5th International Conference on Advanced Technologies
For Signal and Image Processing, ATSIP' 2020
September 02-05, 2020, Sfax, Tunisia

Efficient Routing Protocol for Wireless Sensor Network based on Reinforcement Learning

S.E. Bouzid*†‡, Y. Serrestou†, K.Raoof†, M.N.Omri*
* MARS Laboratory, LR 17ES05, University of Sousse, ISITCom, 4011, Hammam Sousse, Tunisia
† LAUM Laboratory of University of Le Mans, UMR CNRS 6613, Le Mans 72017 Cedex, France
‡ *salah_eddine.bouzid.etu@univ-lemans.fr*

***Abstract*—Wireless sensor nodes are battery-powered devices which makes the design of energy-efficient Wireless Sensor Networks (WSNs) a very challenging issue. In this paper, we propose a new routing protocol for WSN based on distributed Reinforcement Learning (RL). The proposed approach optimises WSN lifetime and energy consumption. This routing protocol learns, over time, the optimal path to the sink node(s). With a dynamic path selection, our algorithm ensures higher energy efficiency, postpones nodes death and isolation. We consider while routing messages the distance between nodes, available energy and hop count to the sink node. The effectiveness of the proposed protocol is demonstrated through simulations and comparisons with some existing algorithms over different lifetime definitions.**

***Index Terms*—WSN, Lifetime, Energy-efficiency, Routing protocol, Reinforcement learning.**

I. Introduction

Wireless Sensor Networks (WSNs) have a wide spectrum of applications such as monitoring, surveillance, domestic, etc. [1]–[4]. These networks are composed of hundreds of interconnected nodes. They consist of low power devices. These nodes are deployed over the Area of Interest which is, generally, a difficult-to-access area. Once energy sources of nodes are drained, the replacement of these sources is difficult or sometimes impossible. Therefore, it is better to design an energy-efficient algorithm to save energy and extend network LifeTime (LT). Hence, energy-efficiency is an important challenge for WSN and must be considered not only while network designing but also after its deployment [5]. In WSN, nodes dissipate energy while processing, sensing, transmitting or receiving packets. Experimental results confirm that communication is the greediest source of energy consumption. In such a heavy network, there are many alternatives to send a packet to the desired node. So, the determination of the best one, namely, the routing process is an important issue that can present an efficient solution for energy saving [6]. This problem of optimising routing problem while maintaining the quality of service is considered as an NP-Hard problem [7].

There have been tremendous works for the development of routing protocols in WSNs. However, several limitations still to be resolved. Traditional routing protocols assume homogeneous devices. In this case, old approaches such as the minimum hop route, maximum available power, minimum energy route works may work properly. Besides in heterogeneous network, they will lead to low network lifetime and low energy-efficiency. That is why the heterogeneity of nodes in terms of communication and energy capacities must be considered. Most of the proposed protocols are either flat or hierarchical one. Regarding to optimisation of LT, the most adopted LT definition is the time when the first node is dead [8], [9]. Whereas, this time is not very important because when a single node is dead, the entire network continues to work.

According to these limitations, we aim to propose a new routing protocol based on reinforcement learning for LT optimisation. Nodes will learn over time and will be self-configured. No prior knowledge of the topology is needed. They will continuously learn and update their information in order to make up-to-date and intelligent path choices. In order to cover different applications, our goal is to optimise LT in three aspects and in terms of energy efficiency.

The remainder of this paper is organized as follows: In section II, we present an overview of the routing problem. In Section III, we present our proposed protocol, named Reinforcement Learning for LT Optimisation (*R2LTO*). Performance evaluation is discussed, and obtained results are compared to other works in section IV. Finally, section V concludes the paper and presents future work.

II. Overview of routing problem

In literature, different methods of routing protocol were proposed [10]–[12]. JA Boyan and ML Littman proposed a hop-by-hop routing algorithm based on Q-learning, called Q-routing [13]. The goal of this work is increasing packet delivery and minimising delivery time. This protocol suffers from freshness. In the case where a route was not selected for a long period of time, the agent has no update of its current condition. Its knowledge is limited. As a result, it may become unreliable learning process. Adaptive Reinforcement Based Routing (ARBR) [14] is a routing protocol for delay tolerant networks. Cooperatively, nodes work together to choose the forwarding node based on node's knowledge, time statistics and network congestion. However, this type of protocol is only dedicated to some specific scenario. Multi-agent Reinforcement Learning based on QoS Routing Protocol (MRL-QRP) [15] is a routing protocol with QoS support in WSNs. What differs this approach from others is the collaboration between agents when they compute their Q-values. Implementing MRL-QRP in practice is discussed in [16] and judged to be difficult. Among routing protocols that consider network LT

978-1-7281-7513-3/20/$31.00 ©2020 IEEE

metrics, energy-aware routing (EAR) [17] has been proven to be an efficient routing protocol. To avoid using the same minimum energy path, EAR saves different paths that link the sender to the destination. Each time, it chooses one path from the saved ones following a parabolic law. Whereas this protocol provides interesting results, it only takes into account energy consumption. Unlike EAR, balanced energy-efficient routing (BEER) considers in addition to energy consumption, remaining energy. EAR and BEER both suffer from choosing the path according to the routing table initially built. However, this table does not mirror the current state of the network. More recently, [10] a reinforcement learning based routing framework, named RLBR, was proposed. RLBR considers distance, remaining energy and hop count to choose the next forwarder. Despite that RLBR presents results better than Q-routing and BEER, it has some limits. While selecting next forwarder, RLBR does not consider nodes with higher hop counts or greater distance than the current node as a candidate node. However, this restriction allows minimizing energy consumption but not maximizing LT. After having studied the above-cited approaches, we aim to propose a new approach that overcomes these limitations.

With the goal to optimise LT in routing problem, many definitions are adopted [9], [18], [19]. These definitions vary depending on the application and on the network topology. In our work, we retain the following LT definitions:

Definition 1: The Time until the first dead node is revealed. A node is considered dead if it depletes its energy source.

Definition 2: The Time until the first isolated node is revealed. An isolated node is a node that still has energy but has no path to sink node.

Definition 3: The Time until no more packet can be delivered.

Added to LT metric, energy-efficiency is another important metric that defines the performance of a network [10]. It is defined as follows:

Definition 4: The ratio of the number of packets to total energy consumption.

In order to cover more applications, the proposed approach must take into consideration different definitions and ensure high energy-efficiency.

III. New routing algorithm

Reinforcement Learning (RL) is the problem faced by an agent that must learn behaviour through trial-and-error with a dynamic environment [20]. The agent selects an action among a set of possible actions according to what he has already learned. Then, it receives a reward. By a learning process, the policy of selecting the optimal actions is constructed. Motivated by these aspects, RL represents an efficient solution and a perfect method to solve distributed real-time decisions problems which is the case in routing problems.

Agents in our routing problem are wireless nodes and actions are choosing the next forwarder node. In WSN, a sensor node detects an event or receives a packet from one of its neighbours then it has to choose from his neighbours, the next forwarder. This action depends on the states of neighbours and the reward function.

Our routing protocol consists of two processes; the discovery process and the routing with a continuous learning process.

A. Discovery process

To have a holistic approach that optimises energy consumption and LT of WSN with no prior knowledge of the topology, a discovery process is developed. Through this process, nodes will learn over time and will be self-configured. The first step is to initialize different nodes. The sink starts by sending a notification packet to his neighbours. A notification packet received by a node is formed by its ID and the previous node's information (node ID, location, remaining energy and hop count). While receiving a notification packet, each node extracts the sender information from this packet. It calculates the path quality (Q-value) of the sender according to equation 1, and its hop count to the sink according to equation 2. Then, the node encodes its own information in the notification packet and broadcasts it to all its neighbours and so on.

$$Q(sender) = \frac{Energy(sender)}{Hop(sender)} \tag{1}$$

$$Hop(current) = Hop(sender) + 1 \tag{2}$$

This step allows each node to initialize its neighbouring table. The informations saved for each neighbour in this table are:

- *ID*: Neighbour's unique identifier.
- *Location*: Neighbour's coordinate.
- *Energy*: Neighbour's remaining energy.
- *Hop*: Neighbour's hop count.
- *Q-value*: Neighbour's path quality to sink.

B. Routing and learning process

After the discovery process, routing the data packet with a continuous learning process starts. A node in the network can either receive or generate a data packet. Data packet, in addition to the same attributes of a notification packet, has Q-value, next node ID and data attributes.

In the case where a node receives a data packet with *Next ID* value different from its own *ID*, it updates its neighbouring table and, then the packet is dropped. With this overhearing mechanism, each node has an up-to-date neighbouring table, and so the routing process is in accordance with the real conditions of the network. Otherwise, it chooses the best next candidates from its neighbouring table. If a sink node is reachable, the packet will be sent directly. If not, it nominates the next forwarder. The next forwarder should not be an isolated node or dead one. If no neighbour has this condition, the packet will be dropped and, the current node is considered as an isolated node. Its hop count and Q-value are then set to 0. In the case where there are many possible candidates, the next forwarder is chosen based on Q-values.

Let i and j be respectively, the current node and its neighbour. Node i calculates the quality of path to sink via neighbour j according to equation 3.

$$Q_{t+1}(i,j) = (1-\alpha)\ Q_t(i,j) + \alpha(R(i,j) + Q(j)) \quad (3)$$

where:

- α: the learning rate,
- $Q(i,j)$: the estimated path quality from node i to the sink via node j,
- $R(i,j)$: the reward given to i if we choose to send a packet via j,
- $Q(j)$: path quality from node j to sink,

We denote that $Q(j)$ can be retrieved from the neighbour table. $R(i,j)$ is the reward function. This latter should be chosen carefully because it highly influences the network lifetime and energy consumption. It must be proportional to remaining energy and is inversely proportional to hop count and to the needed energy to send a packet (eventually distance). Thus, The proposed reward function is as follows:

$$R(i,j) = \frac{Energy(j)}{\left(\frac{Tx(i,j)}{Tx(rang)}\right)\ hop(j)} \quad (4)$$

$Energy(j)$ and $hop(j)$ are respectively remaining energy of node j and hop count to the sink via this node. These two informations are retrieved from the neighbour table of the node i. $Tx(i,j)$ is the estimation of the required energy to send a packet from the node i to the node j after adjusting it transmit power. $Tx(rang)$ is the highest energy consumption.

After calculating the new Q-value for each neighbour node by equation 3, the sender i chooses from these values the node with maximum Q-value and then it updates its Q-value and hop count (equation 5).

$$\begin{cases} Q(i) & = \max\limits_{j\ \in\ neighbours} Q(i,j) \\ hop(i) & = hop(neigh) + 1 \end{cases} \quad (5)$$

where *neigh* is the chosen node. By the next, the current node updates the packet header by putting its own informations and send the packet to the next forwarder. Other neighbours overhear this packet, and then they update their neighbour tables.

As an energy model, we adopt the first-order model which is an acceptable model for WSN [10]. Consumed energy for transmitting or receiving a packet is calculated by equation (6).

$$\begin{cases} E_{Tx}(n,d) & = E_{elec}.n + \epsilon_{amp}.n.d^{\eta} \\ E_{Rx}(n) & = E_{elec}.n \end{cases} \quad (6)$$

where n and d are respectively packet length in bits and distance separating the receiver and the sender. E_{elec} and ϵ_{amp} are the electronic energy and the transmitter amplifier. η and $d(i,j)$ are the path loss exponent and distance between two nodes.

IV. Simulation results and discussions

In order to measure the performance of our algorithm, we will evaluate it via different scenarios. Moreover, we will compare it to some other approaches to prove its efficiency. Approach evaluations will be based on the three retained definitions of LT (Definition. 1 - 3), and in terms of packet delivery, total consumed energy and energy-efficiency 4. In the following, we denote by N and CE respectively the number of packet and the total consumed energy. Energy-efficiency, denoted by E is calculated as follows:

$$E = \frac{N}{CE} \quad (7)$$

For this purpose, we select the routing protocol "*Q-routing*" and the recent "*RLBR*" protocol. Our protocol is coded in python under "PyCharm" development environment and executed on a PC with an Intel Core 7-5500U, 2.4 GHz processor and 8 GB of RAM. Simulation parameters are represented in Table I.

TABLE I: Simulation parameters.

Parameters	Values
Number of nodes	100
Deployment space	100m x 100m
Initial energy	0.5
E_{elec}	50nJ/bit
ϵ_{amp}	100pJ/bit/m^2
Packet generation rate	$1/$ episode
Packet size	512 bits
learning rate	0.5

We first generate 100 node placements. The same configuration are used for different approaches. With a fixed seed value, we generate the same event. This events can be for example motion events detected by nodes in a monitoring application. We start by setting the communication range of different nodes to 30m. For the three routing protocols, we run the simulator until there no more packet delivered to sink (Third definition of LT). Figure 1 compares the number of alive nodes in the function of episodes.

It is apparent from Fig. 1 that our protocol succeeded to keep a higher number of alive nodes over episodes with a small shift. This is due to a low number of hops to the sink with a communication radius equals to 30m in an environment of 100m x 100m. Changing the forwarder node do not highly affect performances. Whatever forwarder is selected, it will be able to reach the sink node in some hops. Table II presents the obtained results for the three approaches. According to

TABLE II: Obtained results (Communication rang = 30m)

Approach	LT Definition			CE	N	E
	First	Second	Third			
R2LTO	2249	3101	3796	35,407	3645	102,943
RLBR	2235	2970	3403	35,139	3245	92,34
Q-routing	1647	2964	3112	45,323	2994	66,058

Table II, R2LTO performs better than other protocols in terms

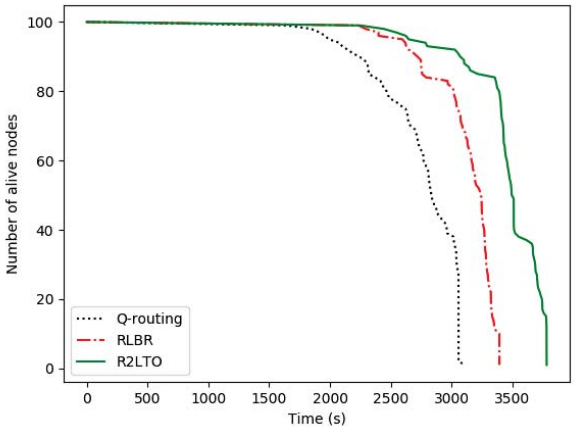

Fig. 1: Variation of the number of living nodes as a function of time (30 m)

of different LT definitions, total energy consumed and energy efficiency.

Additional simulations are needed to fairly evaluate different approaches. We set communication range to 15m and then 10m to increase routing problem complexity. With this configuration, finding the optimal path is harder. This is because a high number of hops are needed to reach the sink and then more decisions (choosing next forwarder) to be taken. A bad choice can lead to more energy consumption and delays packet delivery. Other parameters are the same as defined in Table I. Fig. 2 and Fig. 3 illustrate respectively the number of alive nodes for communication range equals to 15m and 10m.

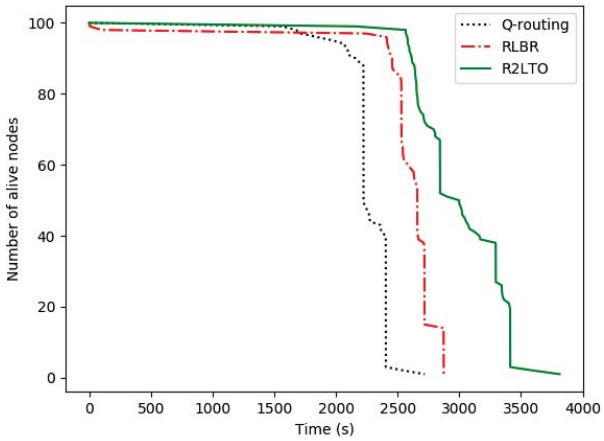

Fig. 2: Variation of the number of living nodes as a function of time (15 m)

The results, as shown in Fig. 2 and Fig. 3, indicate that even with low communication range, our protocol still performs

TABLE III: Obtained results (Communication rang = 15m)

Approach	LT Definition			CE	N	E
	First	Second	Third			
R2LTO	2264	2567	3814	20,135	3669	182,218
RLBR	2245	7	2887	20,798	2759	132,653
Q-routing	1567	2101	2721	33,886	2615	77,169

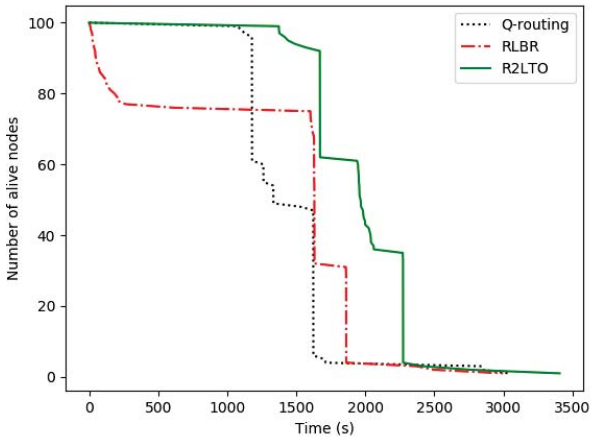

Fig. 3: Variation of the number of living nodes as a function of time (10 m)

better than *Q-Routing* and *RLBR* protocols. In both Tables IV and III, it is clear that RLBR appeared to be ineffective for many network topologies. We can notice that their used strategy works well in many node organizations and may allow energy saving. But, restricting nodes that have distances to sink node higher than the current one from candidates list, makes these nodes isolated at early time which explains having a bad network LT according to definition 2. Routing strategy in *Q-Routing* aims to test all undiscovered neighbours. After discovering delivery time for each neighbour, the node with the lowest delivery time will be chosen at every packet delivery. Accordingly, this node will be drained out quickly. Moreover, having undiscovered neighbours makes learning process longer. Overall, these results indicate that our approach surpasses *Q-Routing* and *RLBR* in different aspects of LT and energy-efficiency. Via simulations and analytics, our approach presents an efficient routing protocol. A real implementation of this protocol is required to study hardware requirements of different nodes such as memory and computing time.

V. Conclusion and Future Works

In this paper, we proposed a reinforcement learning for lifetime optimisation, named R2LTO, that optimises energy usage by choosing the optimal path to the sink in a dynamic and intelligent way. R2LTO outperforms other routing algorithms not only in terms of different LT definitions but also in term of energy efficiency. Our protocol does not require a

TABLE IV: Obtained results (Communication rang = 10m)

Approach	LT Definition			CE	N	E
	First	Second	Third			
R2LTO	1374	1375	3408	14,667	3275	223,288
RLBR	1602	9	2998	12,939	2863	221,252
Q-routing	1071	1180	3036	27,942	2933	104,964

priori knowledge of the network. Via a discovery process, each node explores its neighbours. Then, the learning process allows each node to select the optimal next forwarder according to remaining energy, needed energy (distance) and hop count. This selection strategy allows avoiding node isolations and balance energy consumptions of all sensor nodes and ensuring packet delivery. With this overhearing process, network status is always up-to-date. We have validated the performance of R2LTO through different simulations. R2LTO outperforms performance over Q-Routing and RLBR in terms of the percent of alive nodes, LT under different three definitions, the number of packets delivery and energy efficiency. In future, we intend to implement this protocol under real wireless sensor network.

ACKNOWLEDGEMENT

This work was supported by the Tunisian Ministry of Higher Education and Scientific Research.

REFERENCES

[1] Kasim Al-Aubidy, A.W. Al Mutairi, and Ahmad Derbas. Real-time healthcare monitoring system using wireless sensor network. *International Journal of Digital Signals and Smart Systems*, 1:26, 01 2017.

[2] Jacques Bahi, Wiem Elghazel, Christophe Guyeux, Mourad Hakem, Kamal Medjaher, and Noureddine Zerhouni. Reliable diagnostics using wireless sensor networks. *Computers in Industry*, 104:103 – 115, 2019.

[3] S.E. Bouzid, M. Mbarki, C. Dridi, and M. N. Omri. Smart adaptable indoor lighting system (SAILS). In *2019 IEEE International Conference on Design Test of Integrated Micro Nano-Systems (DTS)*, pages 1–6, April 2019.

[4] S.E. Bouzid, Y. Serrestou, K. Raoof, M. Mbarki, M. N. Omri, and C. Dridi. Wireless Sensor Network Deployment Optimisation based on Coverage, Connectivity and Cost Metrics . *International Journal of Sensor Networks*, 2020.

[5] S.E. Bouzid, Y. Serrestou, K. Raoof, M. Mbarki, M. N. Omri, and C. Dridi. Wireless sensor and actuator network deployment optimization for a lighting control. *International Journal of Computer and Communication Engineering*, 9(2):15, 2020.

[6] Zoubir Mammeri. Reinforcement Learning Based Routing in Networks: Review and Classification of Approaches. *IEEE Access*, 7:55916–55950, 2019.

[7] Zheng Wang and J. Crowcroft. Quality-of-service routing for supporting multimedia applications. *IEEE Journal on Selected Areas in Communications*, 14(7):1228–1234, Sep. 1996.

[8] J. W. Jung and M. A. Weitnauer. On using cooperative routing for lifetime optimization of multi-hop wireless sensor networks: Analysis and guidelines. *IEEE Transactions on Communications*, 61(8):3413–3423, August 2013.

[9] C. G. Cassandras, T. Wang, and S. Pourazarm. Optimal routing and energy allocation for lifetime maximization of wireless sensor networks with nonideal batteries. *IEEE Transactions on Control of Network Systems*, 1(1):86–98, March 2014.

[10] Wenjing Guo, Cairong Yan, and Ting Lu. Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing. *International Journal of Distributed Sensor Networks*, 15(2):1550147719833541, 2019.

[11] Ting Lu, Guohua Liu, and Shan Chang. Energy-efficient data sensing and routing in unreliable energy-harvesting wireless sensor network. *Wireless Networks*, 24(2):611–625, Feb 2018.

[12] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, pages 10 pp. vol.2–, Jan 2000.

[13] Justin A Boyan and Michael L Littman. Packet routing in dynamically changing networks: A reinforcement learning approach. In *Advances in neural information processing systems*, pages 671–678, 1994.

[14] A. Elwhishi, Pin-Han Ho, K. Naik, and B. Shihada. Arbr: Adaptive reinforcement-based routing for dtn. In *2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 376–385, Oct 2010.

[15] Xuedong Liang, I. Balasingham, and Sang-Seon Byun. A multi-agent reinforcement learning based routing protocol for wireless sensor networks. In *2008 IEEE International Symposium on Wireless Communication Systems*, pages 552–557, Oct 2008.

[16] H. Yetgin, K. T. K. Cheung, M. El-Hajjar, and L. H. Hanzo. A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Communications Surveys Tutorials*, 19(2):828–854, Secondquarter 2017.

[17] R. C. Shah and J. M. Rabaey. Energy aware routing for low energy ad hoc sensor networks. In *2002 IEEE Wireless Communications and Networking Conference Record. WCNC 2002 (Cat. No.02TH8609)*, volume 1, pages 350–355 vol.1, March 2002.

[18] M. Najimi, A. Ebrahimzadeh, S. M. H. Andargoli, and A. Fallahi. Lifetime maximization in cognitive sensor networks based on the node selection. *IEEE Sensors Journal*, 14(7):2376–2383, July 2014.

[19] H. Salarian, K. Chin, and F. Naghdy. An energy-efficient mobile-sink path selection strategy for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 63(5):2407–2419, Jun 2014.

[20] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, May 1996.