. 97, no. 4, pp. 6541–6557, Dec. 2017, doi: 10.1007/s11277-017-4854-3.

[6] K. Liu and Q. Zhu, "Machine learning based adaptive modulation scheme for energy harvesting cooperative relay networks," *Wirel.Netw.*, vol. 26, no. 3, pp. 2027–2036, Apr. 2020, doi: 10.1007/s11276-019-01967-y.

[7] V. Saranya, S. Shankar, and G. R. Kanagachidambaresan, "Energy Efficient Clustering Scheme (EECS) for Wireless Sensor Network with Mobile Sink," *Wirel. Pers. Commun.*, vol. 100, no. 4, pp. 1553–1567, Jun. 2018, doi: 10.1007/s11277-018-5653-1.

[8] V. Saranya, S. Shankar, and G. R. Kanagachidambaresan, "Energy Efficient Data Collection Algorithm for Mobile Wireless Sensor Network," *Wirel. Pers. Commun.*, vol. 105, no. 1, pp. 219–232, Mar. 2019, doi: 10.1007/s11277-018-6109-3.

[9] P. Jayarajan, R. Maheswar, and G. R. Kanagachidambaresan, "Modified energy minimization scheme using queue threshold based on priority queueing model," *Clust.Comput.*, vol. 22, no. 5, pp. 12111–12118, Sep. 2019, doi: 10.1007/s10586-017-1564-2.

[10] P. Jayarajan, G. R. Kanagachidambaresan, T. V. P. Sundararajan, K. Sakthipandi, R. Maheswar, and A. Karthikeyan, "An energy-aware buffer management (EABM) routing protocol for WSN," *J. Supercomput.*, vol. 76, no. 6, pp. 4543–4555, Jun. 2020, doi: 10.1007/s11227-018-2582-4.

[11] P. Thirumoorthy, P. Kalyanasundaram, R. Maheswar, P. Jayarajan, G. R. Kanagachidambaresan, and I. S. Amiri, "Time-critical energy minimization protocol using PQM (TCEM-PQM) for wireless body sensor network," *J. Supercomput.*, vol. 76, no. 8, pp. 5862–5872, Aug. 2020, doi: 10.1007/s11227-019-03042-x.

[12] F. AitAoudia, M. Gautier, and O. Berder, "RLMan: An Energy Manager Based on Reinforcement Learning for Energy Harvesting Wireless Sensor Networks," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 408–417, Jun. 2018, doi: 10.1109/TGCN.2018.2801725.

[13] W. Xia, C. Di, H. Guo, and S. Li, "Reinforcement Learning Based Stochastic Shortest Path Finding in Wireless Sensor Networks," *IEEE Access*, vol. 7, pp. 157807–157817, 2019, doi: 10.1109/ACCESS.2019.2950055.

[14] H. Chen, X. Li, and F. Zhao, "A Reinforcement Learning-Based Sleep Scheduling Algorithm for Desired Area Coverage in Solar-Powered Wireless Sensor Networks," *IEEE Sens. J.*, vol. 16, no. 8, pp. 2763–2774, Apr. 2016, doi: 10.1109/JSEN.2016.2517084.

[15] C. Savaglio, P. Pace, G. Aloi, A. Liotta, and G. Fortino, "Lightweight Reinforcement Learning for Energy Efficient Communications in Wireless Sensor Networks," *IEEE Access*, vol. 7, pp. 29355–29364, 2019, doi: 10.1109/ACCESS.2019.2902371.

[16] W. Guo, C. Yan, and T. Lu, "Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 2, p. 155014771983354, Feb. 2019, doi: 10.1177/1550147719833541.

[17] R. C. Hsu, C.-T.Liu, and H.-L. Wang, "A Reinforcement Learning-Based ToD Provisioning Dynamic Power Management for Sustainable Operation of Energy Harvesting Wireless Sensor Node," *IEEE Trans. Emerg.Top.Comput.*, vol. 2, no. 2, pp. 181–191, Jun. 2014, doi: 10.1109/TETC.2014.2316518.

[18] A. Ortiz, H. Al-Shatri, X. Li, T. Weber, and A. Klein, "Reinforcement learning for energy harvesting point-to-point communications," in *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6, doi: 10.1109/ICC.2016.7511405.

[19] A. Ortiz, H. Al-Shatri, X. Li, T. Weber, and A. Klein, "Reinforcement Learning for Energy Harvesting Decode-and-Forward Two-Hop Communications," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 309–319, Sep. 2017, doi: 10.1109/TGCN.2017.2703855.

[20] K.-L. A. Yau, H. G. Goh, D. Chieng, and K. H. Kwong, "Application of reinforcement learning to wireless sensor networks: models and algorithms," *Computing*, vol. 97, no. 11, pp. 1045–1075, Nov. 2015, doi: 10.1007/s00607-014-0438-1.

[21] W. J. Guo, C. R. Yan, Y. L. Gan, and T. Lu, "An Intelligent Routing Algorithm in Wireless Sensor Networks Based on Reinforcement Learning," *Appl. Mech. Mater.*, vol. 678, pp. 487–493, Oct. 2014, doi: 10.4028/www.scientific.net/AMM.678.487.

[22] H. A. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, "Application of reinforcement learning to routing in distributed wireless networks: a review," *Artif. Intell. Rev.*, vol. 43, no. 3, pp. 381–416, Mar. 2015, doi: 10.1007/s10462-012-9383-6.

[23] M. Azharuddin and P. K. Jana, "A PSO Based Fault Tolerant Routing Algorithm for Wireless Sensor Networks," in *Information Systems Design and Intelligent Applications*, vol. 339, J. K.

Mandal, S. C. Satapathy, M. Kumar Sanyal, P. P. Sarkar, and A. Mukhopadhyay, Eds. New Delhi: Springer India, 2015, pp. 329–336.

[24] C. Vimalarani, R. Subramanian, and S. N. Sivanandam, "An Enhanced PSO-Based Clustering Energy Optimization Algorithm for Wireless Sensor Network," *Sci. World J.*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/8658760.

[25] R. Chaoming Hsu, C.-T.Liu, and W.-M. Lee, "Reinforcement Learning-Based Dynamic Power Management for Energy Harvesting Wireless Sensor Network," in *Next-Generation Applied Intelligence*, vol. 5579, B.-C.Chien, T.-P.Hong, S.-M. Chen, and M. Ali, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 399–408.

[26] F. Fraternali, B. Balaji, Y. Agarwal, and R. K. Gupta, "ACES -- Automatic Configuration of Energy Harvesting Sensors with Reinforcement Learning," *ACM Trans. Sens. Netw.*, vol. 16, no. 4, pp. 1–31, Sep. 2020, doi: 10.1145/3404191.