Wireless Networks (2022) 28:2669–2693
https://doi.org/10.1007/s11276-022-02997-9

ORIGINAL PAPER

Energy efficient and reliable routing in wireless body area networks based on reinforcement learning and fuzzy logic

Wenjing Guo[1] · Yiran Wang[1] · Yanglan Gan[1] · Ting Lu[1]

Accepted: 2 May 2022 / Published online: 24 May 2022

Abstract

In Wireless Body Area Networks (WBANs), on the one hand, the energy of nodes is limited. On the other hand, the network topology often changes due to human movement or posture changes. Unstable network topology is easy to cause packet loss, and packet loss will cause inaccurate data collection. Therefore, how to effectively use energy to transmit data reliably becomes a key issue. For this problem, we propose an optimized routing protocol namely Energy Efficient and Reliable Routing based on Reinforcement Learning and Fuzzy Logic (EERR-RLFL). In EERR-RLFL, considering the heterogeneity of nodes in WBANs, we first establish a node rank division mechanism, by which sensor nodes are divided into different ranks from three aspects. Each rank is considered to be one of the factors that affect the link quality. Then, we propose the Fuzzy-Logic-based Link Quality Evaluation (FLLQE) algorithm. It makes use of the fuzzy evaluation method of fuzzy logic and considers the comprehensive influence of multiple factors to evaluate the link quality between two nodes, which will provide reference for routing path selection. In the process of data transmission, based on the FLLQE algorithm, we use a hybrid data transmission mode, in which the time when a forwarding node is needed is first determined, and then the Reinforcement Learning algorithm is used to select the global optimized routing path. Simulation results show that EERR-RLFL outperforms Single Hop Transmission and Optimized Cost Effective and Energy Efficient Routing in terms of network lifetime, packet loss ratio and energy efficiency.

Keywords Wireless body area networks · Energy efficient · Reliable routing · Reinforcement learning · Fuzzy logic · Link quality evaluation

1 Introduction

With the development of wireless communication technology, embedded computer technology and biosensor technology, Wireless Body Area Networks (WBANs) emerges as a branch of Wireless Sensor Networks (WSNs). In WBANs, sensors are placed on the body surface or implanted into the body to collect and transmit such physiological data as blood pressure, heart rate, body temperature, Electrocardiogram (ECG) and Electroencephalogram (EEG) to the sink node. The sink node eventually sends the data to the hospital monitoring center, relatives' phones or community service center. By such a way, WBANs can continuously monitor the human health. Now the problem of global aging is becoming increasingly prominent. For the elderly, their health is of particular concern. With the growth of age, some functions of the elderly body begin to weaken. Many elderly people suffer from hypertension, hyperglycemia, coronary heart disease and other diseases. The health of the elderly is a top priority for their children. However, young people need to work and live, and can not always accompany their parents. The health of parents often becomes the most concerned thing for their children. Especially for the elderly living alone, their health is even more worrying. In some places, there are even such tragedies that the death of the elderly is not discovered until many days later. Therefore, a technology is needed to remotely monitor the health status of the elderly. In addition, for patients in hospitals, such a technology is also needed to enable medical workers to monitor their physical conditions and understand their

✉ Wenjing Guo
wjguo@dhu.edu.cn

[1] School of Computer Science and Technology, Donghua University, Shanghai 201620, China

needs at any time. WBANs have become the effective way to solve these problems.

In WBANs, the most challenging problem is the limited energy resources. The sensor nodes in WBANs are smaller than that in conventional WSNs, which limits their battery capacity. For sensor nodes implanted in the human body, it is very difficult to replace the battery, which is basically impossible. Once the energy of a sensor node is exhausted, this node will have no ability to work and the associated link will be invalid. For a sensor node, the wireless communication module used to transmit data consumes the most energy. Therefore, WBANs must use an energy-efficient way to transmit data to ensure the battery life and maximize the network lifetime. In addition, in WBANs, since the sensor nodes are attached to the human body, the movement of the human body or the change of posture will cause the change of network topology. This unstable network topology can easily cause packet loss. In WBANs, the deployment of sensor nodes is not redundant. Packet loss will lead to inaccurate data received by the monitoring center, which will affect the monitoring effect of the human body, and even more, will endanger human life. Therefore, in WBANs, how to efficiently use energy and reliably transmit data to the sink node becomes a key issue.

In the early research of WBANs, the single-hop transmission method was usually adopted, that is, the sensor nodes sent the collected data directly to the sink node. However, for the sensor nodes far away from the sink node, such a single-hop transmission method will consume a lot of energy. This is because the signal transmission path of the body channel is attenuated. Referring to the path loss model of the human body communication [1], in free space, the basic path loss rate of data propagation between nodes is 2, the path loss rate of data propagation in the line of sight range is 3 to 4 and that in non-line of sight range is 7. It can be seen that if the sensor node is on the back side of the human body and the sink node is on the front side, the path loss rate is as high as 7. If they communicate directly, a large transmit power is required, which will consume a lot of energy and may cause unacceptable electromagnetic radiation to the human body. For this case, the multi-hop transmission method has obvious advantages, in which the forwarding of other nodes is used to reduce the energy consumption. This requires a routing protocol that looks for the path from the source node to the sink node and sends the data along the selected path. Therefore, in WBANs, it is necessary to design an effective routing protocol to complete data transmission.

In the current study, there are some routing protocols specially designed for WBANs, which can be classified as protocols of temperature-based, cluster-based, cross-layer-based and cost-function-based [2]. The temperature-based routing protocols take temperature as the primary factor to deal with the routing problem. Some biomedical sensors are implanted into the human body in WBANs. The generation of radio signals during wireless communication will cause an increase in temperature, which can damage temperature-sensitive human tissue. Therefore, these kind of protocols aim to minimize the bio-thermal effects. Typical representatives of such protocols are Thermal Aware Routing Algorithm (TARA) [3], Least Temperature Routing (LTR) [4], Trust and Thermal-aware Routing Protocol (TTRP) [5] and so on. The cluster-based routing protocols divide sensor nodes in a WBAN into multiple clusters, and the representative protocols are A Self-organization Protocol for Body Area Networks (Anybody) [6], Hybrid Indirect Transmission (HIT) [7], Energy-efficient and Reliable Routing Scheme (ERRS) [8], etc. The cross-layer-based routing protocols improve the overall performance of the network by simultaneously handling the problems of the network layer and the MAC layer. Timezone Coordinated Sleeping Scheduling (TICOSS) [9], Wireless Autonomous Spanning Tree Protocol (WASP) [10] and Priority-based Cross Layer Routing Protocol (PCLRP) [11] are such protocols. Besides these, there is a type of routing protocol in WBANs that first defines a cost function and then chooses the forwarder node based on the cost function. Such protocols include Improved Stable Increased-throughput Multi-hop Link Efficient (iM-SIMPLE) [12], Optimized Cost Effective and Energy Efficient Routing (OCER) [13], Energy Efficient Thermal and Power Aware Routing (ETPA) [14], Cooperative Link Aware and Energy Efficient Protocol for Wireless Body Area Networks (Co-LAEEBA) [15] and Energy Efficient Routing Scheme (EERS) [16]. We will discuss these protocols in Sect. 2. In designing routing protocols for WBANs, it is still a challenging issue to balance energy efficiency and data transmission reliability.

In this paper, to ensure the reliability and energy efficiency of data transmission, we propose an optimized routing protocol namely Energy Efficient and Reliable Routing based on Reinforcement Learning and Fuzzy Logic (EERR-RLFL). In EERR-RLFL, firstly, the link quality between two nodes is evaluated based on the fuzzy evaluation method of fuzzy logic [17]. Through this step, the probability that the link between two nodes is in various states can be obtained. Then, in the process of data transmission, based on the link quality evaluation, a hybrid data transmission mode is adopted. The network topology of a WBAN changes dynamically. In order to ensure the accuracy of the data received by the monitoring center, an effective data transmission method is needed. If all nodes send the collected data directly to the sink, it is likely to cause packet loss and high energy consumption. On the contrary, if a multi-hop data transmission mode is always adopted, it will cause unnecessary energy consumption.

Therefore, EERR-RLFL adopts a hybrid data transmission mode, in which it is first judged whether a forwarding node is required. When a forwarding node is needed, the Reinforcement Learning (RL) [18] algorithm is used to select the routing path. RL is a sub-field of machine learning, which mainly deals with how agents take actions to obtain the maximum long-term benefits. This is consistent with our goal: how to choose the routing path to maximize the reliability and energy efficiency of data transmission. The RL algorithm has strong robustness, it can still guarantee the reliable transmission of data even in the case of topology changes, and it requires less overhead. Therefore, with the advantages of the RL algorithm, the routing selection can be optimized continuously with less additional energy consumption. We summarize the main contributions of this paper as follows:

(1) We establish a node rank division mechanism. In this mechanism, sensor nodes are divided into different ranks from three perspectives. Each rank is considered to be one of the factors that affect the link quality, which can balance the energy consumption to a certain extent, protect important nodes, and make full use of the nodes that have completed the data collection work.
(2) We propose the Fuzzy-Logic-based Link Quality Evaluation (FLLQE) algorithm. This algorithm makes use of the fuzzy evaluation method of fuzzy logic, and considers the comprehensive influence of multiple factors to evaluate the link quality between two nodes, which can provide reference for probabilistic path selection when transmitting data.
(3) We propose the EERR-RLFL protocol. It adopts a hybrid data transmission mode and uses the RL algorithm to select the routing path when forwarding nodes are needed. In addition to the general advantages of RL-based routing protocol, EERR-RLFL also takes some special optimization measures, which help to decrease the total energy consumption, balance the energy consumption between nodes and enhance the reliability of data transmission.

The rest of the paper is organized as follows. In Sect. 2, we present the related works. Then, in Sects. 3 and 4, we introduce the proposed protocol in detail, and verify its performance in Castalia. Finally, Sect. 5 concludes the paper.

2 Related works

WBAN is a body-centric WSN, and is also considered to be an extension of the Wireless Personal Area Network (WPAN). As early as 1998, the IEEE organization established the 802.15 working group to work on the standardization of the physical layer and media access layer of WPANs. The IEEE 802.15 working component is composed of six task groups. The sixth task group was established in February 2007 and began to develop the IEEE 802.15.6 standard. In 2012, the IEEE organization officially approved the IEEE 802.15.6 communication standard for WBANs. The IEEE 802.15.6 standard specifies the standards for the physical layer and media access layer of WBAN, and the routing protocol defines the communication specification of the network layer based on the IEEE 802.15.6 standard. Some scholars have designed routing protocols specifically for WBANs. These routing protocols are generally divided into the following four categories:

(1) Temperature-based routing protocols. TARA [3] is the first routing protocol to deal with the overheating problem of implanted biosensors in WBANs. In TARA, each node monitors the number of data packets sent and received by neighbor nodes and calculates the energy consumption of communication to obtain the temperature change of neighbor nodes. Then, data transmission in areas with excessive temperature are reduced to avoid human burns. In fact, temperature can also reflect the problem of energy consumption. Thus, this protocol can balance energy consumption between nodes to a certain extent. But it does not take optimization measures to reduce the energy consumption and the Packet Loss Rate (PLR). LTR [4] is a minimum temperature routing protocol. In this protocol, each node obtains the temperature of neighbor nodes by discovering the communication activities of these nodes, and selects the node with the least temperature as the next hop to forward data. Like TARA, LTR does not take the optimization of decreasing energy consumption and PLR. Moreover, LTR uses greedy algorithm to select forwarding nodes, which will result in local optimization rather than global optimization. Optimum Path Optimum Temperature (OPOT) [19] is also a temperature-based routing protocol, in which a temperature threshold is set, and only nodes whose temperature are lower than the threshold will be considered by the routing protocol. At the same time, in order to avoid the loss of important packet, OPOT allows packets with critical priority to pass through the hot-spot region temporarily. Thus, OPOT can balance energy consumption between nodes and reduce the PLR to a certain extent. But when there are too many hot-spot regions, the routing protocol only considers a few nodes to establish the routing path, which may lead to hypo-optimum path and

increase of energy consumption. Other temperature-based routing protocols include Mobility-supporting Adaptive Threshold-based Thermal-aware Energy-efficient Multi-hop Protocol (M-ATTEMPT) [20] and TTRP [5]. These protocols all consider temperature as the basic weight to achieve the balance of communication between nodes, but their shortcomings are the lack of comprehensive consideration of reliability and energy efficiency. For example, in order to avoid the increase of delay caused by the disconnection of the previous established links, M-ATTEMPT places the nodes with high data rate on the less mobile positions of human body. However, these nodes also tend to be relatively close to the sink, which will significantly increase the probability that they are selected as relay nodes and eventually lead to energy imbalance. Similarly, the energy consumption of TTRP under different traffic conditions has not been significantly optimized.

(2) Cluster-based routing protocols. Anybody [6] is one of the typical representatives. This protocol randomly selects cluster heads periodically, and organizes multiple cluster heads into one backbone network. The cluster heads in the backbone network are responsible for communication with the base station. In this way, the direct data transmission to the remote base station can be reduced, and the high energy consumption that may be formed on the cluster heads is evenly dispersed within the backbone network. However, this protocol only focuses on the number of cluster heads, but lacks a comprehensive consideration of energy efficiency and reliability. HIT [7] is based on a hybrid structure of one or more clusters. In this protocol, direct transmission to the sink node is minimized, and parallel processing is adopted in the intra-cluster and inter-cluster communication processes. HIT can reduce energy consumption and improve the network lifetime in a loose WBAN. But in a compact WBAN, collisions between communication paths in this protocol will result in higher energy consumption. In addition, this protocol does not consider the reliability of data transmission. Ant Colony Optimization for Body Area Network (ACOBAN) [21] is a routing protocol based on the cluster structure and the ant colony algorithm. It divides the sensor nodes in a WBAN into multiple clusters. The member nodes in the cluster transmit data to the cluster head, and the cluster head finally forwards the data to the cluster head of other clusters. For transmission between cluster heads, the ant colony algorithm is used to select the relay node, which can make nodes consume energy more evenly. However, this way that sending forward ants and backward ants will bring additional energy overhead. Nano Cluster Composition Algorithm (NCCA) [22] is a protocol with a multi-layer cluster structure. Firstly, the sensor nodes are divided into multiple layers according to the distance between the nodes and the sink node. Then, at each layer, the sensor nodes are divided into multiple clusters. For a WBAN with a small number of nodes, it is necessary to consider whether it is worthwhile to divide layers, build and maintain clusters. In addition, the protocol proposed in [23] uses multi-objective genetic algorithm to dynamically select cluster heads and routing path. But the complexity of the algorithm is high, and the cost of frequent repeated calculation caused by human movement needs to be considered. ERRS [8] is also a cluster-based routing protocol proposed for WBANs in recent years. ERRS takes advantage of these two solutions of Forwarder Node Selection and Forwarder Node Rotation to enhance the stability period and the reliability of WBAN. However, it may cause frequent changes of forwarding nodes and accordingly increase the energy consumption.

(3) Cross-layer-based routing protocols. TICOSS [9] is a cross-layer-based multi-hop protocol building on IEEE 802.15.4. This protocol aims to enhance IEEE 802.15.4 by making sensor nodes switch activity status to preserve energy, reducing the packet collisions due to hidden terminals, and searching for the shortest routing path to the closest sink node. It has been proved that the network lifetime achieved by TICOSS is almost twice as long as that by IEEE 802.15.4. But TICOSS can not effectively guarantee the reliability of data transmission. WASP [10] is a slotted multi-hop approach to medium access control and routing. The WASP protocol first establishes a spanning tree, splits the time axis into time slots, and allocates time slots to each node in a distributed manner. Then, each node sends a proprietary message to its child nodes in its own time slot. This message is used to inform the child nodes of the time slots in which their data can be sent. This protocol can achieve high throughput, but it has low energy efficiency and does not support two-way communication. PCLRP [11] is an adaptive cross-layer-based routing protocol. In this protocol, slot assignment techniques, sleep and wake-up mechanism are used to deal with topology changes, and priority guaranteed CSMA/CA approaches are used to access channel. The PCLRP protocol is used in conjunction with the PCLMAC protocol to ensure the reliability of data transmission. However, like many cross-layer-based routing protocols, PCLRP does not

consider the problem of energy imbalance, which is prone to generate energy black hole and lead to the end of network lifetime. Cooperative Multi-path Routing (CMR) [24] is also a cross-layer-based routing protocol, which is designed to improve the network performance of WBANs by jointly processing multiple layers of the protocol stack. It considers the potential inter-network communication and cooperation between WBANs, and mitigates the radio interference between WBANs. However, this protocol is based on the Open Shortest Path First (OSPF) protocol and uses the Dijkstra algorithm, which means that each node needs to maintain link state information and build its own minimum spanning tree. When the network topology changes frequently, there may be a great energy overhead.

(4) Cost-function-based routing protocols. iM-SIMPLE [12] is one of these cost-function-based routing protocols. In iM-SIMPLE, a cost function with residual energy and distance to sink as parameters is designed. According to the cost function, the sink node calculates the cost value for each node and selects the node with the minimum value as the forwarder. This means that this protocol mainly considers the problems of energy balance and energy consumption, but can not guarantee the reliability of data transmission. OCER [13] is another cost-function-based routing protocol. In this protocol, when a node needs to transmit data, a multi-objective cost function determined by residual energy, path loss and link reliability is used to select the forwarding node. The weights of the parameters in this cost function are adjusted by genetic algorithm. OCER has a good performance in various comparative experiments. However, it selects the forwarding node according to the local link cost rather than the overall path cost,which may fall into local optimization. The protocol of ETPA [14] takes a node's temperature, energy level and received power from neighbor nodes into account to define the cost function, and selects the neighbor node with the minimum cost as the next hop to transmit data. But it does not take effective optimization measures to ensure transmission reliability.The comparative experiments with PCLRP [11] show that ETPA performs well in a loose WBAN, but in a compact WBAN, the average packet delivery rate decreases from 78.72% to 68.64%. Co-LAEEBA [15] and EERS [16] are also routing protocols of this type. These protocols design cost function according to performance requirements, and always use multi-hop scheme to transmit data. However, unnecessary multi-hop transmission will bring more energy consumption. In addition, an energy efficient protocol for routing and scheduling in WBANs is proposed in [25]. In this protocol, each sensor node calculates the cost of each related link, and selects a routing path with the minimum link cost. The link cost is determined by these factors such as node energy, path loss, traffic type, and so on. In addition to the routing scheme, the authors also propose a channel competition procedure and an adaptive slot assignment method. The experiment results show that this protocol performs well in comprehensive energy consumption, but there is still a problem of energy imbalance. In this protocol, the second node failed at 7300 rounds, but the first node failed at 3500 rounds. This is because this protocol uses the greedy algorithm to select forwarding nodes. It always selects the node with the minimum related link cost as the next hop, which tends to make some nodes die prematurely.

In summary, among the above representative protocols, some protocols do not comprehensively consider the effective measures to reduce the energy consumption, balance the energy consumption between nodes and reduce the PLR. Although some protocols take optimization measures in these three aspects, they often select the forwarding node according to the local link cost, which may not necessarily lead to the optimal overall network performance. Different from the previous work, our work takes optimization measures in these three aspects, and uses the global optimization algorithm to optimize the selection of routing path. Table 1 summarizes our optimization measures, which are further explained as follows.

(1) In our proposed protocol EERR-RLFL, the FLLQE algorithm is used to establish a mathematical model to evaluate the link quality. The existing researches on link quality often use specific scores to evaluate the link quality according to the historical link conditions. If the routing path is selected based on this, it is prone to cause an uneven energy consumption between nodes, and it will bring about a high PLR in the case of frequent changes in the network topology. While the FLLQE algorithm uses fuzzy logic to evaluate the probability that the link quality is in each state, which can provide reference for probabilistic routing path selection. For the link quality evaluation, node ranks, residual energy and link distance are taken into account as impact factors, which can protect important nodes as much as possible, balance the energy consumption between nodes and reduce the total energy consumption. In addition, the hop count from each candidate relay node to the sink and the channel state between the current node and the candidate relay node are also

Table 1 Optimization measures in our proposed protocol EERR-RLFL

Objective	Optimization measures
To decrease the energy consumption	1. For the link quality evaluation, link distance is taken into account
	2. Using a combination of single-hop and multi-hop transmission to reduce the energy consumption caused by unnecessary data forwarding
	3. Carrying a feedback in the data packet
To balance the energy consumption between nodes	1. For the link quality evaluation, node ranks are taken into account
	2. For the link quality evaluation, residual energy is taken into account
	3. Using the FLLQE algorithm to evaluate the probability that the link quality is in each state
To reduce the PLR	1. For the link quality evaluation, hop count is considered
	2. For the link quality evaluation, channel state is considered
	3. In the selection of routing path, both the historical and the current network conditions are considered
To achieve a global optimization of the network	1. Using the RL algorithm to select the forwarding node according to the global quality of the whole path rather than the local link quality
	2. Through continuous learning, the selection of routing path tends to global optimization

considered when evaluating the link quality, which can decrease the PLR as much as possible.

(2) Based on the link quality evaluation model, a hybrid data transmission mode is used. When a node needs to transmit data, it first judges whether a forwarding node is needed according to the link quality between itself and the sink. In this way, the energy consumption caused by unnecessary forwarding can be further reduced.When a forwarding node is needed, the forwarding node is selected not according to the local link quality between the current node and each candidate relay node, but according to the global quality of the whole path that used by the current node to reach the sink through each candidate relay node. The path quality is determined by two parts, one is the quality of the link between the current node and the candidate relay node, and the other is the quality of the optimal path from the candidate relay node to the sink that can be found currently. Such a way not only considers the historical situation, but also considers the current network situation, which can further reduce the PLR.

(3) In EERR-RLFL, after the current node selects the forwarding node, it will carry a feedback in the data packet. The way of carrying the feedback in the data packet can save the energy consumption of specially sending control packet. This feedback represents the quality of the selected path from the current node to the sink. The neighbor nodes receive the feedback and update the quality of the local stored correlation path. Through such a continuous learning process, the selection of routing path is getting closer to the global optimization, that is, the global energy efficient and reliable data transmission path.

3 Proposed protocol: EERR-RLFL

In this section, we first introduce the relevant technical background including RL algorithm and fuzzy logic. Then, we give the system model used by the proposed protocol EERR-RLFL. Finally, we present the protocol operation of EERR-RLFL in detail, including the node rank division mechanism, the FLLQE algorithm and the specific working process of the protocol. Before introducing these contents, we show the general overview of EERR-RLFL, as shown in Fig. 1. The work of the protocol contains two phases: network initialization and data transmission. For the detailed operation, we will present it in Sect. 3.3.

3.1 Technical background

3.1.1 RL algorithm

The goal of the RL algorithm is to learn an optimal action strategy to obtain the maximum long-term benefit. In the RL algorithm, each agent can take different actions. However, the agent does not know the advantages and disadvantages of each action in advance. So it tries different actions, and then gets feedback from the environment. According to the environment feedback, the agent adjusts the action strategy to make a better action choice in the future [26]. The related concepts [18] are defined as follows.

Definition 1 (*Agent*) It refers to a device that can sense the state of the environment and execute an action according to the action strategy. In WBANs, each sensor node is treated as an agent.

Definition 2 (*Action*) It refers to the behavior of the agent in the environment. In WBANs, the behavior that a node

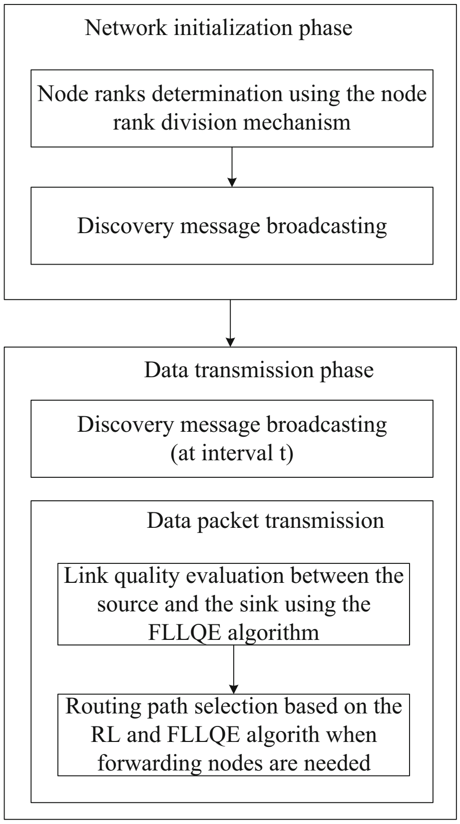


Fig. 1 Overview of EERR-RLFL

selects a neighbor node as the next hop to forward packet is regarded as an action.

Definition 3 (***Markov Decision Process(MDP)***) A RL task can be described as a MDP(S; A; P; R), where S is the set of all possible states, A is the set of all possible actions, P is the probability of state transition, and R is the reward obtained from the environment. When an agent takes an action a_i at state s_i, it will receive a reward r_i from the environment. A MDP is a set of this series of action a_i, state s_i and reward r_i.

Definition 4 (***Action policy***) It defines how an agent takes the next action at a certain state.

Definition 5 (***Reward function***) It defines the feedback value from the environment after an agent takes an action at a certain state.

Figure 2 shows the framework of the RL algorithm. It mainly includes two parts: agent and environment. The agent senses the current state of the environment, and chooses an action according to the current action policy. After the agent executes an action, the environment state of the agent is changed. The environment evaluates the impact of this action and returns a reward to the agent.

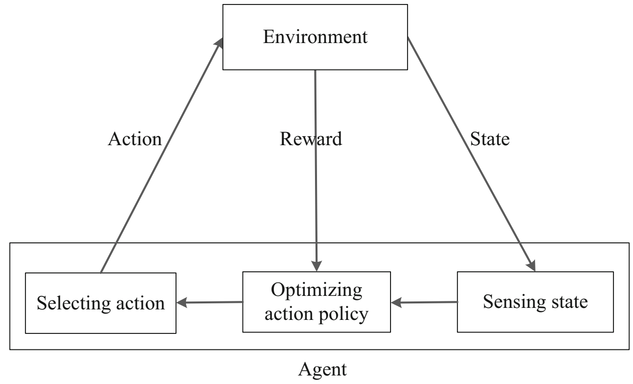


Fig. 2 Framework of the RL algorithm

According to the reward, the agent optimizes the action policy to better adapt to the environment.

3.1.2 Fuzzy logic

For the evaluation of some objects, it is not good to directly determine the specific score. On the contrary, there are often some fuzzy language concepts to evaluate an object, which needs to be judged by the way of fuzzy logic. The basic steps of fuzzy evaluation method [17, 27] are as follows:

(1) Establish a factor set: the factors affecting the evaluation object are determined.
(2) Establish a weight set: the importance of each factor in the factor set in the final evaluation is determined.
(3) Establish a fuzzy evaluation set: all evaluation descriptions need to be contained in this set.
(4) Establish a single-factor evaluation matrix: each element in this matrix represents the membership degree of each element in the factor set to each element in the evaluation set. In this matrix, the sum of elements in each row is 1.
(5) Establish a comprehensive evaluation matrix: based on the single-factor evaluation matrix, the importance of each factor is also considered to reflect the comprehensive influence of each factor.

3.2 System model

In EERR-RLFL, two types of nodes are involved, one is sensor node, the other is sink node. The sensor nodes, which are deployed in the ankle, leg, wrist, arm and other parts of of the human body, collect the blood pressure, blood glucose, pulse and other physiological data, and then transmit the data to the sink node in the way of single hop or multi hop. The sink node finally sends the data to hospital monitoring center, relatives' mobile phones or community service center.

3.2.1 Path loss model

Due to the characteristics of wireless channel in WBANs, there will be signal attenuation between the receiving node and the sending node. Therefore, it is important to estimate the path loss between two nodes. The path loss in WBANs is not only related to distance, but also to frequency. The influence of human shadow effect should also be considered. In the proposed protocol, the calculation of the path loss (dB) between the sending node and the receiving node follows the accepted formula [1]:

$$PL(d) = PL(d_0) + 10\eta \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma \tag{1}$$

where PL(d) is the path loss at distance d, PL(d_0) is the known path loss at a reference distance d_0, which can be computed by Eq. (2). In addition, η is the path loss exponent, and Xσ is a gaussian zero-mean random variable with standard deviation σ.

$$PL(d_0) = 20\log_{10}\left(\frac{4\pi d_0 f}{c}\right) \tag{2}$$

where f is the frequency of the propagating wave, and c is the speed of light.

In the experimental platform Castalia used by our protocol, the following default values are used: η = 2.4, PL(d_0) = 55 dB, d_0 = 1 m. Xσ can be obtained from the measured documents provided by Castalia.

3.2.2 Energy model

In the proposed protocol, the energy model in [28] is used to compute the energy consumption. As shown in Eq. (3), E_{Tx}(k,d) represents the energy consumption to transmit k bits data to a node at a distance of d, and E_{Rx}(k) denotes the energy consumption to receive k bits data.

$$\begin{cases} E_{Tx}(k,d) = E_{elec}k + \varepsilon_{amp}kd^m \\ E_{Rx}(k) = E_{elec}k \end{cases} \tag{3}$$

where k is the length of a packet, d is the transmission distance, E_{elec} represents the energy consumption for the transmitter or receiver circuitry to transmit or receive unit data, and ε_{amp} symbolizes the energy consumption for transmitter amplifier. These two factors are both constant: E_{elec} = 50nJ/bit, ε_{amp} = 100pJ/bit/m^2. In addition, m is an exponent of propagation attenuation, and its value is 2 or 4.When transmission distance is greater than a certain threshold, the value of m is 4. In our protocol, the distance of direct transmission is short, so the value of m is 2.

3.3 Protocol operation

In EERR-RLFL, sensor nodes are divided into different ranks from three aspects by the node rank division mechanism. Each rank is considered to be one of the factors that affect the link quality. Then, the link quality between two nodes is evaluated by the FLLQE algorithm, which provides reference for routing path selection. Based on the link quality evaluation, in the data transmission phase, when a forwarding node is needed, the RL algorithm is used to select the optimized routing path.

3.3.1 Node rank division mechanism

In WBANs, sensor nodes are heterogeneous. First, the deployment location of each node is different, and the difficulty of replenishing energy is different. It is more difficult for in-body nodes to supplement energy. Second, the data services collected by each node are different, including electrocardiogram (ECG), electroencephalogram (EEG), blood pressure, blood oxygen, and so on. These sensor nodes have different importance. Third, the workload of each node is different, some nodes need continuous monitoring data, while others only need to collect data in a short time. For example, the ECG node needs one day or several days to continuously monitor the ECG state of human body, while blood oxygen node can complete the routine detection in a short time. In view of the heterogeneity of nodes, we divide the nodes into different ranks from three perspectives, as shown in Fig. 3.

(1) According to the deployment location of sensor nodes, they are divided into two ranks. The rank of on-body node is defined as high-rank, while that of in-body node is defined as low-rank. When data needs to be forwarded, a high-rank node will be preferentially selected as a relay node. Since it is very difficult for in-body nodes to supplement energy, this division mechanism is to avoid using in-body nodes as relay nodes.

(2) According to the importance of sensor nodes, they are divided into two ranks. If a node collects important physiological information related to life and no other node collects the same information, the node is regarded as an important node. The rank of important nodes is defined as low-rank to make such nodes undertake less forwarding tasks. On the contrary, unimportant nodes are defined as high-rank nodes.

(3) According to the workload of sensor nodes, they are divided into two ranks. If a node can complete the collection task in a short time, the node is defined as a high-rank node, so that this type of nodes is more

```
NodeRankDivisionMechanism
// Global variables: n, Node[];
//n: the number of sensor nodes in the network;
//Node[]: an array recording the information of sensor nodes, in which the type of each element is struct;
1. for ( i = 1 to n)
2.     if (node i is an on-body node) Node[i].RankL = 1;
3.     else Node[i].RankL = 0;
// The member RankL of Node[i] records the node rank in terms of deployment location
4.     if (node i is an unimportant node) Node[i].RankI = 1;
5.     else Node[i].RankI = 0;
//The member RankI of Node[i] records the node rank in terms of node importance
6.     if (node i can complete the collection task in a short time) Node[i].RankW = 1;
7.     else Node[i].RankW = 0;
//The member RankW of Node[i] records the node rank in terms of node workload
8. End for
```

Fig. 3 Node rank division mechanism

responsible for forwarding tasks. Otherwise, the node is defined as a low-rank node.

3.3.2 FLLQE algorithm

In WBANs, in order to ensure the reliability of data transmission, it is necessary to transmit data along a better path. However, what is a better path? First of all, it is necessary to determine the evaluation index of link quality. For link quality, it is actually a fuzzy phenomenon. For example, we evaluate the link quality between node i and node j as good. But what is good? It is a vague concept. For this problem, we propose the FLLQE algorithm, which uses fuzzy evaluation method of fuzzy logic to evaluate the link quality between two nodes in WBANs. Take the link between the current node i and any of its neighbors j as an example, the FLLQE algorithm works as follows:

(1) Create a factor set U, in which each element affects the link quality. In this paper, when relay nodes are needed to forward data, link quality determines the selection of the next-hop node, which directly affects the reliability and energy efficiency of data transmission. For WBANs, the factors related to these performances includes the residual energy and the ranks of the relay node, the link distance and the channel state between the current node and the relay node, and the hop count from the relay node to the sink. Thus, the factor set affecting the link quality is defined as

$$U = \{R\_e, R\_d, R\_i, R\_w, H\_c, L\_d, C\_s\} \tag{4}$$

where R_e, H_c, L_d and C_s denote these influencing factors such as residual energy, hop count, link distance and channel state. R_d, R_i, and R_w respectively represent the rank factors in terms of deployment location, importance and workload.

(2) Establish a weight set W according to the influence degree of each factor in U on the link quality, as shown in Eq. (5).

$$W = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7\} \tag{5}$$

where $w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 = 1$. Each weight in W reflects the importance of each factor in U, which can be set according to the specific application scenarios.

(3) Build a fuzzy set V to evaluate link quality. In WBANs, the link quality between two nodes can be simply divided into such two states as good and bad. Thus, the evaluation set describing link quality of WBANs is defined as:

$$V = \{good, bad\} \tag{6}$$

(4) On basis of U and V, establish a single-factor evaluation matrix M_{ij} for the link between node i and node j. For each element in the factor set U, the membership degree corresponding to each element in the evaluation set V is calculated and recorded in a matrix M. The matrix M corresponding to the link between node i and node j is defined as shown in

Eq. (7), and each element in M_{ij} is computed as shown in Table 2.

$$M_{ij} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \\ m_{41} & m_{42} \\ m_{51} & m_{52} \\ m_{61} & m_{62} \\ m_{71} & m_{72} \end{bmatrix} \quad (7)$$

where $m_{i1} + m_{i2} = 1$ ($i = 1, 2 \ldots 7$).

As shown in Table 2, when evaluating the link quality between the current node i and any of its neighbors j, the matrix M_{ij} reflects the membership degree of each factor affecting link quality to each element in evaluation set.

- In terms of the impact of the residual energy on the link quality, the higher the residual energy of node j, the greater the probability that the link quality between node i and node j is good. This is because for the current node i, when it needs a relay node to forward data, the higher the energy of the relay node, the greater the probability of its data being successfully forwarded. In this respect, the best state of link quality is that the remaining energy of node j is equal to the initial energy. Therefore, the probability that the link quality between node i and node j is good is defined as RE_j / IE_j. RE_j and IE_j represent the values of residual energy and initial energy of node j.
- Regarding the impact of the node rank on the link quality based on the deployment location, the probability that the link between node i and node j is a good link is defined as the RD value of node j. As shown in Fig. 4, if node j is an in-body node, its RD value is equal to 0. Link quality is the basis for selecting relay nodes. Such a design is to avoid using in-body nodes as relay nodes. In the same way, for the impact of the node ranks on the link quality based on the node importance and the node workload, the probability that the link quality between node i and node j is good is defined as node j's RI value and RW value, which is to avoid using important nodes and nodes that need to work continuously as relay nodes.
- With regard to the impact of the hop count on the link quality, we define the probability of being good link between node i and nodes j as inversely proportional to the hop count from node j to sink. As shown in Table 2, HC_j denotes the hop count from node j to sink. This is to encourage the selection of nodes with less hops to sink as relay nodes since link quality is the basis for selecting relay nodes.
- For the impact of the link distance on the link quality, we design it by combining the energy consumption model. According to the energy consumption model in Eq. (3), the energy consumed by a sensor node to send data can be roughly calculated as proportional to the square of the sending distance. Since energy limitation is a significant feature for WBANs, the probability that the link quality between two nodes is bad can be defined as a certain relationship with the square of the distance between the two nodes. In addition, node j is a neighbor node of node i, so the maximum distance between node i and node j is equal to the

Table 2 Each element in M_{ij}

U	V	
	Good	Bad
R_*e*	m_{11} = RE_j / IE_j	m_{12} = 1 − RE_j / IE_j
R_d	m_{21} = RD_j	m_{22} = 1 − RD_j
R_i	m_{31} = RI_j	m_{32} = 1 − RI_j
R_w	m_{41} = RW_j	m_{42} = 1 − RW_j
H_c	m_{51} = 1/ HC_j	m_{52} = 1–1/ HC_j
L_d	m_{61} = 1-D_{ij}^2/CR^2	m_{62} = D_{ij}^2/CR^2
C_s	m_{71} = CS_{ij}	m_{72} = 1 − CS_{ij}

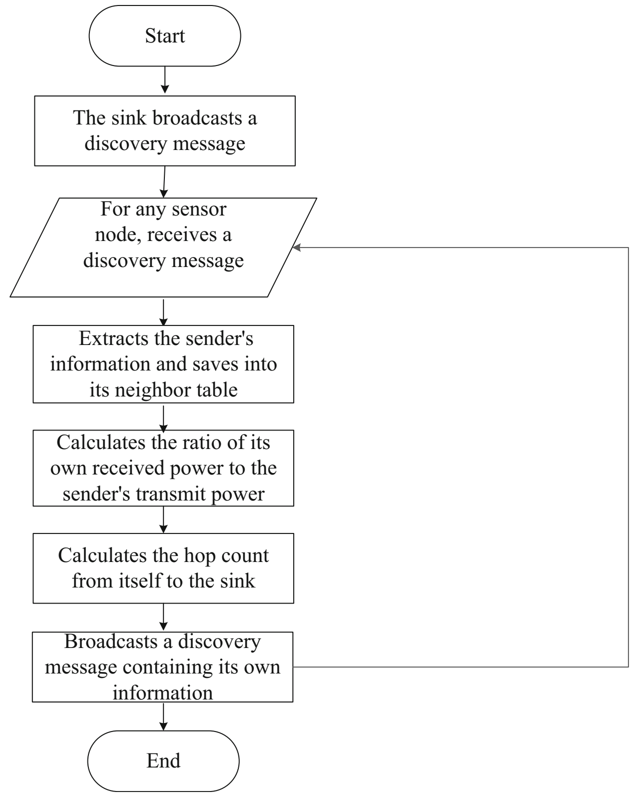


Fig. 4 Broadcasting process of the discovery message

communication radius. Thus, for the link quality between node i and node j, the probability of being bad is defined as D_{ij}^2/CR^2, and the probability of being good is defined as $1-D_{ij}^2/CR^2$. D_{ij}^2 represents the square of the distance between node i and node j, which can be computed by Eq. (8), and CR denotes the communication radius.

$$D_{ij}^2 = (X_j - X_i)^2 + (Y_j - Y_i)^2 \tag{8}$$

where (X_i, Y_i) and (X_j, Y_j) are the location coordinates of node i and node j.

- Finally, from the perspective of the impact of the channel state on the link quality, the probability that the link quality between node i and node j is good is defined as CS_{ij}. The calculation is shown in Eq. (9). The reason for such a design is that $RSSI_j$ reflects the recent channel state between node i and node j, and the larger the value, the better the channel state. The maximum value of $RSSI_j$ is equal to Tx_power_i, which is an ideal condition without considering path loss. In addition, in the simulation platform Castalia used in our experiment, dBm is taken as the unit of receiving power and transmitting power, and the receiving power and transmitting power are both negative. Consequently, Tx_power_i / $RSSI_j$ can be regarded as the probability of a good link between node i and node j.

$$CS_{ij} = Tx\_power_i / RSSI_j \tag{9}$$

where $RSSI_j$ is the Received Signal Strength Indicator (RSSI) of node j the last time it received a signal from node i, and Tx_power_i is the corresponding transmitting power of node i. The values of $RSSI_j$ and Tx_power_i are both negative. The larger the value of $RSSI_j$, the smaller the absolute value of $RSSI_j$, and accordingly the larger the value of Tx_power_i /$RSSI_j$, indicating that the channel state between node i and node j is better.

(5) Build a comprehensive evaluation matrix CM_{ij} for the link between node i and node j. On basis of the single-factor evaluation matrix M_{ij}, considering the influencing weight of each factor, a comprehensive evaluation matrix CM_{ij} is established by Eq. (10).

$$\begin{aligned} CM_{ij} &= W \circ M_{ij} \\ &= \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 \end{bmatrix} \circ \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \\ m_{41} & m_{42} \\ m_{51} & m_{52} \\ m_{61} & m_{62} \\ m_{71} & m_{72} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^{7} w_i m_{i1} & \sum_{i=1}^{7} w_i m_{i2} \end{bmatrix} \end{aligned} \tag{10}$$

where w_i is the element of the weight set W, as shown in Eq. (5). m_{i1} and m_{i2} are the elements of matrix M, which can be obtained by calculation in Table 2.

The matrix CM reflects the comprehensive influence of all factors, which represents the membership degree of the current link quality to each element in the evaluation set. As shown in Eq. (10), for the link between node i and node j, the first element of the matrix CM is the probability of good link quality, while the second element is the probability of bad link quality.

3.3.3 Working process

In EERR-RLFL, a hybrid data transmission method is used. First, whether relay nodes are needed is determined. If relay nodes are needed, the reinforcement learning algorithm is used to find the optimized routing path. Reinforcement learning deals with how agents take actions to obtain the long-term maximum benefit. This is in line with our goal: how nodes choose routing paths to maximize the reliability and energy efficiency of data transmission. The reinforcement learning algorithm has strong robustness, even in the case of topology changes, it can still guarantee reliable data transmission and requires less additional overhead.Therefore, in EERR-RLFL, the advantages of reinforcement learning are used to continuously optimize routing path with less additional overhead. The working process of EERR-RLFL includes network initialization phase and data transmission phase. In the network initialization phase, a discovery message is broadcast starting from the sink. In the data transmission phase, in addition to transmitting data packets, the discovery message is still broadcast regularly. The broadcasting process of the discovery message is shown in Fig. 4, and the flowchart of EERR-RLFL in data transmission phase is shown in Fig. 5.

(1) In the network initialization phase, each node discovers the information of its neighbor nodes. As

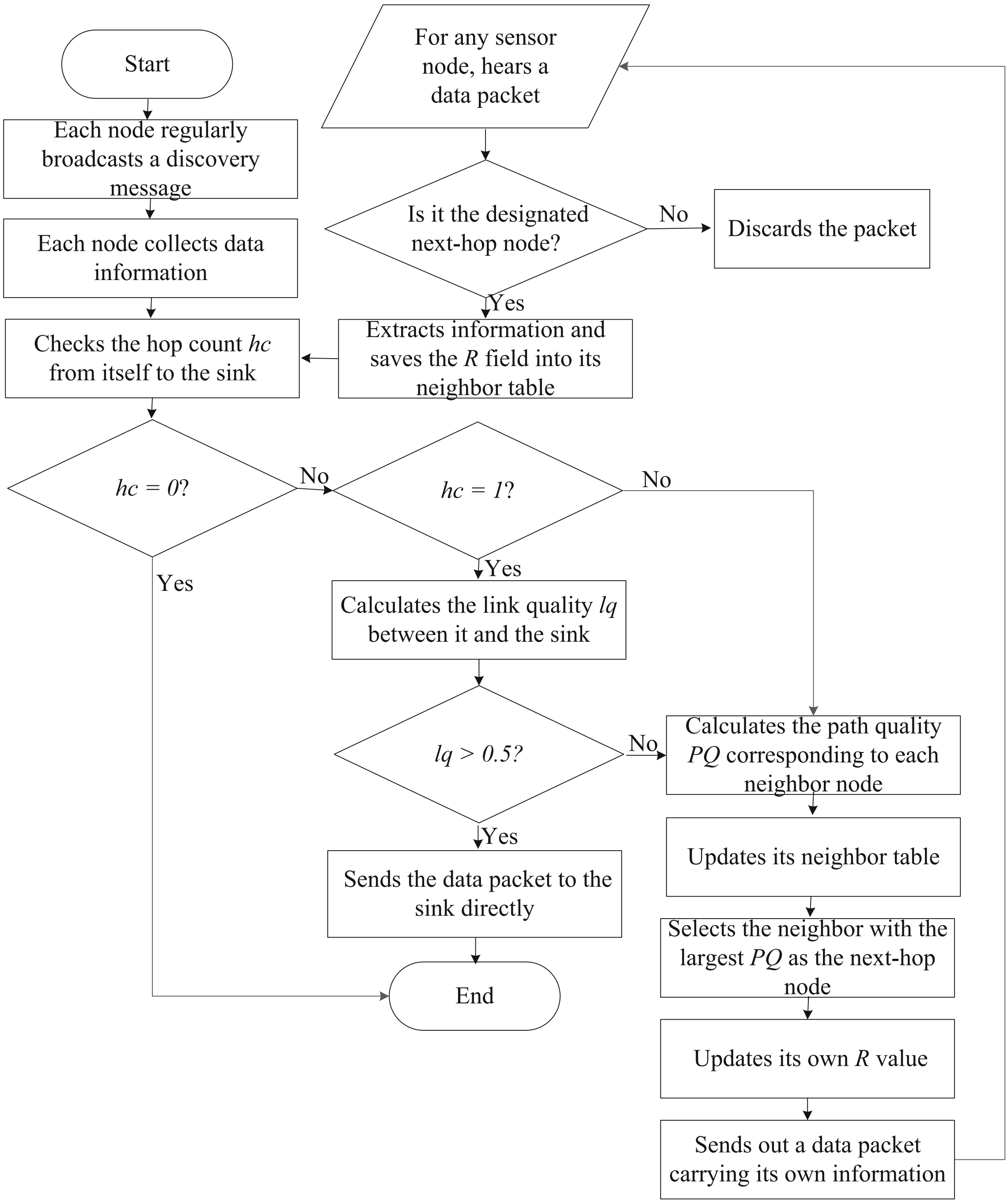


Fig. 5 Flowchart of EERR-RLFL in data transmission phase

shown in Figs. 6 and 7, the network initialization process is carried out through the following steps.

- A discovery message is locally broadcast from the sink node. The discovery message is a small control packet, and its structure is shown in Fig. 8. The first two fields indicate the type and identification of the packet, and the remaining fields indicate the sender's information. NodeID, (X,Y) and ResidualEnergy represent the identification, the location coordinates and the residual energy of the sender. RankL, RankI and RankW are the ranks of this sender in terms of deployment location, importance and workload. HopCount denotes the hop count from this sender to the sink, and Tx_power indicates the transmitting power when this sender sent this message.

```
NetworkInitialization
// Global variables: n, Node[], phase;
//n: the number of sensor nodes in the network;
//Node[]: an array recording the information of sensor nodes, in which the type of each element is struct;
// phase: a marker of working phase;
1. phase = 1;
2. Call NodeRankDivisionMechanism;
3. for ( i = 1 to n) Node[i].NeighborNum=0;
4. Call BroadcastingControlPacket;
```

Fig. 6 Network initialization process

- When a node receives a discovery message, it first extracts the sender's information and saves the information in its neighbor table. In addition, it calculates the ratio of its own received power to the sender's transmit power and also records it in its neighbor table. Then, on the basis of the hop count from the sender to the sink, it adds one to calculate the hop count from itself to the sink. Finally, it broadcasts a discovery message containing its own information. In this way, each node can know the hop count from itself to the sink, as well as the information of its each neighbor. For each node, each item in its neighbor table is shown in Fig. 9. The first seven items have the same meanings as those in Fig. 8. ChannelState represents the latest ratio of the power received by the current node to the corresponding transmitting power of this neighbor. Beyond that, R denotes the quality of the best path that can be found currently from this neighbor to the sink, and PQ indicates the quality of the path taken by the current node to send data to the sink through this neighbor. The R field is carried in the data packet, and is extracted by the receiving node and stored in the corresponding entry in the neighbor table. PQ is calculated by the current node in the data transmission phase and then saved in the neighbor table. The specific definitions and calculation methods of R and PQ will be introduced in detail in the following data transmission phase. In the initialization phase, the values of these two fields are set to 0.

(2) Each node collects the physiological information it is responsible for, and the network enters the data transmission phase. Figure 10 shows the data transmission process, and Fig. 11 lists the steps of the hybrid data transmission mechanism. In the data transmission phase, due to the dynamic changes of network topology in WBANs, each node still regularly broadcasts the discovery message. The process is the same as that in step (1), the neighbor table and the hop count of each node are updated regularly. While propagating the control packet, each sensor node collects data information and sends it out. As shown in Fig. 12, the header of data packet contains the type and the identification of the packet, the node identification of the sender, the quality of the best path that can be found currently from this sender to the sink (marked by R), and the designated next-hop node.

(3) For any sensor node i, when it hears a data packet, it first checks whether it is the designated next-hop node. If not, it discards the packet. Otherwise, it extracts the information in the data packet and saves the R field into the corresponding entry in its neighbor table. And proceed to step (4). When a node i generates a data packet, it also goes to step (4).

(4) The current node i checks the hop count from itself to the sink. There are several situations:

- If the hop count is equal to 0, this round of data transmission ends.
- If the hop count is equal to 1, the FLLQE algorithm is used to calculate the link quality between the current node and the sink. If the probability that this link quality is good is greater than 0.5, the data packet is sent directly to the sink, and this round of data transmission ends. Otherwise, go to step (5).
- If the hop count is greater than 1, go to step (5).

(5) Based on the reinforcement learning algorithm, the routing path from the current node i to the sink is established and the path selection is continuously optimized. Figures 13 and 14 show the routing selection algorithm and the FLLQE algorithm. In the routing selection algorithm, the related terms and symbols are defined as follows:

```
BroadcastingControlPacket
// Global variables: n, Node[];
// n: the number of sensor nodes in the network;
// Node[]: an array recording the information of sensor nodes, in which the type of each element is struct;
The sink node broadcasts a control packet;
for ( i =1 to n)
    if (node i hears a control packet cp) { // the type of cp is struct, and its structure is shown in Fig.4
        for (j=1 to Node[i].NeighborNum)
            if (Node[i].NeighborNodes[j].NodeID = = cp.NodeID) break;
        End for
// The member NeighborNum of Node[i] records the number of node i's neighbor nodes
// The member NeighborNodes[ ] of Node[i] records the information of node i's neighbor nodes
// The type of each element in NeighborNodes[ ] is struct, and its structure is shown in Fig.5
        if (j > Num) {Node[i].NeighborNum++; Node[i].NeighborNodes[j].NodeID=cp.NodeID;}
        Node[i].NeighborNodes[j].X=cp.X;
        Node[i].NeighborNodes[j].Y=cp.Y;
        Node[i].NeighborNodes[j].ResidualEnergy=cp.ResidualEnergy;
        Node[i].NeighborNodes[j].RankL=cp.RankL;
        Node[i].NeighborNodes[j].RankI=cp.RankL;
        Node[i].NeighborNodes[j].RankW=cp.RankW;
        Node[i].NeighborNodes[j].HopCount=cp.HopCount;
        Node[i].NeighborNodes[j].ChannelState=Node[i].RSSI/cp.Tx_power;
//The member RSSI of Node[i] records the current received power of node i
        if (phase = = 1){ Node[i].NeighborNodes[j].R=0; Node[i].NeighborNodes[j].PQ=0;}
        Node[i].HopCount=cp.HopCount+1;
//The member HopCount of Node[i] records the hop count from node i to the sink
        Node i updates the control packet cp with its own information, and then broadcasts it.
    }
End for
```

Fig. 7 Control packet broadcasting process

Type	PacketID	NodeID	(X,Y)	ResidualEnergy	RankL	RankI	RankW	HopCount	Tx_power

Fig. 8 Structure of control packet

NodeID	(X,Y)	ResidualEnergy	RankL	RankI	RankW	HopCount	ChannelState	R	PQ

Fig. 9 Each item in the neighbor table

```
DataTransmission
// Global variables: n, Node[];
//n: the number of sensor nodes in the network;
//Node[]: an array recording the information of sensor nodes, in which the type of each element is struct;
1. phase = 2;
2. Call BroadcastingControlPacket at interval t;//Control packet and data packet use different channels
3. for (i = 1 to n)
4.    if (node i generates a data packet dp) //the type of dp is struct, and its structure is shown in Fig.8
5.       Call HybridDataTransmissionMechanism( i,dp );
6.    if (node i hears a data packet dp) {
7.        if (i == dp. NextID) {
8.            for ( j = 1 to Node[i].NeighborNum)
//The member NeighborNum of Node[i] records the number of node i's neighbor nodes
9.                if ( Node[i].NeighborNodes[j].NodeID == dp.NodeID)
10.                   Node[i].NeighborNodes[j].R = dp.R
//The member NeighborNodes[ ] of Node[i] records the information of node i's neighbor nodes
// The type of each element in NeighborNodes[ ] is struct, and its structure is shown in Fig.5
11.           End for
12.           Call HybridDataTransmissionMechanism( i,dp );
13.       }
14.       else Discard the data packet dp;
15.   }
16. End for
```

Fig. 10 Data transmission process

```
HybridDataTransmissionMechanism(i, dp)
// i:the ID of the current node; dp: the current data packet
// Global variables: Node[];
//Node[]: an array recording the information of sensor nodes, in which the type of each element is struct.
1. if (Node[i].HopCount ==0) Finish this round of data transmission;
//The member HopCount of Node[i] records the hop count from node i to the sink
2. else if (Node[i].HopCount ==1) {
3.         Call FLLQEAlgorithm(i,sinkID);
4.         if (Node[i].R > 0.5) Send the data packet dp directly to the sink;
5.         else Call RoutingSelectionAlgorithm(i,dp);
6.   }
7.   else Call RoutingSelectionAlgorithm(i,dp);
```

Fig. 11 Hybrid data transmission mechanism

Type	PacketID	NodeID	R	NextID	PayloadData

Fig. 12 Structure of data packet

```
RoutingSelectionAlgorithm(i,dp)
// i: the ID of the current node; dp: the current data packet
// Global variables: Node[]; an array recording the information of sensor nodes, in which the type of each element is struct.
1. for ( j = 1 to Node[i].NeighborNum)
// The member NeighborNum of Node[i] records the number of node i's neighbor nodes
2.     Call FLLQEAlgorithm(i, Node[i].NeighborNodes[j].NodeID);
// The member NeighborNodes[ ] of Node[i] records the information of node i's neighbor nodes
// The type of each element in NeighborNodes[ ] is struct, and its structure is shown in Fig.5
3.     Compute PQ(i, Node[i].NeighborNodes[j].NodeID) according to Eq.11 and Eq.12;
4.     Node[i].NeighborNodes[j].PQ = PQ(i, Node[i].NeighborNodes[j].NodeID);
// Save the PQ value in the corresponding entry in node i's neighbor table
5.     if (Node[i].NeighborNodes[j].PQ > MaxPQ) {
6.         MaxPQ = Node[i].NeighborNodes[j].PQ;
7.         Next = Node[i].NeighborNodes[j].NodeID;
8.     }
9.  End for
10. Node[i].R = MaxPQ;
11. Node[i].HopCount = Node[Next].HopCount+1;
12. dp.NodeID=i;
13. dp.R=Node[i].R;
14. dp.NextID=Next;
// node i updates the data packet dp with its own information and the selected next- hop node
15. Send out the updated data packet dp;
16. Return to DataTransmission;
```

Fig. 13 Routing selection algorithm

Definition 1 (***Path quality PQ(i, j)***) The quality of the path taken by node i to send data to the sink through node j. More specifically, this path means that node i directly sends data to node j, and node j sends the data to the sink through one or more hops.

Definition 2 (***Link quality LQ(i, j)***) The quality of the direct link between node i and node j.

Definition 3 (***Feedback value R_j***) The quality of the best path that can be found currently from node j to the sink.

According to Eq. (11), the current node i calculates the path quality PQ_{new}(i, j) corresponding to each neighbor node j. The evaluation of path quality includes two parts: historical path information and current link information. The learning rate determines the weight of the two parts.

$$PQ_{new}(i,j) = (1-\eta)PQ_{old}(i,j) + \eta(LQ(i,j) + R_j) \quad (11)$$

where $\boldsymbol{\eta}$ is the learning rate, which can be determined through theoretical analysis and experiments according to practical applications. If the human posture changes more frequently, the value of$\boldsymbol{\eta}$is larger, indicating that the current information has a higher weight. Conversely, if the monitoring object is a bedridden patient, the historical information has a higher weight. The historical PQ_{old}(i,j) and R_j can be obtained from the entry corresponding to node j in the neighbor table of node i, and their initial

```
FLLQEAlgorithm(i, j)
// i: the ID of the current node; j: the ID of the neighbor node of node i
// Global variables: Node[];
//Node[]: an array recording the information of sensor nodes, in which the type of each element is struct.
1. for ( k = 1 to Node[i].NeighborNum)
// The member NeighborNum of Node[i] records the number of node i's neighbor nodes
2.      if ( Node[i].NeighborNodes[k].NodeID = = j) {
// The member NeighborNodes[ ] of Node[i] records the information of node i's neighbor nodes
// The type of each element in NeighborNodes[ ] is struct, and its structure is shown in Fig.5
3.           entryNum = k;
4.           break;
5.      }
6. End for
7. m[1][1] = Node[i].NeighborNodes[entryNum].ResidualEnergy/Node[i].NeighborNodes[entryNum].InitialEnergy;
8. m[2][1] = Node[i].NeighborNodes[entryNum].RankL;
9. m[3][1 ]= Node[i].NeighborNodes[entryNum].RankI;
10. m[4][1] = Node[i].NeighborNodes[entryNum].RankW;
11. m[5][1] = Node[i].NeighborNodes[entryNum].HopCount;
12. Dis2=power(Node[i].NeighborNodes[entryNum].X-Node[i].X, 2) + power(Node[i].NeighborNodes[entryNum].Y-Node[i].Y, 2);
13. m[6][1] =1 - Dis2 / power(radius,2);
14. m[7][1] =Node[i].NeighborNodes[entryNum].ChannelState;
15. for ( k = 1 to 7) cin>>w[k];
16. End for
17. LQ = 0;
18. for ( k = 1 to 7) LQ += w[k]*m[k][1];
19. return LQ;
```

Fig. 14 FLLQE algorithm

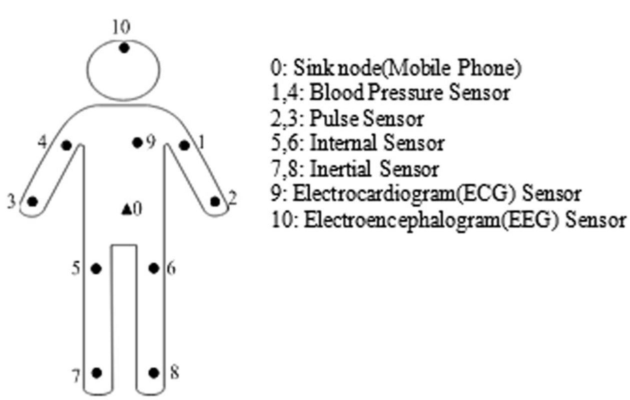


Fig. 15 Node deployment position on the human body

values are 0. In addition, the link quality LQ(i, j) can be calculated by Eq. (12).

$$LQ(i,j) = CM_{ij}(0,0) \tag{12}$$

where CM_{ij} is the comprehensive evaluation matrix of the link between node i and node j, which can be obtained by the FLLQE algorithm. CM_{ij} (0,0) represents the element in row 0 and column 0 of matrix CM_{ij}, that is, the probability of good link between node i and node j.

Although the FLLQE algorithm used here will increase the computing costs of nodes, the communication costs saved by this scheme are far greater than the additional computing costs. This is because for a sensor node, the energy costs of sensor and processor modules are far less than that of wireless communication module. The FLLQE algorithm evaluates the link quality and provides a basis for the selection of each hop path for data transmission. This algorithm enables nodes to transmit data along an optimized path, thereby saving energy costs for wireless communication.

(6) The current node i calculates the PQ value corresponding to each neighbor node and updates it to the neighbor table. Then, it selects the neighbor node with the largest PQ as the next-hop node. Most of the previous schemes select relay nodes based on the local information. Unlike these schemes, the EERR-RLFL protocol selects the routing path based on the evaluated global information. This is because that the path quality PQ(i, j) calculated by Eq. (11) indicates the overall quality of the path from node i to the sink through node j.

(7) If node i chooses node n as the next-hop node, it updates its own R value according to Eq. (13). It means that the optimal path from node i to the sink that can be found currently is to reach the sink through node n. Then, node i puts its identification, the value of Ri and the identification of node n into the corresponding fields of the data packet, as shown in Fig. 12. Here, Ri can be understood as a feedback to node i0 which is the previous-hop node of node i. This is because data packet is sent by local broadcast, the previous-hop node i0 can also hear the data packet. Node i0 extracts Ri from the data packet and updates the entry corresponding to node i in its neighbor table, which can prepare for the next path selection. Through such a learning process, the R value of each node can be continuously updated, thereby optimizing the selection of the routing path. And the energy cost of sending control packet can be saved to a certain extent by carrying feedback value in data packet.

$$R_i = PQ(i, n) \tag{13}$$

where R_i is the quality of the best path that can be found currently from node i to the sink, and PQ(i, n) is the quality of the path taken by node i to send data to the sink through node n.

The current node i sends out the updated data packet, and then the workflow of the EERR-RLFL protocol returns to step (3).

4 Performance evaluation

In this section, we evaluate the performance of EERR-RLFL in terms of network lifetime, throughput, packet loss ratio and energy efficiency. For performance comparison, we implemented these three protocols of Single Hop Transmission (SHT) [29], OCER [13] and EERR-RLFL in Castalia. Castalia, a simulator for WSNs, WBANs and generally networks of low-power embedded devices, has been used as the experimental platform by many researches on WBANs [30]. In WBANs, there are mainly two types of data transmission methods. One is that all nodes directly transmit data to the sink, and the other is that nodes use multi-hop transmission methods. Therefore, we first take the SHT protocol as a comparison object. For the multi-hop protocols, OCER is a representative energy-efficient protocol in recent years, and it is a cost-function-based protocol just as our proposed protocol. Thus, we choose these protocols as comparison objects.

In our experiment, some key simulation parameters are summarized in Tables 3 and 4. In Table 3, the elements in W represent the weights of these influence factors: residual energy, node rank in terms of deployment location, node rank in terms of importance, node rank in terms of workload, hop count to the sink, link distance and channel state. In FLLQE, the factors affecting link quality include residual energy, node ranks, hop count to the sink, link distance and channel state. In our experiment, to consider the universality, the weights of these five aspects are set to the same, i.e. 1/5. For node ranks, there are three aspects including node rank in terms of deployment location, node rank in terms of importance and node rank in terms of workload. The weight of each aspect is also set to the same, which is (1/5) * (1/3) = 1/15. In Table 4, the values of node ranks are shown. As mentioned in Sect. 3.3.1, RD, RI and RW represent the node ranks in terms of deployment location, node importance and node workload, respectively. As shown in Fig. 3 in Sect. 3.3.1, the corresponding rank value of high-rank node is 1, and the rank value of low-rank node is 0. The use of low-rank nodes for data forwarding should be avoided as far as possible.

The deployment position of the nodes on the human body is shown in Fig. 15. Taking the location of the sink node as the origin in a two-dimensional coordinate system, the location coordinates of these nodes are shown in Fig. 16.

For the simulation of path loss between nodes, on the one hand, based on the relative location of nodes, according to Eq. (1), the average path loss between nodes can be obtained, as shown in Fig. 17. The x-axis and y-axis represent the identification of the sending node and the receiving node respectively, and the color intensity indicates the average path loss between the sending node and the receiving node. The bottom color bar reveals the gradual change of color. The green color represents the minimum value of the average path loss (the average path loss between the same two nodes is 0), the red color represents the maximum value (62 dB), and the yellow color represents the intermediate value. On the other hand, the temporal variation model document, provided by the Castalia platform, can be used to simulate the path loss component caused by temporal variation. According to the two parts, we can simulate the path loss between nodes at each time. Under the same parameter settings above, the simulation was performed 10 times, and the average results are as follows.

Table 3 Simulation parameters

Parameter	Value
Topology area	2 m × 2 m
Number of nodes	1 sink node,10 sensor nodes
Initial energy	Sink node: 20 J
	Normal node: 0.5 J
Deployment location	As shown in Figs. 14 and 15
Transmitting power	− 25dBm
Packet generation rate	Different packet generation rate: (5, 10, 15, 20, 25, 30) packets/s
Data packet size	100 bytes
MAC protocol	IEEE 802.15.6
Learning rate α	0.5
Weight set W	$\{w_{R\_e}, w_{R\_d}, w_{R\_i}, w_{R\_w}, w_{H\_c}, w_{L\_d}, w_{C\_s}\}$: {1/5, 1/15, 1/15, 1/15, 1/5, 1/5, 1/5}

Table 4 Node ranks

NodeID	RD	RI	RW
1	1	1	1
2	1	0	1
3	1	0	1
4	1	1	1
5	0	1	1
6	0	1	1
7	1	1	1
8	1	1	1
9	1	1	0
10	1	1	0

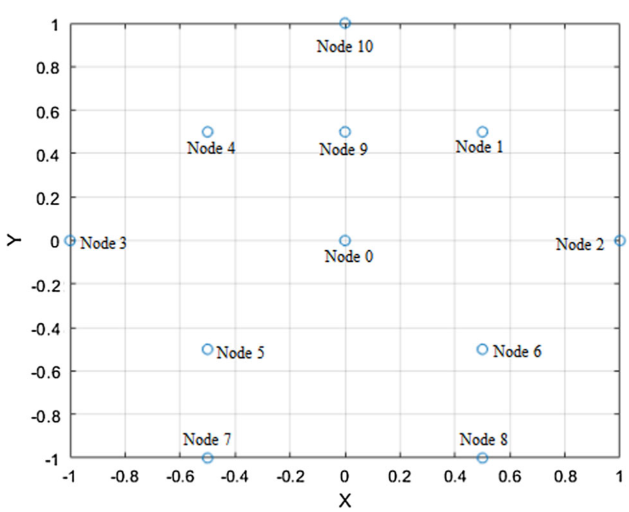


Fig. 16 Node location coordinates

4.1 Network lifetime

Under different data rates (data rate = 5,10,15,20,25,30 packets/s), we test the proportion of living nodes in the network every 10 s in the three protocols. As shown in Fig. 18, at the same data rate, the decline of the proportion of living nodes in EERR-FLRL is the latest, followed by OCER and SHT. For example, when the data rate is 5 packets/s, as shown in Fig. 18(a), for the time when the proportion of living nodes drops to 0.9, 0.7, 0.5, 0.3 and 0.1, compared with OCER, EERR-FLRL delays the time by 12%, 3%, 5%, 8% and 6% respectively. Compared with SHT, EERR-FLRL delays this time by 22%, 6%, 6%, 13% and 7% respectively. For other data rates, the comparison results of the three protocols are similar to the case when the data rate is 5 packets/s, as shown in Fig. 18(b)–(f).

Figures 19–21 show more intuitively the network lifetime comparisons of the three protocols under different data rates. We define the network lifetime from three perspectives, which are the time when the first node dies in the network, the time when 50% of the nodes in the network die, and the time when all the nodes in the network die. Figure 19 shows the network lifetime in terms of the death time of the first node. When the data rate is 5 packet/s, compared with OCER and SHT, EERR-RLFL delays the death time of the first node by 12% and 23% respectively. For the case of data rate = (10, 15, 20, 25, 30) packet/s, EERR-RLFL yields 16%, 3%, 10%, 5% and 5% longer lifetime over OCER, and 19%, 6%,17%,10% and 7% longer lifetime over SHT. It can be concluded that if the network lifetime is defined by the death time of the first node in the network, EERR-RLFL can extend the network lifetime no matter what the data rate is. In addition, Fig. 20 shows a comparison of network lifetime defined in terms of the time when 50% of the nodes in the network die. When data rate = (5, 10, 15, 20, 25, 30) packet/s, EERR-RLFL extends the network lifetime by 5%, 4%, 5%, 2%, 2% and 3% over OCER, and by 6%, 5%, 11%, 4%, 3% and 6% over SHT. After that, we also captured the time when all nodes in the three protocols died under different data rates, as shown in Fig. 21. When data rate = (5, 10, 15, 20, 25, 30) packet/s, compared with OCER, EERR-RLFL delays the time when all nodes die by 6%, 1%, 1%, 1%, 4% and

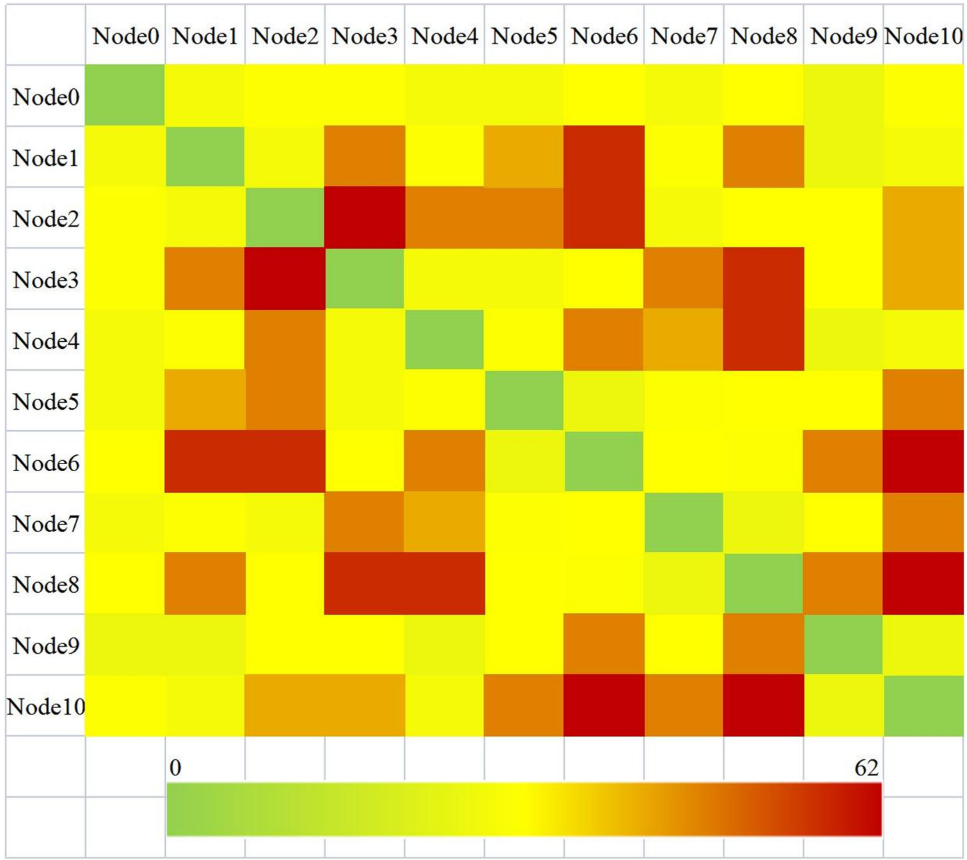


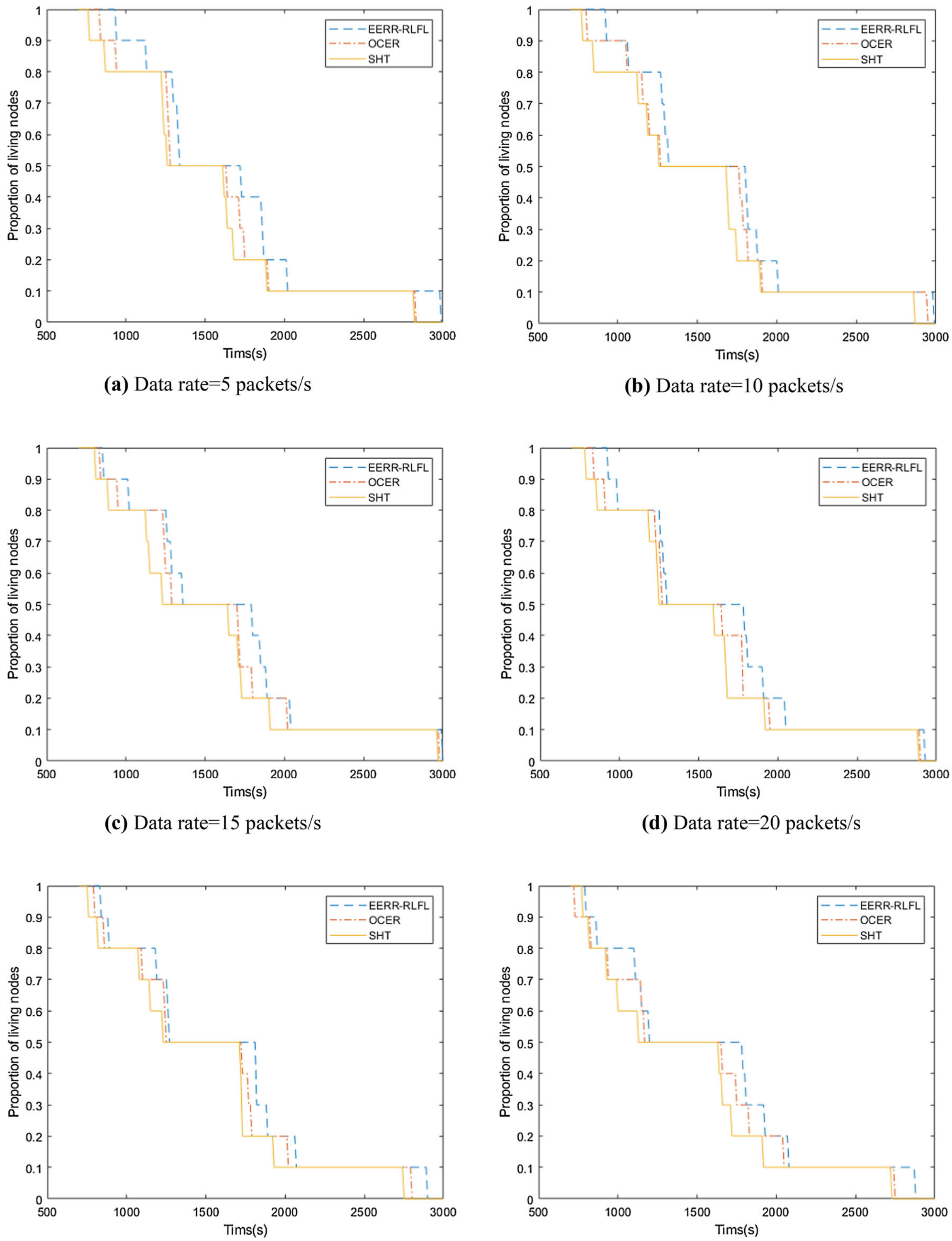
Fig. 17 Average path loss between nodes

5%. Compared with SHT, EERR-RLFL prolongs this time by 6%, 4%, 1%, 1%, 5% and 5%. As can be seen from Figs. 20 and 21, on the one hand, no matter what the data rate is, the time when half or all of the nodes die in SHT is the earliest, while this time in EERR-RLFL is the latest. On the other hand, for the same protocol, when the data rate is 5, 10 or 15 packets/s, the nodes' death time has no obvious characteristics. But when the data rate increases to 20 packets/s, with the increase of data rate, half or all nodes in the same protocol die earlier and earlier. This is due to the significant increase of the amount of data to be transmitted in the network when the data rate increases to a certain extent, which will lead to a significant increase in the energy consumption of sending and receiving data. Therefore, most of the nodes run out of energy much earlier.

From the above comparisons, we can see that EERR-RLFL can prolong the network lifetime. This is mainly because we have taken measures to reduce the total energy consumption of the network and balance the energy consumption between nodes. Firstly, we use fuzzy evaluation method to calculate the probability of good and bad link quality between nodes. In the definition of the factor set U which affects link quality, the residual energy of nodes is taken into account. This will tend to choose nodes with higher energy as forwarding nodes, which can balance the energy consumption between nodes. Secondly, taking the ranks of nodes in three perspectives as the factors in U can avoid the premature death of in-body nodes and important nodes, and balance the energy consumption by reducing the forwarding workload of continuous working nodes. Thirdly, link distance is also considered as one of the factors of U, which can promote the selection of closer nodes as forwarding nodes and thus reduce the energy consumption for transmitting data. Finally, on the basis of the link quality evaluation, the reinforcement learning algorithm is used to continuously learn and obtain the global optimal routing path, in which each node does not need global network information but can still approximate to the global optimization without additional energy.

4.2 PLR

To verify the reliability of our protocol, we test the performance of network throughput and PLR in these three protocols. In this paper, network throughput is defined as the total number of packets successfully received at the sink node when all nodes cannot complete any packet delivery. Figure 22 shows the network throughput in EERR-RLFL,OCER and SHT under different data rates. It can be seen that the network throughput in EERR-RLFL is higher than that in OCER and SHT at the same data rate.



(a) Data rate=5 packets/s

(b) Data rate=10 packets/s

(c) Data rate=15 packets/s

(d) Data rate=20 packets/s

(e) Data rate=25 packets/s

(f) Data rate=30 packets/s

Fig. 18 Proportion of living nodes over time

With the increase of data rate, EERR-RLFL has more and more obvious superiority. Specifically, when the data rates are 5, 10, 15, 20, 25 and 30 packets/s, EERR-RLFL increases the network throughput by 9%, 7%, 7%, 8%, 9% and 11% over OCER, and by 23%, 23%, 24%, 23%, 24% and 28% over SHT.

The reason why EERR-RLFL has such an advantage in network throughput is that the network throughput is

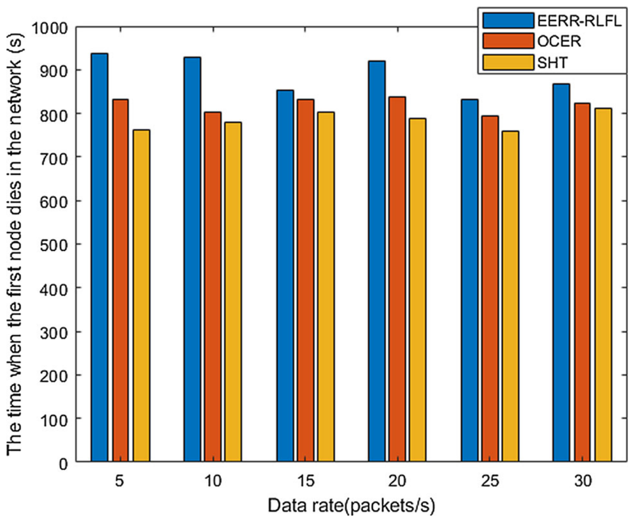


Fig. 19 The time when the first node dies in the network under different data rates

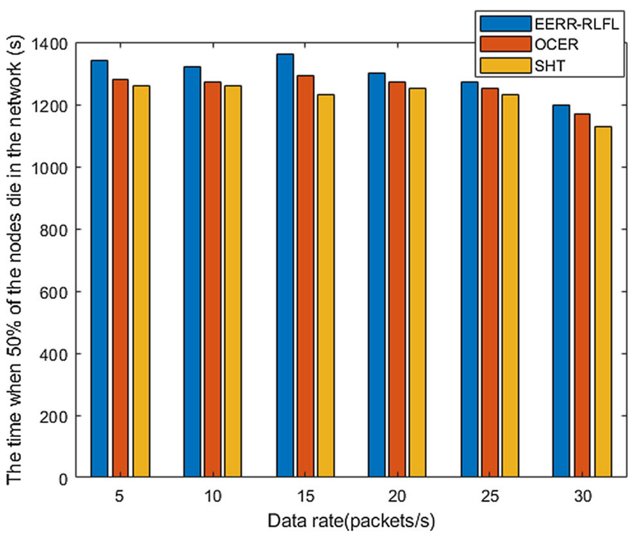


Fig. 20 The time when 50% of the nodes die in the network under different data rates

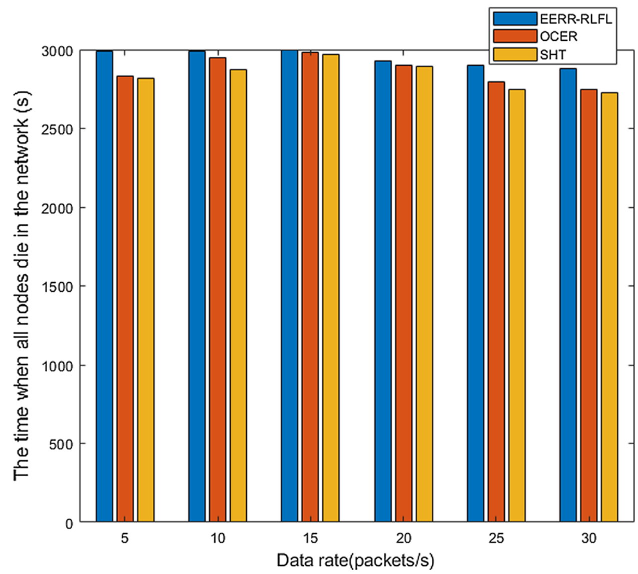


Fig. 21 The time when all nodes die in the network under different data rates

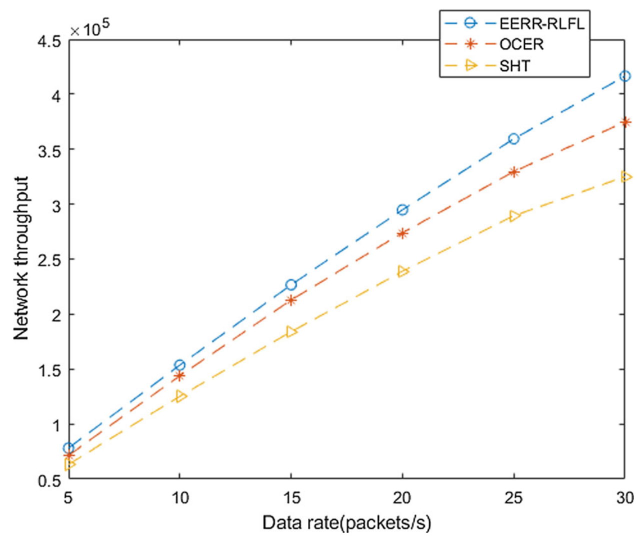


Fig. 22 Network throughput under different data rates

mainly determined by two factors. One is the number of living nodes in the network, and the other is the PLR in the network. The former determines the total number of packets generated in the network, while the latter determines the ratio of packets that can be successfully delivered to the sink node. EERR-RLFL has some optimization measures in these two aspects.

On the one hand, as described in Sect. 4.1, EERR-RLFL takes measures to reduce the total energy consumption and balance the energy consumption of nodes, so the number of living nodes in EERR-RLFL is more than that in OCER and SHT at the same time. Therefore, at a fixed data generation rate, the total number of packets generated in EERR-RLFL is larger.

On the other hand, in EERR-RLFL, the node with less hops to the sink and better channel state with the current node is prone to be chosen as the forwarding node, which can reduce the probability of packet loss. Thus, EERR-RLFL has some advantages in reducing PLR. As shown in Fig. 23, when the data rate is 5 packets/s, the PLR in EERR-RLFL, OCER and SHT is 0.051, 0.07 and 0.153 respectively. EERR-RLFL decreases PLR to 73% of OCER and 33% of SHT. At other data rates, the PLR in EERR-RLFL is still relatively low. According to our experiment, when the data rate = (10, 15, 20, 25, 30) packets/s, EERR-RLFL reduces the PLR by 27%, 26%, 24%, 32% and 31% compared with OCER, and by 64%,

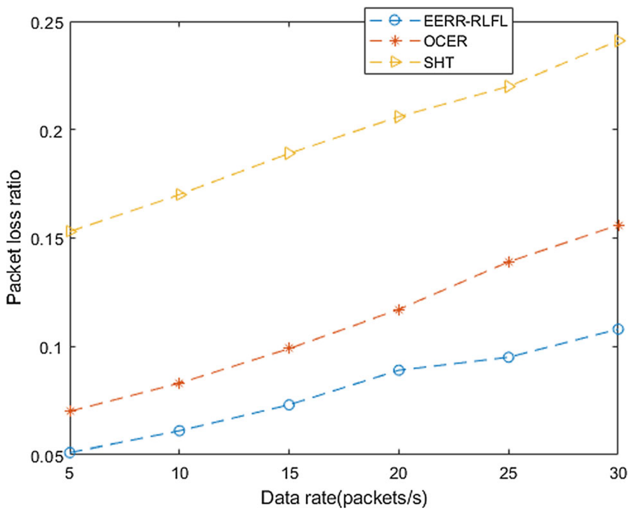


Fig. 23 Packet loss ratio under different data rates

61%, 57%, 57% and 55% compared with SHT. In addition, it can be seen that for the same curve in Fig. 23, that is, for a certain protocol, the PLR in this protocol increases with the increase of data rate. This is mainly because that the network load and the data transmission conflicts will increase when the data rate becomes larger and larger. Thus, the PLR also increases accordingly.

4.3 Energy efficiency

Finally, we test the work efficiency of these routing protocols, using energy efficiency as a measure. Here, we define energy efficiency as the average number of bits that

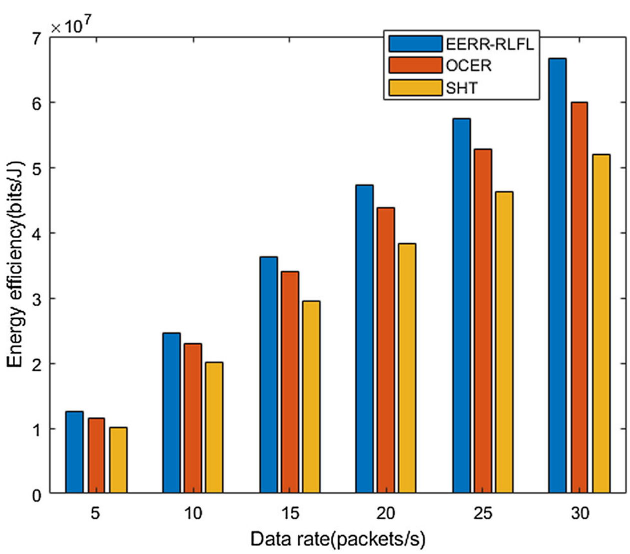


Fig. 24 Energy efficiency under different data rates

can be successfully delivered to the sink per unit of energy consumption. Figure 24 shows the energy efficiency in EERR-RLFL, OCER and SHT under different data rates. As shown in Fig. 24, no matter what the data rate is, the energy efficiency in EERR-RLFL is the highest and that in SHT is the lowest at the same data rate. For the same protocol, when the data rate increases from 5 packets/s to 30 packets/s, the energy efficiency in this protocol also increases.

In EERR-RLFL, on the premise that other conditions are the same, the node with shorter link distance is more likely to be selected as the forwarding node, which can decrease the energy consumption for transmitting data as far as possible. Moreover, the current node selects the forwarding node not only according to the local link quality, but also according to the quality of the whole path used by the current node to send data to the sink through each candidate relay node. In the process of data transmission, each node updates and carries the feedback value when sending data packets, which can save the energy consumption for sending control packet. Then, the quality of the related path is updated in the light of the feedback value. Through such a continuous learning process, the path selection is getting closer and closer to the global optimum. In addition, for the factors that affect the link quality, the residual energy of the candidate relay node, the ranks of the candidate relay node in three perspectives, the hop count from the candidate relay node to the sink and the channel state between the current node and the candidate relay node are also considered, which can balance the energy consumption and reduce the the probability of packet loss.

As explained in Sects. 4.1 and 4.2, due to the advantages in reducing and balancing energy consumption, the number of living nodes in EERR-RLFL is more than that in OCER and SHT at the same time. Accordingly, the total number of packets generated in EERR-RLFL is larger under the same data rate. Moreover, EERR-RLFL has a lower PLR, so it has the highest network throughput among these three protocols. Because of the predominance in saving energy and improving network throughput, EERR-RLFL can successfully deliver more data with less energy.

On the other hand, as described in Sect. 4.2 , for the same protocol, with the increase of data rate, the PLR increases, but the network throughput increases greatly. As the number of packets sent increases, the number of packets successfully received by the sink node is getting more. No matter what the data rate is, the total available energy in the network is basically the same. Therefore, as the data rate increases, the energy efficiency in the same protocol is enhanced.

5 Conclusions

Energy efficiency and reliability are two important performances for data transmission in WBANs. In this paper, we propose an energy efficient and reliable routing protocol based on reinforcement learning and fuzzy logic. In the proposed protocol EERR-RLFL, according to the deployment location, the importance and the workload of sensor nodes,we divide sensor nodes into different ranks from these three perspectives. Then, we use the proposed FLLQE algorithm to evaluate the link quality between two nodes. For the link quality evaluation, the factors of node ranks, residual energy, link distance, hop count and channel state are taken into account, and the fuzzy evaluation method of fuzzy logic are used to balance the comprehensive influence of these factors. When a node needs to transmit data, based on the link quality evaluation algorithm, we first judge whether a forwarding node is needed. When a forwarding node is needed, the routing path is selected based on the RL algorithm and the FLLQE algorithm, in which forwarding node is selected not according to the local link quality but according to the global path quality. Moreover, once a forwarding node is determined by the current node, a feedback will be carried in the data packet. When the neighbor nodes of the current node receive the feedback, they update the relevant path quality stored in their local tables. Through such a continuous learning process, the selection of routing path tends to the global optimization,which is the global energy efficient and reliable data transmission path. Our on-going and future work is to make use of the kinetic energy harvesting technology to solve the problem of limited energy in WBANs, and to solve the problem of reliable data transmission in this kind of network.

Acknowledgements The work was supported by Shanghai Municipal Natural Science Foundation (Grant No.18ZR1401200).

Availability of data and material Not applicable.

Declarations

Conflicts of interest The authors have no relevant financial or nonfinancial interests to disclose.

Code availability Not applicable.

References

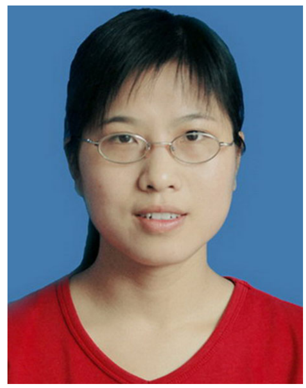
1. Fort, A., Ryckaert, J., & Desset, C. (2006). Ultra-wideband channel model for communication around the human body. *IEEE Journal on Selected Areas in Communications, 24*(4), 927–933.
2. Zuhra, F. T., Bakar, K. A., Ahmed, A., & Tunio, M. A. (2017). Routing protocols in wireless body sensor networks: A comprehensive survey. *Journal of Network and Computer Applications, 99*, 73–97.
3. Tang, Q., Tummala, N., Gupta, S. K., & Schwiebert, L. (2005). TARA: thermal-aware routing algorithm for implanted sensor networks. In *IEEE International Conference on Distributed Computing in Sensor Systems* (pp. 206–217).
4. Bag, A., & Bassiouni, M.A. (2006). Energy efficient thermal aware routing algorithms for embedded biomedical sensor networks. *In Proceedings of the Mobile Ad hoc and Sensor Systems* (pp. 604–609).
5. Bhangwar, A. R., Kumar, P., Ahmed, A., & Channa, M. I. (2017). Trust and thermal aware routing protocol (TTRP) for wireless body area networks. *Wireless Personal Communications, 97*(12), 1–16.
6. Watteyne, T., Auge-Blum, I., Dohler, M., & Barthel, D. (2007). Anybody: a self-organization protocol for body area networks. In *Proceedings of the ICST 2nd international conference on Body area networks* (pp. 641–647).
7. Culpepper, B. J., Dung, L., & Moh, W. M. (2004). Design and analysis of hybrid indirect transmissions (HIT) for data gathering in wireless micro sensor networks. *ACM Sigmobile Mobile Computing & Communications Review, 8*(1), 61–83.
8. Ullah, F., Khan, M. Z., Faisal, M., Rehman, H. U., Abbas, S., & Mubarek, F. S. (2021). Energy efficient and reliable routing scheme to enhance the stability period in wireless body area networks. *Computer Communications, 165*, 20–32.
9. Ruzzelli, A.G., Jurdak, R., Ohare, G.M., & Stok, P.V.D. (2007). Energy-efficient multi-hop medical sensor networking. In *ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments* (pp. 37–42).
10. Braem, B., Latre, B., Moerman, I., Blondia, C., & Demeester, P. (2006). The wireless autonomous spanning tree protocol for multihop wireless body area networks. In *Proceeding of the Third Annual International Conference on Mobile and Ubiquitous Systems: Networking Services* (pp.1–8).
11. Elhadj, H. B., Elias, J., Chaari, L., & Kamoun, L. (2015). A priority based cross layer routing protocol for healthcare applications. *Ad Hoc Networks, 42*, 1–18.
12. Javaid, N., Ahmad, A., Nadeem, Q., Imran, M., & Haider, N. (2015). iM-SIMPLE: Improved stable increased-throughput multi-hop link efficient routing protocol for wireless body area networks. *Computers in Human Behavior, 51*, 1003–1011.
13. Kaur, N., & Singh, S. (2017). Optimized cost effective and energy efficient routing protocol for wireless body area networks. *Ad Hoc Networks, 61*, 65–84.
14. Movassaghi, S., Abolhasan, M., & Lipman, J. (2012). Energy efficient thermal and power aware (ETPA) routing in body area networks. In *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications* (pp. 1108–1113).
15. Ahmed, S., Javaid, N., Yousaf, S., Ahmad, A., Sandhu, M. M., Imran, M., Khan, Z. A., & Alrajeh, N. (2015). Co-LAEEBA: Cooperative link aware and energy efficient protocol for wireless body area networks. *Computers in Human Behavior, 51*, 1205–1215.
16. Liang, L., Ge, Y., Feng, G., Ni, W., & Wai, A. A. P. (2014). A low overhead tree-based energy-efficient routing scheme for multi-hop wireless body area networks. *Computer Networks, 70*, 45–58.
17. Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications* (pp. 20–62). Prentice Hall.
18. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*(1), 237–285.

19. Banuselvasaraswathy, B., & Rathinasabapathy, V. (2020). Self-heat controlling energy efficient OPOT routing protocol for WBAN. *Wireless Networks, 26*(5), 3781–3792.
20. Javaid, N., Abbas, Z., Fareed, M. S., Khan, Z. A., & Alrajeh, N. (2013). M-ATTEMPT: A new energy-efficient routing protocol for wireless body area sensor networks. *Procedia Computer, 19*, 224–231.
21. Srinivas, M. B. (2016). Cluster based energy efficient routing protocol using ANT colony optimization and breadth first search. *Procedia Computer Science, 89*, 124–133.
22. Afsana, F., Asif-ur-rahman, M. D., Ahmed, M. R., Mahmud, M., & Kaiser, M. S. (2018). An energy conserving routing scheme for wireless body sensor nanonetwork communication. *IEEE Access, 6*, 9186–9200.
23. Kumar, K. R. (2020). Energy efficient dynamic cluster head and routing path selection strategy for WBANs. *Wireless Personal Communications, 113*, 33–58.
24. Shimly, S., Smith, D.B., & Movassaghi, S. (2017). Cross-layer optimized routing with low duty cycle TDMA across multiple wireless body area networks. In *IEEE International Conference on Communications* (pp.1–6).
25. Yang, G. S., Wu, X. W., Li, Y., & Ye, Q. B. (2020). Energy efficient protocol for routing and scheduling in wireless body area networks. *Wireless Networks, 26*(2), 1265–1273.
26. Littman, M. L. (2015). Reinforcement learning improves behaviour from evaluative feedback. *Nature, 521*(7553), 445–451.
27. Krieken, E. V., Acar, E., & Harmelen, F. V. (2022). Analyzing differentiable fuzzy logic operators. *Artificial Intelligence, 302*, 1–46.
28. Heinzelman, W.R., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. In *The 33rd Annual Hawaii International Conference on System Science* (pp. 1–10).
29. Astrin, A.(2012). IEEE Std 802.15.6–2012. *IEEE standard for local and metropolitan area networks part 15.6: Wireless body area networks* (pp. 1–271).
30. Benmansour, T., Ahmed, T., Moussaoui, S., & Doukha, Z. (2020). Performance analyses of the IEEE 802.15.6 wireless body area network with heterogeneous traffic. *Journal of Network and Computer Applications, 163*, 1–19.

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

Dr Wenjing Guo received the Ph.D. degree in computer science from East China Normal University (ECNU) in 2013. Since 2014, she has been a teacher in the School of Computer Science and Technology, Donghua University, Shanghai. Her research interests include wireless sensor networks, wireless body area networks, algorithm optimization and artificial intelligence.

Yiran Wang is currently pursuing the M.E. degree in Donghua University, Shanghai. His research interests include wireless body area networks and resource allocation.



Dr Yanglan Gan is a Professor in the school of Computer Science and Technology, Donghua University, Shanghai, China. She received her Ph.D. in Computer Science from Tongji University, Shanghai, China, 2012. Her research interests include data mining, bioinformatics and web services. She has published more than 50 papers on premier international journals and conferences, including Knowledge-bases System, Briefings in Bioinformatics, Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics and Neurocomputing. She served as a program committee member on varieties of international conferences. She worked as a reviewer for Briefings in Bioinformatics, Knowledge-Based Systems, APBC and ADMA.

Dr Ting Lu received the Ph.D. degree in electronic information and electrical engineering from Shanghai Jiao Tong University, China in 2013. She was a visiting researcher with the Department of Electrical and Computer Engineering, State University of New York at Stony Brook from 2018 to 2019. Since 2013, she has been a teacher in the School of Computer Science and Technology, Donghua University, Shanghai. Her research interests include wireless networks and communications, cloud computing and mobile cloud, mobile and distributed computing, and network management and control.