IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON COMMUNICATION AND FOG/EDGE COMPUTING TOWARDS INTELLIGENT CONNECTED VEHICLES (ICVS)

Received November 25, 2019, accepted December 6, 2019, date of publication December 23, 2019,
date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2961331

A Novel QoS-Awared Grid Routing Protocol in the Sensing Layer of Internet of Vehicles Based on Reinforcement Learning

DENGHUI WANG[1], QINGMIAO ZHANG[1], JIAN LIU[1], AND DEZHONG YAO[2]
[1]Department of Information Engineering, East China Jiaotong University, Nanchang 330013, China
[2]RR@NTU Corporate Laboratory, School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Denghui Wang (wangdenghui1856@ecjtu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61862024, and in part by the Scientific and Technological Research Project of Educational Department in Jiangxi Province under Grant GJJ180351.

ABSTRACT This paper proposes a novel Quality of Service (QoS) grid routing protocol in Wireless Multimedia Sensor Networks (WMSN) based on reinforcement learning to guarantee Quality of Service in WMSN based on the sensing layer of the Internet of Vehicles (IoV). The sensing layer of IoV acquires abundant information to handle complex road traffic problems. Moreover, WMSN is rich in perceptual data. This suggests a need for complex acquisition, processing, storage, transfer of text and video data. These issues are elevated due, impart, increased requirements for QoS in WMSN. However, WMSN is heterogeneous, and its network topology is changing dynamically. Therefore, ensuring high QoS in a complex environment has become a challenge. This research suggests that least delay can be accomplished by calculating the distance among the nodes through grid identification number (GID) to acquire the nearest path from the source to the sink. Additionally, optimal grid coordinators with the highest reliability can be elected by making all the nodes in the grid for reinforcement learning to acquire their performance knowledge of reliability and delay. This enables high QoS performance in terms of reliability and end-to-end delay. The results indicate that the QoS of QoS-awared grid routing (QAGR) protocol is higher compared with the traditional grid-based clustering routing protocol.

INDEX TERMS WMSN, grid, reinforcement learning, QoS.

I. INTRODUCTION

Wireless Multimedia Sensor Network (WMSN) has drawn tremendous attention in the scientific literature for its enhanced multimedia sensing capability and dominant data collection and analysis ability [1]–[3]. WMSN is a widely used wireless network system in the area of wireless networking and sensor technologies providing powerful support for a large range of fields distributed wireless communication systems such as Internet-of-Vehicles (IoV)'s sensing layer [4].

With an ever-growing number of traffic accidents and the high frequency of heavy traffic during peak hours, there is an urgent need for vehicles to effectively reduce traffic accidents and ease traffic congestion on the road utilizing exchanging reliable, real-time, safe and efficient information with other vehicles and infrastructures based on vehicular communication [5]. With the rapid development of computation and communication technologies, IoV as an open and integrated network system has been proved that having huge potential to handle safety and efficiency traffic issues with lower overhead costs [6]. There are many approaches to guarantee communication quality [7]–[9]. The application of WMSN in the sensing layer of IoV enables vehicles to communicate with nearby vehicles not only by data but also by image and video information, where the various sensors in the network tend to work together to monitor physical parameters such as temperature, sound, vibration, video and other environmental conditions. In a smart environment, the sensing layer of IoV interacts with WMSN, thereby utilizing diverse networks with heterogeneous sensor nodes in terms of battery capacity, mobility, and processing capabilities to render them smarter and more efficient. Thus, WMSN brings the sensing layer of IoV richer capabilities for both sensing and actuation.

The application of IoV acquires high QoS performance such as in edge computing [10]–[12]. With WMSN, the sensing layer of IoV can acquire, collect, process and compute a variety of complex and dynamic data of humans, vehicles, infrastructures and environments to get a better reliability

The associate editor coordinating the review of this manuscript and approving it for publication was Junhui Zhao.

guarantee [13]. Additionally, recent research suggests that WMSN can be arranged in large cities and highways [14] to monitor traffic flow, making it possible for IoV sensor to obtain a variety of complex road condition information through WMSN to reduce traffic jams [15]. Beyond aiding in the reduction of congestion, a more critical role of WMSN in the sensing layer of IoV is monitoring traffic hazards, especially in real-time emergencies. In such cases, efficient transmission of wireless signals and accurate data acquisition and process is vital in dealing with traffic victims. This suggests a need for the sensing layer of IoV to have the capability for automatic identify as well as storage through WMSN and transmission to interdependent systems such as traffic court systems. The application of WMSN in the sensing layer of IoV offers an effective solution to the heavy traffic jams and provides security for fewer traffic accidents, which requires the sensing devices to be high in computation ability, low in energy consumption, small in size, as well as superior in overhead cost [16]. At the same time, offering real-time and reliable traffic information is important for the transport department and other relevant departments.

WMSN's coverage area is wide that the sink is not always within the data transmission range of each network node. Far away nodes can only transmit data to the sink by establishing multi-hop routes. Thus, the routing design of WMSN should obtain more attention, and apart from those general requirements mentioned above, the routing protocol in this paper is also expected to possess one core capability that is higher QoS including end-to-end delay, reliability and so on [17] especially in the sensing layer of IoV. Since the network studied in this paper is heterogeneous, and all of them can transform into a standby sleep mode according to the energy consumption. For this reason, the topology of the network is changing continuously. Thus it is a tough task for the routing protocol of WMSN to provide reliable and real-time road condition information in such a complex environment requiring highly smart and accurate data traffic.

Moreover, this highly dynamic topology structure is still open to many routing and messages forwarding challenges [18]. Thus, the WMSN routing must be built on a hierarchical architecture that provides an efficient and scalable network structure for collaborating sensor nodes by grouping them into a hierarchy in the complex WMSN where any node displacement may change the entire network topology [19]. Additionally, the grid distribution of nodes can not only solve the problem of data communication among cluster heads but also provides convenience for the failure recovery.

Artificial intelligence (AI) has made a remarkable breakthrough in the field of heterogeneous networks, which has played a vital role in implementing higher efficiency and stronger stability towards big data. The importance of AI applied to the routing protocol of heterogeneous WMSN lies in a practical and efficient approach to making an adaptable judgment to complex environments according to real-time changing information. In particular, reinforcement learning (RL) is an AI-based algorithm that allows the agents to learn, interact and take actions to expect the maximum rewards in the long term.

In this paper, RL approaches to grid-based clustering routing protocol is used to study the reliability and delay in WMSN, which can effectively improve the data delivery rate, reduce the network delay, expend the network communication capacity thus possessing a relatively strong processing ability in the heterogeneous WMSN.

The rest of this paper is organized as follows. Section II discusses the existing clustering routing protocols related to WMSN and the application of RL to the routing protocols of WMSN. Section III provides a detail description of the suggested system model and its assumptions. Section IV presents the specific implementation of the algorithm. Section V shows the simulation results and Section VI concludes this paper.

II. RELATED WORKS

Review of suggests increasing interests in uneven clustering routing protocols in WSN with Low-Energy Adaptive Clustering Hierarchy (LEACH) being considered the new kid on the block. [20] proposes a new energy-efficient protocol based on LEACH considering the network's purpose and providing enhanced performance in terms of QoS. The simulated results indicate that the proposed protocol has a better performance than original LEACH in terms of throughput and latency, there exist several disadvantages. For example, the uneven clusters need to be organized in the set-up phase every round, which can breed too much energy consumption. It is not always a practical theory for clusters to the route with long communication ranges since it is hard to locate in LEACH, which does not apply to the sensing layer of IoV with mobile sensor nodes.

However, geographic forwarding can be used to effectively address these issues. All the location of the network sensor nodes can be known by using GPS as their virtual coordinates. The mode-switched grid-based sustainable routing (MSGR) [21] protocol divides the whole sensing area into virtual equal-sized grids and selects one node per grid as the grid coordinator (GC). Then the routing path to the sink can be established using grid coordinators. This grid-based clustering method avoids the problem of clustering every round and uneven clustering, which is energy-efficient for routing of data packets. However, this routing protocol does not take the case of multiple mobile sink nodes into consideration, which may lead to several control traffic packets.

[22] presents the energy-aware grid-based routing scheme (EAGER) for WSN, which uses a rerouting method to reduce rerouting frequency and also a time-scheduling method to manage the energy consumption of the grid. The election of GC follows the first-mover rule, which means that each node invokes a timer with random intervals and then broadcasts an election packet with its GID, and if the node makes an election attempt before it receives an election packet from any other member, then the node becomes the GC. This approach enhances the network lifetime and stability of homogeneous

WSN. However, the EAGER protocol is limited by the fact that it does not address the heterogeneity of a network.

[23] proposes grid-based clustering and routing algorithms called grid-based fault-tolerant clustering and routing algorithms (GFTCRA), which takes care of the failure of the cluster heads (CHs). The election of the CH is based on the minimal distance from other sensor nodes inside the cluster and also based on its residual energy which is greater than a threshold value. GFTCRA addresses elements of energy efficiency, load balancing and fault-tolerant routing issues together. However, the main disadvantage of this approach has to do with (i) lack of consideration of any delay/hop count for data delivery to the sink and (ii) lack of supporting infrastructure for heterogeneous WSN.

The virtual grid-based data dissemination (VGDD) scheme [24] aims to optimize the tradeoff between network lifetime and data delivery performance while adhering to the low-cost theme of WSN. The criterion of the GCs election is based on the distance that the node close to the center of the grid. This approach allows a limited number of GCs taking part in the routers readjustment process causing minimal network control overhead while preserving near-optimal data delivery routes. However, this approach does not support the heterogeneity of the VGDD in WSN.

The present research places emphasis on WMSN and dynamic complex environment. Unlike the planar model studied in the virtual grid, the actual sensing layer of IoV is spacial model involving the transmission of any two distant nodes too hard to complete. Achieving QoS guarantee in heterogeneous WMSN for tiny sensing nodes is a novel but challenging task. Therefore, searching for a new method to address this issue is urgent. In [25], researchers have found that AI techniques can achieve outstanding performance in terms of dynamic, heterogeneous, large-scale, and complex wireless network to optimize networks in diverse scenarios and complicated environments. AI techniques open a new research direction for the heterogeneous networks with complex structures and dynamic topology.

[26] also highlights the excellent advantages of AI techniques, which can be successfully applied both at the routing protocol of network and node level, thus enabling intelligent behaviors and adaptivity. In particular, [26] proposed an effective AI-based approach called reinforcement learning (RL) that enables a decision-maker to observe, learn, and take actions in its operating environment to increase its accumulated reward. And the simulation results prove that the use of RL technique can largely improve the network performance significantly.

Energy and quality-of-service aware-based cooperative communication routing protocol is proposed in [27]. The proposed algorithm is an energy and QoS aware routing protocol for that it ensures better performance in terms of end-to-end delay and packet loss rate and takes into account the consumed energy through the network. This routing protocol is designed for the wildfire application, which is similar to the sensing layer of IoV researched in this paper, where QoS

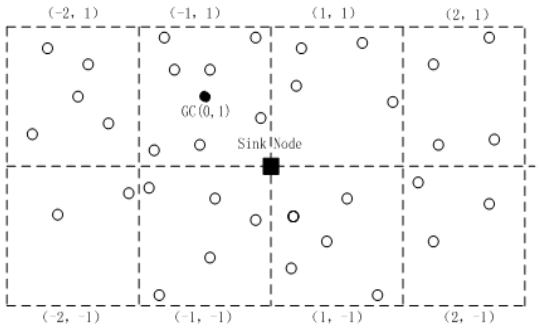


FIGURE 1. Grid construction.

is the core factor that should be taken into consideration in a complex environment.

[28] proposes a scalable scheme for an energy-efficient data flow control algorithm that yields the energy-efficient strategy by using the RL technique. The main novelty of this algorithm is the division of the routing problem into two subproblems by adopting the cluster structure respectively, which allows the achievement of scalability and the application of heterogeneous approaches to each problem. The proposed scheme is adaptive to the dynamic changes of the network topology and traffic generation pattern and yields near-optimal results.

III. SYSTEM MODEL AND ASSUMPTIONS

This section articulates a number of necessary assumptions for the suggested WMSN model. First, we assume that the wireless multimedia sensor nodes are randomly distributed in the network, and each node has a single omnidirectional antenna. All the nodes are battery-powered. Second, we assume that the links between any two sensor nodes is subject to narrowband Rayleigh fading, propagation path-loss, and additive white Gaussian noise (AWGN). It is rational to assume that the channel fading of different links is statistically independent of each other because nodes are usually spatially separate. Also, we assume that the nodes use any CSMA/CA based medium access control protocol to transmit over a single communication channel. The sensing nodes are not mobile, and once the overall arrangement of the entire sensor network is determined, the positions of the sensor nodes will not be changed, so does the location of the base station. Also, each node knows the position of the base station.

We also assume that the topology of this network model adopts a hierarchical grid-based network structure. Both the nodes and the sink are supplied with a GPS positioning module, through which the position information of them can be obtained. The simulation area that the nodes are distributed in is a rectangle. As demonstrated in Fig.1, the simulation network area is divided into several virtual grids with the same size, and each of them is identified by a unique grid identification number (GID). Each node determines its GID through GPS. And all the sensor nodes located in the same

grid cell share the same GID. Information exchanges can be achieved between the nodes. In each grid, there exists a GC, which is responsible for disseminating data among other GCs and managing all the members in its grid. Since all the other members in the grid do not participate in the process of routing, they keep their sensing channel on and turn off their radio until they sense any stimuli generated from an external event.

Since WMSN studied in this paper is heterogeneous, the capability of each node is different. The field research is the sensing layer of IoV, therefore the nodes in the heterogeneous WMSN are expected to have the ability to transmit text information as well as video information. Based on this, this paper divides the initial energy of nodes into three different energy levels categorized into 1, 2 and 3 energy levels. The higher the number, the higher the energy level. Therefore, 1-energy-level (E_1) corresponds to the low energy level, 2-energy-level (E_2) corresponds to the medium energy level, and 3-energy-level (E_3) corresponds to the high energy level. E_1 can only afford to deliver text information from the sender to the receiver without extra energy for image and video information. And E_2 is not only fit for transmitting simple text information, but also suitable for delivering abundant image and video information. While E_3 has enough energy to handle high definition pictures.

It is of practical significance to handle high definition pictures at the node ports to achieve real-time in the sensing layer of IoV. Also, each node is designed to know its energy level as well as its remaining energy at any given time. And the nodes can decide whether to go into standby sleep modes according to energy consumption. In this case, the topology of the network changes dynamically, making it impossible to elect fixed GCs. Therefore, an adaptive solution to the dynamic changes in the environment is required.

The total energy consumption in WMSN can be expressed mathematically as follows:

$$E = E_1 + E_2 + E_3 \tag{1}$$

So the remaining energy is defined as follows:

$$E_{res} = 1 - E \tag{2}$$

And the energy consumption of the nodes belonged to E_1 when transmitting a single s-bit packet is given by Equation (3).

$$E_1 = s(E_{receive} + E_{send}) + E_{RL} \tag{3}$$

where, $E_{receive}$ is the energy used for receiving per packet, E_{send} is the energy used for sending per packet, and E_{RL} is the energy used for running RL algorithm.

The energy consumption of the nodes belonged to E_2 when transmitting a single s-bit packet can be written as Equation (4).

$$E_2 = s(E_{receive} + E_{send} + E_{store}) + E_{RL} + E_{route} \tag{4}$$

where, E_{store} is the energy used for storing per packet, and E_{route} is the energy used for routing.

And the energy consumption of the nodes belonged to E_3 when transmitting a single s-bit packet can be written as Equation (5).

$$E_3 = s(E_{receive} + E_{send} + E_{store} + E_{process}) + E_{RL} + E_{route} \tag{5}$$

where, $E_{process}$ is the energy used for processing per packet.

Since the sensing layer of IoV acquires high reliable and real-time information, the QoS in this model is only involved in reliability and delay. The reliability- and delay-constrained sensing data packets are transmitted in a multihop way from the source to the sink all the time. Network packets have the the same size. And each data packet is restricted to a predefined reliability requirement (C_{rel}) and delay-deadline (C_{del}) by the application layer, where, C_{rel} is the reliability metric, which defines the minimum percentage of packets that should be transmitted to the sink for the reliability; and, C_{del} is the delay metric, which defines the allowable maximum delay percentage of packets that should be transmitted to the sink. C_{rel} and C_{del} vary over time. This leads to the following QoS model expression:

$$C_{QoS}(t) = (1 - \alpha)C_{rel}(t) + \alpha C_{del}(t) \tag{6}$$

where, α is the weight factor, and in this model $\alpha = 0.5$. t is an arbitrary time slot, $C_{QoS}(t)$, $C_{rel}(t)$, $C_{del}(t)$ are all between 0 and 1.

IV. GRID-BASED CLUSTERING ROUTING USING REINFORCEMENT LEARNING

The ultimate goal of the grid-based clustering routing protocol using lightweight reinforcement learning routing algorithm is to achieve high QoS. The network can calculate the nearest path quickly through the source to the sink by the GID, which can guarantee the shortest delay in the network. And then select the GCs with the highest reliability to transmit information. Meanwhile, the election for the GCs should not be too frequent. Excessive election will not only lead to unnecessary energy waste, but also cause too much delay. Therefore, this problem can be divided into two phases to discuss respectively: the first phase is grid construction; and the second one is RL-based grid coordinators election. The following will highlight the solution for the two phases.

A. GRID CONSTRUCTION PHASE

There exists a limitation on the size of each grid. If the size of grids is too small, the QoS guarantee and the scalability of the grid cannot be obtained. Whereas if the size of grids is too large, data collection and transmission inside a grid and among grids will lead to unnecessary source wasting. Thus the distance between any two adjacent grids should not exceed to the maximum transmission distance.

Suppose that the maximum transmission distance of the node is R, then the side length of the grid L can be expressed as follows:

$$L = \frac{R}{2\sqrt{2}} \tag{7}$$

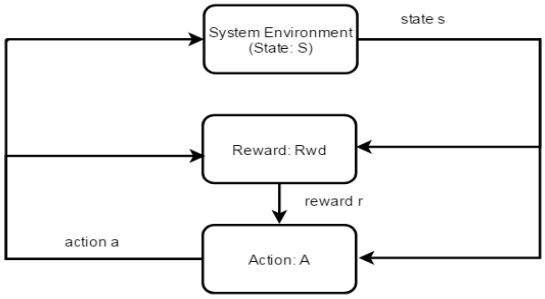


FIGURE 2. The reinforcement learning (RL) system.

This allows it to communicate directly with its eight adjacent grid cells via radio channels. We use a pair of numbers to identify the GID. And the location of the sink node is elected as GID (0,0) for the initial location of the grid. Then the relative geographic location information for the geometric center of each grid is calculated by the sink node according to the location of itself. The location information of the sink node and the geographic location information for the geometric center of each grid are sent to all the nodes within the grid area by broadcast, and then the geographic location information of each node and the geographic location information of the sink node are compared to derive the relative geographic location of the node k. The GID of the grid area that the node k itself belongs to can be calculated by Equation (8).

$$GID(NL_k) = \{(X, Y) | X = \lfloor (x_k)/L \rfloor, Y = \lfloor (y_k)/L \rfloor\} \quad (8)$$

where, L is the side length of the grid, (x_k, y_k) is the location of the node k, and (X, Y) is the GID of the grid that the node k belongs.

The node k can broadcast the geographic location information of the geometric center of its own grid and its GID to all the neighbor nodes within the transmission range. And the distance of any two nodes can be calculated through GID, so the shortest path can be obtained which allows the least delay in the network.

B. RL-BASED GRID COORDINATORS ELECTION PHASE

This paper proposes a reinforcement learning method to elect grid coordinators. Through reinforcement learning, the GCs with the highest reliability can be elected to complete the task of transmission. Reinforcement learning is a kind of learning from environment state to action mapping, so that the cumulative reward values of action from the environment can be maximized. Its main idea is to find the optimal action strategy through trial and error. Reinforcement learning is an online learning technique without requiring prior training examples.

The reinforcement learning algorithm proposed in this paper is a model-free reinforcement learning algorithm, which means that it can directly use the data obtained from interacting with the environment to improve its action. The reinforcement learning framework consists of three modules, namely, state perceptron, action selector and learner. As is shown in Fig. 2, the state perceptron maps the environment state (s) to the internal perception of the agent; the action selector selects actions to act on the environment according to the current strategy; the learner updates the strategy knowledge of the agent according to the reward value (r) of the environment state and the internal perception; the environment under the action (a) will lead to a change in the environment state (s). If an action of the agent leads to a positive reward from the environment, the trend of the agent to produce this action in the future will be strengthened; otherwise, the trend will be weakened.

Then define the system of reinforcement learning based on this paper. Here the agent is sensor nodes in the grids. And the dynamic environment is the wireless channel characteristics and data traffic flows. The state, action and reward are defined as follows:

$$State(s) : s \in S \quad (9)$$

where, S is a set of the feasible policies.

Then we define a set of action A and its element action a. By taking an action, a state may move from one to another. After the selected action is applied, the QoS performance of the changed policy can be estimated. And if the QoS is improved, then the reward would be positive; otherwise, it would be negative. The reward is restored in the action preference matrix $Q(s, a)$, which means the reward that the agent acquires after applying action a under the state s.

$$\begin{aligned} Q(s, a) &= \eta^{\pi}(s, a) \\ &= \lim_{N \to \infty} E\{\frac{1}{N} \sum_{t=0}^{N-1} r(s_k, a_k) \\ &\quad | s_0 = s, \pi\} \end{aligned} \quad (10)$$

where, $\eta^{\pi}(s, a)$ is the average reward function, s_0 is the initial state, s_k is the state at the time of k, a_k is the the action at the time of k, $r(s_k, a_k)$ is the reward at the time of k, and π is the optimal policy to make the maximum average reward.

By iterating this process, the approach to electing the GC for the highest QoS performance can be obtained. The optimal cumulative reward value can be written as follows:

$$Q^*(s, a) = \gamma \sum_{s \in S} T(s, a, s')(r(s, a, s') + \max_{a'} Q^*(s', a')) \quad (11)$$

where, $Q^*(s, a)$ is the cumulative reward that the agent acquires after selecting the action a under the state s, $T(s, a, s')$ is the state transition function when applying the action a. γ is the discount factor, and $0 \leq \gamma \leq 1$.

And the reward is updated according to the Equation (12).

$$\begin{aligned} Q(s_k, a_k) \leftarrow\ & Q(s_k, a_k) + \alpha(r_{k+1} \\ & + \gamma \max_{a} Q(s_{k+1}, a_{k+1}) \\ & - Q(s_k, a_k)) \end{aligned} \quad (12)$$

where, a timestamp k is denoted to indicate the elapsed number of iteration times from the beginning. When an iteration

Algorithm 1 The GCs Election Based on Reinforcement Learning

01: **Initialization:** Initialize S, Q_0, Q_i
02: $\forall s \in S, \forall a \in A, k = 0$
03: **Loop**
04: **for** every node **in** the grid
05: **repeat**
06: **Choose** a_k in s_k
07: Calculate

$$Q^*(s,a) = \gamma \sum_{s \in S} T(s,a,s')(r(s,a,s') + \max_{a'} Q^*(s',a'))$$

08: **Get** cumulative reward from action
09: **Update**Q

$$Q(s_k,a_k) \leftarrow Q(s_k,a_k) + \alpha(r_{k+1} + \gamma \max_a Q(s_{k+1},a_{k+1}) - Q(s_k,a_k))$$

10: For all the Q_i
11: Execute **Choose** a_k for Q_i
Get cumulative reward

$$V(s_{k+1},\pi^*) = \max_{a \in A} Q^*(s_k,a_k) = \lim_{N \to \infty} E\{\sum_{t=0}^{N-1} r(s_t,\pi(s_t)) - \eta^{\pi} \mid s_0 = s, \pi\}$$

Transfer Bellman Equation

$$V(s_{k+1},\pi^*) = r(s,\pi(s)) - Q(s,a) + \sum_{s' \in S} P(s' \mid s,\pi(s)) V^{\pi}(s')$$

Update Q_i
12: Move to the next state s_{k+1}

$$k \leftarrow k+1$$

13: **end Loop**
14: get the optimal policy

$$\pi^*(s_k,a_k) = \arg\max_{a \in A} Q^*(s_k,a_k) = \arg\max_{a \in A} \{r(s,a) - Q^*(s,a) + \sum_{s' \in S} P(s' \mid s,a) V^{\pi^*}(s')\}$$

15: weigh the $C_{rel}(t)$ of the node to determine whether it can be the GC in the grid
16: **until** all the nodes in the grid are traversed once
17: The GCs for the first round are all elected
18: **Loop**
19: Calculate

$$t_s = \frac{E_n - E_m}{P - K(d_n - d_o)}$$

20: **until** the recent GC runs for t_s
21: **for** every node **in** the grid
22: **repeat** step 06-16
23: The new GCs for the next round are all elected
24: **until** all the nodes in the grid are dead
25: **end Loop**
26:**end**

is over, the state s_k transits to the next state s_{k+1} and a new action a_{k+1} is selected.

After building the system, the next is to define an objective function to determine what the optimal action is from a long-term point of view, which usually expressed as the value function of the state or the state-action pair. The objective function of this paper is expressed as follows:

$$V(s_{k+1},\pi^*) = \max_{a \in A} Q^*(s_k,a_k) = \lim_{N \to \infty} E\{\sum_{t=0}^{N-1} r(s_t,\pi(s_t)) - \eta^{\pi} \mid s_0 = s, \pi\} \quad (13)$$

where, $V(s_{k+1},\pi^*)$ is the objective function to make the maximum average reward when taking the optimal policy π^*, and $\pi(s_t)$ is the policy at the state of s_t.

Then $V(s_{k+1},\pi^*)$ satisfies Bellman Equation.

$$V(s_{k+1},\pi^*) = r(s,\pi(s)) - Q(s,a) + \sum_{s' \in S} P(s' \mid s,\pi(s)) V^{\pi}(s') \quad (14)$$

where, $P(s' \mid s,\pi(s))$ is the state transition function, which means that the process at the state of s would transfer from the state of s to the state of s' at the probability of $P(s' \mid s,\pi(s))$ after executing the action $\pi(s)$.

If the objective function is determined, the optimal action can be determined according to the following formula.

$$\pi^*(s_k,a_k) = \arg\max_{a \in A} Q^*(s_k,a_k) = \arg\max_{a \in A} \{r(s,a) - Q^*(s,a) + \sum_{s' \in S} P(s' \mid s,a) V^{\pi^*}(s')\} \quad (15)$$

The above equations have established a relationship to the knowledge about reliability. In the process of solving the optimal strategy, the value of $C_{rel}(t)$ is constantly updated. By solving Equation (15), the optimal strategy of the node can be obtained, as well as the optimal GCs in the grid to guarantee the highest reliability performance can be elected.

Meanwhile, the overhead of the frequent GCs election should be taken into consideration. The initial GCs election can be carried out according to the above steps. What we need to pay attention to is when to elect new GCs. The solution for this problem can be expressed through an equation estimating the time for a GC to run stably as follows.

$$K(d_n - d_o)t_s + E_n = Pt_s + E_m \quad (16)$$

where K is the coefficient, d_o is the distance between any non-cluster node and the original GC, d_n is the distance between this non-cluster node and the new elected GC, t_s is the time for a GC to run stably, E_n is the energy consumption of establishing route after new GC is elected, E_m is the energy consumption of monitoring, and P is the energy consumption

per unit time when the cluster is running stably. The value of K should not be too large or too small. The left of Equation (16) expresses the energy consumption of changing a GC, and the right of it expresses the energy consumption of running stably. When the value of the left is larger than that of the right, it means that the energy consumption of changing GCs is higher than not changing GCs, which is too frequent to change GCs. On the contrary, if the value of the left is smaller than that of the right, it means that the change of the GCs reduces energy consumption. Only when the value of the left is equal to that of the right that can we get the minimal time for a GC to run stably. The equation are as follows.

$$t_s = \frac{E_n - E_m}{P - K(d_n - d_o)} \tag{17}$$

According to the above expression, we can decide when to elect new GCs after the first round. The steps of the algorithm are as follows.

V. EVALUTION METRICS

In this section, the performance of the QoS-awared grid routing (QAGR) protocol is to be evaluated via simulations with respect to the following metrics.

A. QoS

QoS can be evaluated with several metrics such as data delivery ratio, end-to-end delay, reliability and so on, while in this paper, QoS is only the service level parameter of end-to-end delay and reliability. And the measurement of QoS is defined as Equation (6) in Section III.

B. ENERGY CONSUMPTION

Energy consumption is measured as the lifetime of the network. When the lifetime decreases to zero, it means that the total energy in the network is all run out. The lifetime is defined as follows:

$$p = \frac{N_{res}}{N} \tag{18}$$

where N_{res} is the number of remaining nodes in the network, and N is the number of the total nodes initially in the network.

VI. SIMULATION EXPERIMENTS

In this section, the performance of the QoS-awared grid routing (QAGR) protocol is evaluated. The extensive simulations are performed using NS-2 to validate the results. Considering the heterogeneity of WMSN, we have taken a network area of 200m×200m as the sensing operation and set different number of sensor nodes with different energy levels into this area.

In this simulation, we have conducted four groups of experiments, of which two groups are placed with 200 sensor nodes, while the other two groups are placed with 400 sensor nodes. We set up two different proportion for the nodes with three different energy levels. The number of different energy levels of nodes and the density in every group is given in TABLE 1. And the setting of the simulation environment is in the same simulation environment. QoS and energy consumption are the main performance criteria for our research study, thus we compare the QoS and energy consumption with those of the traditional grid routing (TGR) protocol. And since the distribution of the nodes in the simulation area is random, every group of experiments has to be done 10 times. And the simulation results are the average values of the 10 experiments.

TABLE 1. Simulation parameters.

Group \ Number	total	E_1	E_2	E_3	density
1	200	120	60	20	0.005
2	200	80	60	60	0.005
3	400	240	120	40	0.01
4	400	160	120	120	0.01

TABLE 2. Simulation settings.

Specification	Parameters
Network area size	200m × 200m
Deployment type	Random
Network architecture	Heterogeneous
Initial node energy	10J
Buffer size	50
Radio range	100m
Sensing radius	52m
Link layer trans.rate	512Kbps
Transmit power	7.214 W

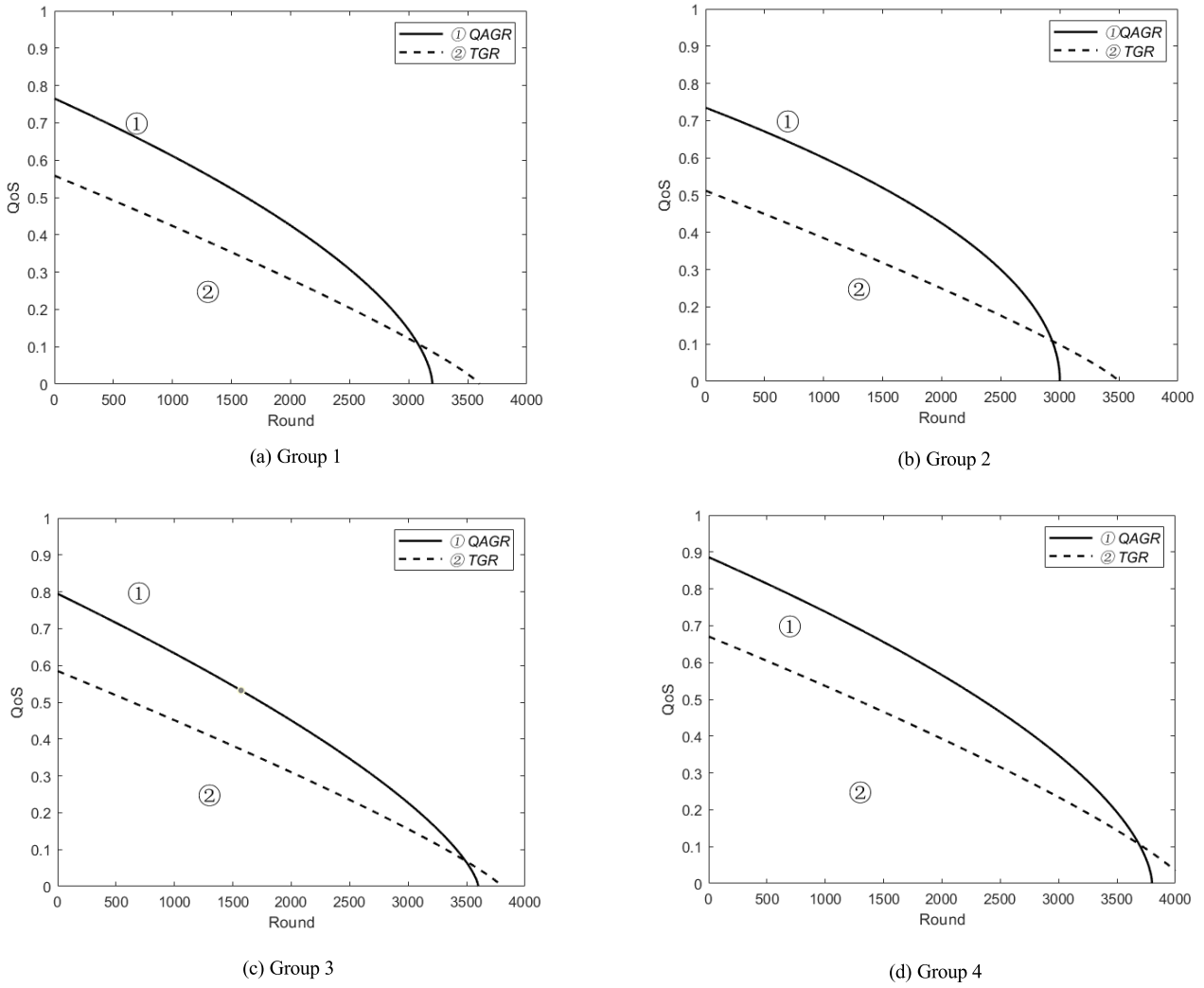
A. QoS

The comparison of the average QoS performance in the four groups is represented in Fig. 3. As is shown in Fig. 3, the X-axis represents the running rounds, and the Y-axis represents the average QoS performance of the nodes in the network.

From Fig. 3 (a) we can see, there are considerable differences in the QoS among the QAGR and TGR protocol. At the beginning, the QoS of the QAGR protocol is always higher than that of the TGR protocol. This is because that by making all the nodes in Vertical lines are optional in tables. Statements that serve as caption grid to run RL algorithm, the optimal GCs can be elected to provide high QoS and low delay to the sink, thus the QoS will be improved substantially. However, after about 2900 rounds, there is a sharp drop in the QoS of the QAGR protocol for the reason that nearly all the nodes in the area are too weak to afford to run RL algorithm.

As is shown in Fig. 3 (b), the QAGR protocol performs better when the number of the nodes with E_3 energy level increases. With the increase of the nodes with E_3 energy level, the initial QoS of both the two routing protocols is increasing, so as to the round that both of them fall to zero.

Comparing with Fig. 3 (a), we can see that the QoS performance in the Fig.3 (c) has improved due to the increase of a



(a) Group 1

(b) Group 2

(c) Group 3

(d) Group 4

FIGURE 3. Comparison of the average QoS.

total number of nodes in the network. And Fig. 3 (d) shows that with the increase of the number of the nodes with E_3 energy level, the QoS performance at the beginning of the two routing protocols is both improved. By comparison we can find that increasing both the number of nodes with stronger ability and the number of overall nodes in the network will lead to an increase to QoS.

B. ENERGY CONSUMPTION

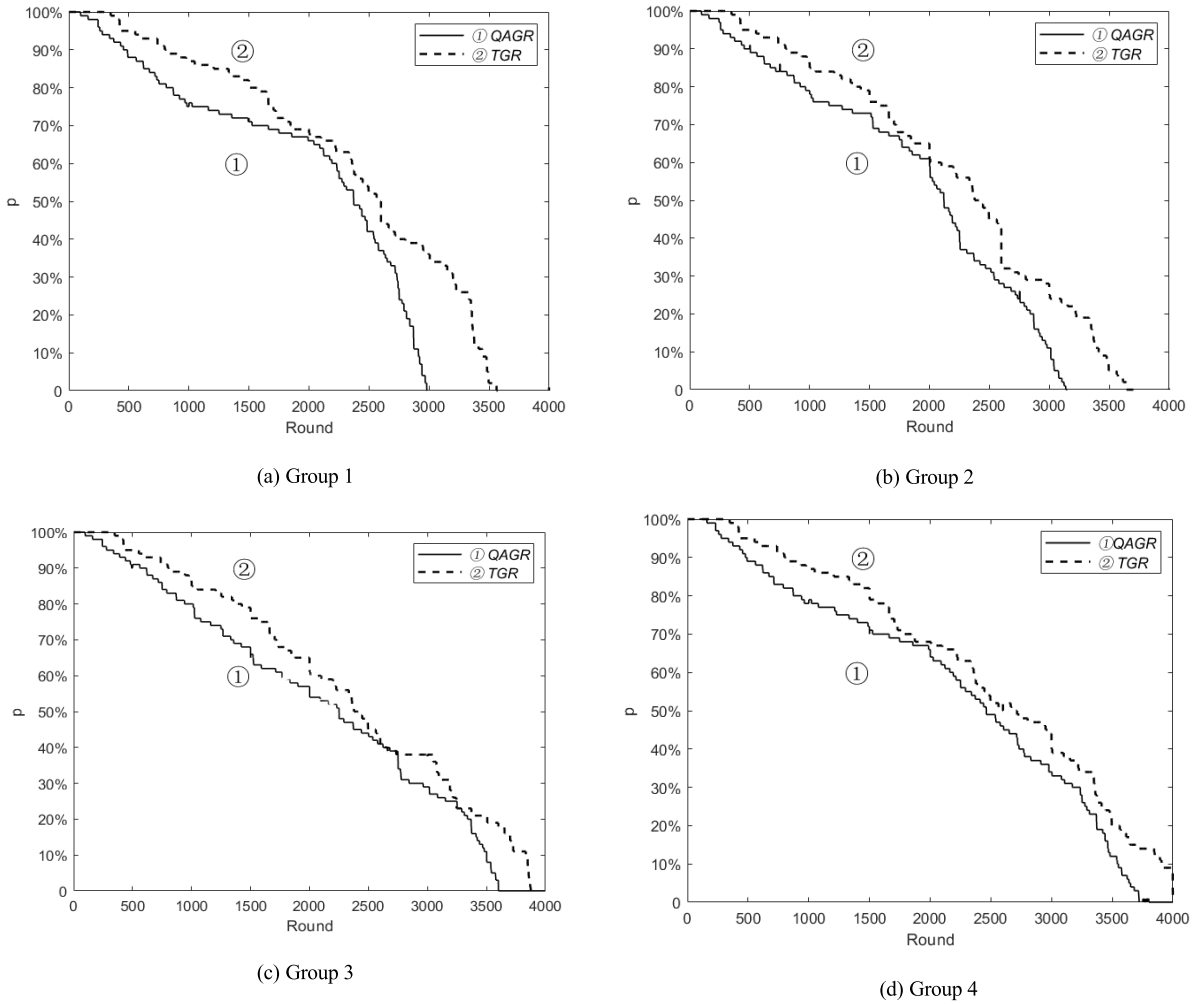
The energy consumption performance comparisons between our proposed QAGR and the TGR protocol is carried out for varying running rounds. Fig. 4 shows the comparison of the average percentage of surviving nodes in the four groups. From Fig. 4 we can see, the X-axis shows the running rounds, and the Y-axis shows the lifetime of the network.

Fig. 4 (a) represents that the lifetime of the TGR protocol is always higher than that of the QAGR protocol. This is because that the QAGR protocol consumes a lot of energy to run RL algorithm to maintain the QoS guarantee while the TGR protocol has no necessity to do that. And Fig. 4 (b), Fig. 4 (c) and Fig. 4 (d) also show this pattern. From Fig. 4 (a) we can see the nodes of the QAGR protocol all died after about 3000 rounds and the nodes of the TGR protocol all died after about 3450 rounds.

As is shown in Fig. 4 (b), the lifetime is increasing with the increase of the number of the nodes with E_3 energy level.

And the same conclusion can be got when comparing Fig. 4(c) and Fig. 4 (d).

Fig. 4 (c) and Fig. 4 (d) show that the lifetime of the QAGR protocol is shorter than that of the TGR protocol. In accordance with the approach above, the QAGR protocol consumes more energy comparing to the TGR protocol. Though energy consumption is not the main performance that has been researched in this paper, as what is shown in Fig. 4, running RL algorithm does not add too much burden to the lifetime of the whole network.



(a) Group 1

(b) Group 2

(c) Group 3

(d) Group 4

FIGURE 4. Comparison of the average percentage of surviving nodes.

VII. CONCLUSION

This paper proposes a novel QoS-awared grid routing protocol in wireless multimedia sensor networks based on reinforcement learning. The application background of this routing protocol is the sensing layer of IoV, which acquires information (i.e., texts, pictures and videos). The network can obtain the optimal path guaranteeing the least delay from the source to the sink through calculating the distance via the given GID of all the grids in the network. Reliable requirement is achieved through selection of GCs with high reliability to accomplish the transmission task. The selection criterion for optimal GCs with high reliability is realized through reinforcement learning. By making all the nodes in every grid of the network run RL algorithm, the knowledge of reliability and delay can be obtained so that the optimal GCs can be decided. The time for selecting new GCs depends on the criterion for the energy consumption of changing a new GC. This time is shorter than the energy consumption of running stably in the same amount of time. The proposed routing protocol provides QoS guarantee especially reliability and delay performance in a heterogeneous WMSN. Simulation results show an improvement in QoS in terms of reliability and delay.

REFERENCES

[1] C. Küçükkeçeci and A. Yazici, "Multilevel object tracking in wireless multimedia sensor networks for surveillance applications using graph-based big data," *IEEE Access*, vol. 7, pp. 67812–67832, 2019.

[2] T.-L. Lin, H.-W. Tseng, Y. Wen, F.-W. Lai, C.-H. Lin, and C.-J. Wang, "Reconstruction algorithm for lost frame of multiview videos in wireless multimedia sensor network based on deep learning multilayer perceptron regression," *IEEE Sensors J.*, vol. 18, no. 23, pp. 9792–9801, Dec. 2018.

[3] T. Mekonnen, P. Porambage, E. Harjula, and M. Ylianttila, "Energy consumption analysis of high quality multi-tier wireless multimedia sensor network," *IEEE Access*, vol. 5, pp. 15848–15858, 2017.

[4] H. Zhu and F. Yu, "A cross-correlation technique for vehicle decetions in wireless magnetic sensor network," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4484–4494, Jan. 2016.

[5] K. M. Alam, M. Saini, and A. E. Saddik, "Toward social Internet of vehicles: Concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, Mar. 2015.

[6] O. kaiwartya, A. H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C.-T. Lin, and X. Liu, "Internet of vehicles: Motivation, layered architecture, network model, challenges, and future aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016.

[7] Z. Junhui, Y. Tao, G. Yi, W. Jiao, and F. Lei, "Power control algorithm of cognitive radio based on non-cooperative game theory," *China Commun.*, vol. 10, no. 11, pp. 143–154, Nov. 2013, doi: 10.1109/CC.2013.6674218.
[8] J. Zhao, X. Guan, and X. P. Li, "Power allocation based on genetic simulated annealing algorithm in cognitive radio networks," *Chin. J. Electron.*, vol. 22, no. 1, pp. 177–180, Jan. 2013.
[9] J. Zhao, S. Ni, L. Yang, Z. Zhang, Y. Gong, and X. Yu, "Multi-band cooperation for 5G HetNets: A promising network paradigm," *IEEE Veh. Technol. Mag.*, vol. 14, no. 4, pp. 85–93, Dec. 2019, doi: 10.1109/MVT.2019.2935793.
[10] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7944–7956, Aug. 2019.
[11] X. Ma, J. Zhao, Y. Gong, and X. Sun, "Carrier sense multiple access with collision avoidance-aware connectivity quality of downlink broadcast in vehicular relay networks," *IET Commun.*, vol. 13, no. 8, pp. 1096–1103, Jul. 2019.
[12] Q. Li, J. Zhao, and Y. Gong, "Computation offloading and resource allocation for mobile edge computing with multiple access points," *IET Commun.*, vol. 13, no. 17, pp. 2668–2677, Oct. 2019.
[13] A. Bouchemel, D. Abed, and A. Moussaoui, "Enhancement of compressed image transmission in WMSNs using modified transformation," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 934–937, May 2018.
[14] T. Mekonnen, M. Komu, R. Morabito, T. Kauppinen, E. Harjula, T. Koskela, and M. Ylianttila, "Energy consumption analysis of edge orchestrated virtualized wireless multimedia sensor networks," *IEEE Access*, vol. 6, pp. 2169–3536, 2017.
[15] A. Siddiqa, M. A. Shah, H. A. Khattak, A. Akhunzada, I. Ali, Z. B. Razak, and A. Gani, "Social Internet of vehicles: Complexity, adaptivity, issues and beyond," *IEEE Access*, vol. 6, pp. 62089–62106, 2018.
[16] M. Civelek and A. Yazici, "Automated moving object classification in wireless multimedia sensor networks," *IEEE Sensors J.*, vol. 17, no. 4, pp. 1116–1131, Dec. 2016.
[17] A. Alanazi and K. Elleithy, "Real-time QoS routing protocols in wireless multimedia sensor networks: Study and analysis," *Sensors*, vol. 15, no. 9, pp. 22209–22233, Sep. 2015.
[18] J. M. Kim, H. S. Seo, and J. Kwak, "Routing protocol for heterogeneous hierarchical wireless multimedia sensor networks," *Wireless Pers. Commun.*, vol. 60, no. 3, pp. 559–569, Oct. 2011.
[19] G. Chao, G. Zhao, J. Lu, and S. Pan, "A grid-based cooperative QoS routing protocol with fading memory optimization for navigation carrier ad hoc networks," *Comput. Netw.*, vol. 76, no. 2, pp. 294–316, Jan. 2015.
[20] A. Rozas and A. Araujo, "An application-aware clustering protocol for wireless sensor networks to provide QoS management," *J. Sensors*, vol. 2019, pp. 1–11, Sep. 2019.
[21] S. Sharma, D. Puthal, S. Tazeen, M. Prasad, and A. Y. Zomaya, "MSGR: A mode-switched grid-based sustainable routing protocol for wireless sensor networks," *IEEE Access*, vol. 5, pp. 19864–19875, 2017.
[22] Y.-P. Chi and H.-P. Chang, "SAMS: An energy-aware grid-based routing scheme for wireless sensor networks," *Telecommun. Syst.*, vol. 54, no. 4, pp. 405–415, Dec. 2013.
[23] S. Jannu and P. K. Jana, "A grid based clustering and routing algorithm for solving hot spot problem in wireless sensor networks," *Wireless Netw.*, vol. 22, no. 6, pp. 1901–1916, Aug. 2016.
[24] A. W. Khan, A. H. Abdullah, M. A. Razzaque, J. I. Bangash, and A. Altameem, "VGDD: A virtual grid based data dissemination scheme for wireless sensor networks with mobile sink," *Int. J. Distrib. Sensor Netw.*, vol. 2015, pp. 1–17, Jan. 2015.
[25] X. X. Wang Li and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.
[26] S. Soni and M. Shrivastava, "Novel learning algorithms for efficient mobile sink data collection using reinforcement learning in wireless sensor network," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–13, Aug. 2018.
[27] M. Maalej, S. Cherif, and H. Besbes, "QoS and energy aware cooperative routing protocol for wildfire monitoring wireless sensor networks," *Sci. World J.*, vol. 2013, pp. 1–11, May 2013.
[28] S.-H. Moon, S. Park, and S.-J. Han, "Energy efficient data collection in sink-centric wireless sensor networks: A cluster-ring approach," *Comput. Commun.*, vol. 101, no. 11, pp. 12–25, Mar. 2017.

DENGHUI WANG received the B.S., M.S., and Ph.D. degrees in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2006, 2008, and 2016, respectively. In 2016, he joined East China Jiaotong University, where he was a Lecturer with the School of Information Engineering Departments. He is researching in the area of heterogeneous wireless sensor networks routing design architecture, deployment, and reliability evaluation.

QINGMIAO ZHANG received the B.E. degree in computer application technology from East China Jiaotong University, Jiangxi, China, in 2008, where she is currently pursuing the Ph.D. degree in control science and engineering. Her research interests include train-ground communication technologies in communication-based train-ground communication systems and performance enhancements for wireless communication for train control.

JIAN LIU is currently pursuing the B.S degree in information engineering with East China Jiaotong University, Nanchang. Her research interest includes wireless multimedia sensor networks routing design.

DEZHONG YAO received the B.S. and Ph.D. degrees from the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2006 and 2016, respectively. From 2010 to 2012, he was a Visiting Scholar at HCII, Carnegie Mellon University, Pittsburgh. He is currently a Research Fellow with Nanyang Technological University, Singapore. His research interests include semantic data mining, machine learning, and mobile computing. He is a member of the ACM.