

Research Article

Q-Learning-Based High Credibility and Stability Routing Algorithm for Internet of Medical Things

Kefeng Wei ^{1,2}, Lincong Zhang ³, Xin Jiang,⁴ and Yi Guo⁴

¹School of Computer Science and Engineering, Northeastern University, Shenyang, China

²Shen Kan Engineering and Technology Corporation, MCC., Shenyang, China

³School of Information Science and Engineering, Shenyang Ligong University, Shenyang, China

⁴The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, China

Correspondence should be addressed to Lincong Zhang; lincongz@foxmail.com

Received 26 September 2020; Revised 7 November 2020; Accepted 11 December 2020; Published 28 December 2020

Academic Editor: Xiaojie Wang

Copyright © 2020 Kefeng Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the outbreak of COVID-19, people's demand for using the Internet of Medical Things (IoMT) for physical health monitoring has increased dramatically. The considerable amount of data requires stable, reliable, and real-time transmission, which has become an urgent problem to be solved. This paper constructs a health monitoring-enabled IoMT network which is composed of several users carrying wearable devices and a coordinator. One of the important problems for the proposed network is the unstable and inefficient transmission of data packets caused by node congestion and link breakage in the routing process. Based on these, we propose a Q-learning-based dynamic routing selection (QDRS) algorithm. First, a mathematical model of path optimization and a solution named Global Routing selection with high Credibility and Stability (GRCS) is proposed to select the optimal path globally. However, during the data transmission through the optimal path, the node and link status may change, causing packet loss or retransmission. This is a problem not considered by standard routing algorithms. Therefore, this paper proposes a local link dynamic adjustment scheme based on GRCS, using the Q-learning algorithm to select the optimal next-hop node for each intermediate forwarding node. If the selected node is not the same as the original path, the chosen node replaces the downstream node in the original path and so corrects the optimal path in time. This paper considers the congestion state, remaining energy, and mobility of the node when selecting the path and considers the network state changes during packet transmission, which is the most significant innovation of this paper. The simulation results show that compared with other similar algorithms, the proposed algorithm can significantly improve the packet forwarding rate without seriously affecting the network energy consumption and delay.

1. Introduction

In recent years, there are more and more kinds of diseases, which cause significant trouble for human beings and make people pay more attention to their health. Traditional medical treatment requires patients to go to the hospital and always takes a long time. The medical test results are usually time-consuming and inefficient. Many diseases need continuous monitoring for patients, but traditional medical treatment cannot achieve real-time observation and doctors' decision-making.

The emergence of wearable devices solves these problems. It allows people to monitor their health anytime and

anywhere, thus promoting the Internet of Medical Things (IoMT) [1, 2]. The wearable device-based IoMT has attracted more and more attention and will become a trend that human beings pay attention to their health in the future. The IoMT can not only continuously monitor the physiological information of the human body through wearable devices but also transmit the detection results to the remote monitoring center or family doctor and even realize the emergency alarm. It is worth mentioning that the IoMT can help doctors propose treatment plans through decision support systems [3, 4]. This medical method can significantly reduce medical examination time, improve detection efficiency, and save human resources.

The outbreak of COVID-19 in 2019 makes people worldwide pay more attention to their health. The demand for monitoring, early warning, and transmission to doctors and family members using the IoMT is growing explosively. The mobility of users leads to the continuous change of network topology and also challenges the data transmission. Frequent user mobility will lead to link breakage and degrade network performance. At present, some scholars have studied the routing algorithms for IoMT [5–9]. However, current routing algorithms mainly consider user mobility’s impact on algorithm performance, such as delay, network energy consumption, and network lifetime. If the link is not reliable, it is easy for data loss, retransmission, and other situations to occur, which pose a serious threat to the monitored personnel. Therefore, medical data need a stable and reliable transmission. Some scholars have studied the stability of link transmission to minimize the probability of link breakage [10]. Nowadays, the IoMT monitors not only patients but also the whole society with a wide range of user groups, different roles, and behaviors. Hence, the security of the multi-hop transmission of medical data becomes a great challenge. The author in [11] introduced the node activity in the routing algorithm, preferring to select the node with more connection times as the next hop to ensure the safe and reliable data transmission. However, the algorithm does not consider reducing the link break probability.

We proposed a routing algorithm based on comprehensive link stability, which can find the most stable link between the source node and the destination node, and provide reliable and durable communication between wearable users [12]. However, some problems have not been solved yet. First of all, the comprehensive link stability only considers the link connection duration. Then, the current congestion degree and the residual energy of nodes will also affect the link stability while not being considered. Finally, to pursue link stability, the algorithm allows too many hops.

The most important thing is that although we have established a reliable and stable path from the source node to the destination node, in the process of data forwarding, the state of intermediate forwarding nodes may change. For example, the new forwarding data from other nodes may lead to congestion for the node in the selected path. The moving trajectory of the node changes suddenly, which may bring a bad link even an interrupted link for the previously selected path. Therefore, it is necessary to adjust the path dynamically to adapt to the changing network environment. This paper proposes a Q-learning-based dynamic routing selection (QDRS) algorithm. Firstly, we establish the mathematical optimization model for routing selection. According to the connection duration, the credibility between the current node and its neighbor nodes, the residual energy, and the congestion degree of the neighbor nodes, the GRCS algorithm is proposed to select the optimal path. Node credibility is the number of times nodes communicate with each other. The higher the credibility, the more reliable the node and the more likely it is to provide reliable and stable forwarding. After that, this paper proposes a Q-learning-based local link dynamic (QLLD) algorithm to solve the congestion and link breakage in the path. The Q-learning algorithm is used to select the

optimal next-hop node for each intermediate node and to modify the original optimal path in time to ensure the stability and reliability of the path.

The main contributions of this paper are as follows.

- (1) This paper builds a wireless network named IoMT based on wearable devices and describes the problems of high credibility and stability in data transmission
- (2) We formulate a mathematical model to maximize the credibility and stability of path and propose a global routing optimization routing algorithm with the constraints of node congestion degree, residual energy rate, credibility between nodes, and connection duration of link and hops
- (3) To meet the requirement of a high packet forwarding rate under user mobility, we propose a local link dynamic adjustment method based on the Q-learning algorithm to locally select the optimal next hop for each intermediate node in the selected path. And the results are used to update the selected path. Thus, the waiting delay and transmission interruption because of node congestion and link breakage can be alleviated

The rest of this paper is organized as follows. The system model illustrates the network construction and related parameters in Section 2. Section 3 provides the problem formulation and the optimal path selection algorithm GRCS. Section 4 specifies a local link adjustment method using the Q-learning algorithm. The simulation and performance evaluation are described in Section 5. Finally, Section 6 concludes this paper.

2. System Model

Figure 1 shows that the Internet of Medical Things includes several users carrying a coordinator node and several wearable devices equipped with wireless sensors. These sensors can monitor the physiological information of different parts of the user’s body (such as electroencephalograph (EEG), electrocardiograph (ECG), blood pressure, and body temperature) and the user’s movement information (motion, including speed, direction, and acceleration) and surrounding environment information (temperature, humidity, toxic gas content, etc.). Each sensor periodically or suddenly sends data to the coordinator node according to the data characteristics by itself. The users’ coordinator nodes can exchange or transmit information to the gateway node for remote transmission via the Internet. Therefore, real-time monitoring and early warning notifications of the user’s physical health can be completed between family members or between the users and their family doctor or the hospital monitoring center. In this paper, we only consider the communication between the coordinators. We assume that the network includes N users (i.e., coordinators), and each user wears M sensor nodes and one coordinator. The gateway is randomly placed in the IoMT. Typically, to reduce energy consumption, the coordinator will send data to the nearest gateway node.

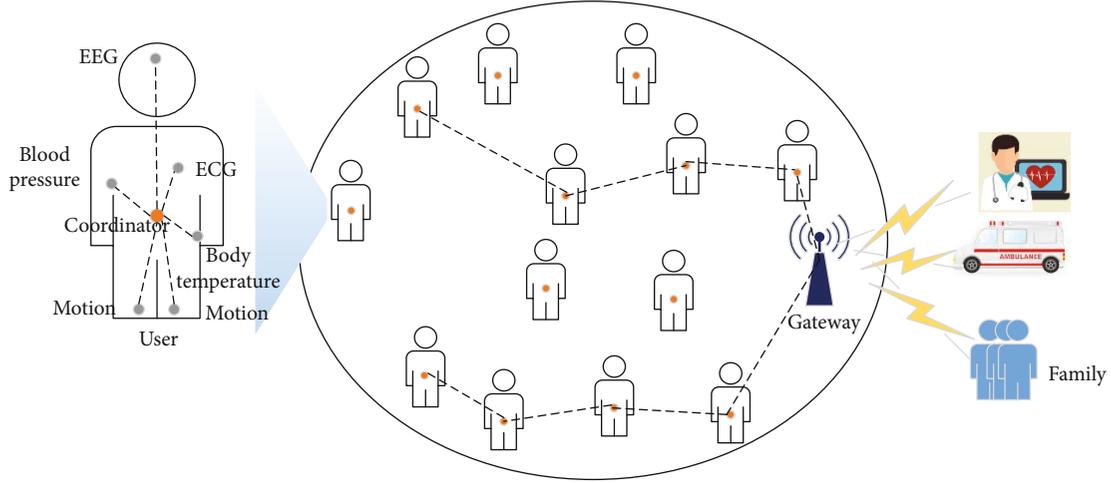


FIGURE 1: Internet of Medical Things.

We assume that there are H hops in the routing path between source node s and destination node d . To simplify the description of the problem, we introduce the following notations:

h : The index of hops,

i, j : Represented any two adjacent nodes of the path from s to d and $j = i + 1$,

(i, j) : Described any link of the path from s to d ,

$\tau_{ij}(h)$: The connection duration of the h_{th} link (i, j) ,

$\tau_{ij(\min)}$: In the path from s to d , the connection duration of the link with the shortest connection duration,

η : Weight coefficient,

m : The number of packets in the buffer of the h_{th} node,

k : The index of packets in the buffer of the h_{th} node,

p_{length}^k : The length of the k_{th} packet in the node buffer,

Q_{\max} : The maximum length of buffer queue,

E_{h_remain} : The residual energy of the h_{th} node,

E_{\max} : The initial energy of the h_{th} node,

N_{ij} : The number of connections between the node i and j ,

N_{th} : The maximum value of the credibility,

$\alpha, \beta, \gamma, \omega$: Different weight factors.

To ensure the routing path's stability, the connection duration of the link denoted by $\tau_{ij}(h)$ is an important factor [12]. Hence, we define the link maintenance L_{sd} from s to d to measure the path's strength, as shown by

$$L_{sd} = \frac{(1 - \eta)\tau_{ij(\min)} + \eta \left(\sum_{i=1}^H \tau_{ij}(h) / H \right)}{\sum_{i=1}^H \tau_{ij}(h)}. \quad (1)$$

The node load state is also a vital influence factor for link stability as the link may break when the packet waits for a too long time due to the high congestion of the node. The node congestion degree of the h_{th} node is represented by λ_h and computed by

$$\lambda_h = \frac{\sum_{k=1}^m p_{\text{length}}^k}{Q_{\max}}. \quad (2)$$

There is no doubt that the node's residual energy is not a negligible factor for link stability because the low-powered node may not finish the packet forwarding. The residual rate of power of the h_{th} node is represented by E_h and computed by

$$\xi_h = \frac{E_{h_remain}}{E_{\max}}. \quad (3)$$

To ensure the safety of the packets, the forwarding node should be trustworthy and will not leak any information significant to medical knowledge. Therefore, we define the credibility to measure the safety of the forwarding node. The credibility of the node i and j denoted by R_{ij} can be computed by

$$\rho_{ij} = \frac{N_{ij}}{N_{th}}. \quad (4)$$

In conclusion, the credibility and stability of the path from the source node s to destination node d are denoted by CS_{sd} and can be computed by

$$CS_{sd} = L_{sd}^{\alpha} \frac{\left(\sum_{h=1}^{H-1} \xi_h \right)^{\gamma} \cdot \left(\sum_{i=1}^{H+1} \rho_{ij} \right)^{\omega}}{\left(\sum_{h=1}^{H-1} \lambda_h \right)^{\beta}}. \quad (5)$$

3. Path Selection with High Credibility and Stability

3.1. Problem Formulation. For the sake of better modeling the credible and stable routing path, we state the problem as follows:

$$\text{Maximize } CS_{sd}. \quad (6)$$

```

if  $i == d$  then
    compute  $L_{sd}$  and  $CS_{sd}$ ;
else
    compute  $d_{i,d}$ ,  $\tau_{ij}(h)$ ,  $\lambda_h$ ,  $\xi_h$ ,  $\rho_{ij}$ ;
    if  $d_{i,d} > d_{s,d} \parallel \lambda_h \leq \lambda_{th} \parallel \xi_h < \xi_{th} \parallel \rho_{ij} < \rho_{th} \parallel H_{sd} > H_{th}$  then
        Reject forwarding;
    else if  $\tau_{ij(\min)} > \tau_{ij}(h)$  then
         $\tau_{ij(\min)} \leftarrow \tau_{ij}(h)$ ;
    end if
     $h \leftarrow h + 1$ ;
     $H_{sd} \leftarrow H_{sd} + 1$ 
    add  $\tau_{ij(\min)}$ ,  $T_{ij}(h)$ ,  $\lambda_h$ ,  $\xi_h$ ,  $\rho_{ij}$  and  $H_{sd}$  into the RREQ packet and broadcast it.
    end if
end if

```

ALGORITHM 1: Judgment for the available node.

Subject to

$$d_{i,j} \leq D_{th}, \quad (7)$$

$$d_{i,d} \leq d_{s,d}, \quad (8)$$

$$\lambda_h \leq \lambda_{th}, \quad (9)$$

$$\xi_h \geq \xi_{th}, \quad (10)$$

$$\rho_{ij} \geq \rho_{th}, \quad (11)$$

$$H_{sd} \leq H_{th}. \quad (12)$$

The objective in (6) is to find the maximum CS by computing (5). The constraint in (7) states that any two communicable nodes should be in the transmission range D_{th} . Equation (8) indicates that the available node should be closer to the destination node than the source node. Equation (9) states that the single node is not very busy. Equation (10) implies that the available node has enough energy to forward packets. Equation (11) indicates the intimate and trustable relationship between two nodes, which should be larger than ρ_{th} . Equation (12) ensures that the path length is no longer than H_{th} .

3.2. GRCS Algorithm. To address the above problem, we propose a traditional algorithm named GRCS, which mainly focuses on selecting available nodes for each hop and delivering node information for each hop. The detail of the GRCS algorithm is as follows.

Step 1. Initialize the related parameters and add them to an RREQ packet. The source node broadcasts the RREQ packet to neighbor nodes in the transmission range. The destination node will reply an RRER packet to the source node after receiving this RREQ packet. Otherwise, go to Step 2.

Step 2. After the neighbor node j receives the RREQ packet from the upstream node i , it will check the information in

the header fields of the RREQ packet. The detailed process for determination is shown in Algorithm 1.

Step 3. After receiving all RREQ packets in a period, the destination node computes L_{sd} and CS_{sd} for each RREQ packet, chooses the path with the largest CS_{sd} as the optimal path, and sends an RREP packet to the source node back to the way the RREQ came.

4. Q-Learning-Based Local Link Dynamic Adapting Algorithm

The data transmission starts after the path $p^*(s, d)$ has been established. However, the intermediate forward node receives the data not only from its upstream node in the path $p^*(s, d)$ but also from other nodes not in the path $p^*(s, d)$, i.e., it may congest. Meanwhile, the users may change their mind to go to another place and result in another motion trail. This sudden change will lead to link breakage. Therefore, we propose the QLLD algorithm to select an optimal node as the new forward node to improve the performance of the transmission.

The Q-learning algorithm is one of the frequently used methods of machine learning and has been used to solve the optimization problem in VANET [13–18], opportunistic networks [19–23], wireless sensor network [24–26], etc. In this paper, we adopt the Q-learning algorithm to select the optimal next-hop selection for the real-time correction for the path selected by the GRCS algorithm. We assume that the coordinator worn by each user is an agent. The cumulative revenue of the agent is affected by the next hop selected by other coordinators. In order to obtain the location, moving speed, direction, residual energy, link state, and additional information of other users, it is necessary to broadcast hello packets periodically between network coordinators for information exchange. The coordinator does not need to know the information of all the coordinators in the network but only needs to ensure that it can receive the information from its neighbors.

We define the neighbor nodes of coordinator x as those nodes which are closer to the destination node than node x and in its communication range. The set of the neighbor nodes of coordinator x is represented by \mathfrak{N}_x ; y is one of the neighbor nodes of coordinator x :

$$\mathfrak{N}_x = \{y \mid d_{x,y} < \min(d_{x,d}, d_{th}), d_{y,d} < d_{x,d}\}. \quad (13)$$

Each coordinator maintains a neighbor node table, and each neighbor node is identified by the node congestion degree, residual energy rate, credibility, and connection duration.

System state: we define that the state $s_x(t)$ is decided by the location $\ell_x(t)$, represented by

$$s_x(t) = (\ell_x(t) \mid y \in \mathfrak{N}_x). \quad (14)$$

Action: the current node x selects the next-hop coordinator denoted by $a_x(t) \in \mathcal{A}_x$.

Reward: node x observes the system status $s_x(t)$; the direct reward obtained by implementing reflection $a_x(t) = \ell$ is $\mathcal{F}_x(t)$, represented by

$$\mathcal{F}_x(t) = \alpha_1 \lambda_y(t) + \beta_1 \xi_y(t) + \delta \rho_y(t) + \sigma \tau_{xy}(t). \quad (15)$$

Among them, α_1 , β_1 , δ , and σ are weighting coefficients and $\alpha_1 + \beta_1 + \delta + \sigma = 1$. The above reward function is defined as the sum of the congestion degree and energy residual rate of node y and the credibility and connection duration between node y and the current node x .

The long-term reward $\mathcal{R}_x^\pi(s)$ obtained by each coordinator is the expected value of the cumulative discount's direct reward, as shown in the following formula:

$$\mathcal{R}_x^\pi(s) = \mathbb{E} \left[\sum_{t=0}^{+\infty} \gamma^t \mathcal{F}_x(s(t), a(t)) \mid s(0) = s, a(0) = a_i \right]. \quad (16)$$

Among them, γ represents the discount rate, which determines the proportion of the direct reward and long-term reward, $0 \leq \gamma < 1$; the greater the γ , the more significant the proportion of the direct reward.

The coordinator as an agent selects action based on the strategy π . Given any coordinator n , in the state s_t , the Q value obtained by selecting a_t according to a specific strategy $\pi(s_t, a_t)$ is defined as $Q_x^\pi(s_t, a_t)$. The strategy is evaluated by Q -learning, in which the Bellman equation is used to obtain the optimal Q value function, expressed in $Q_x^{\pi^*}(s_t, a_t)$, and the calculation is as follows:

$$Q_x^{\pi^*}(s_t, a_t) = \mathbb{E} \left[(R_x(s_t, a_t)) + \gamma \sum_{s_{t+1}} P_{s_t}^{s_{t+1}}(a_{x,t}) \max_{\pi^*} Q_x^{\pi^*}(s_{t+1}, a_{t+1}) \right]. \quad (17)$$

Among them, $P_{s_t}^{s_{t+1}}(a_{x,t})$ is the transition probability from state s_t to state s_{t+1} . The optimal strategy is defined as

$$\pi_x^*(s) = \max_{a_{t+1} \in \mathcal{A}_x} Q_x^{\pi^*}(s, a_{t+1}). \quad (18)$$

Q -learning iteration formula:

$$Q_x(s_t, a_t) = (1 - \alpha) Q_x(s_t, a_t) + \alpha \left[R_{nx}(s_t, a_t) + \gamma \max_{a_{t+1}} Q_x(s_{t+1}, a_{t+1}) \right]. \quad (19)$$

Among them, α is the learning rate and reflects the convergence speed of the iterative process.

For each intermediate forwarding node x , after the neighbor node y with maximum Q -value is selected, we will compare it with the downstream node of x in the original path. If they are not the same, then the downstream node of x is replaced with the selected neighbor node y . In each selected path, the intermediate coordinator node executes the QLLD algorithm and is denoted as node x . The QLLD algorithm is realized by Algorithm 2.

5. Performance Evaluation

The MATLAB software is employed as a simulation platform to verify the effectiveness of the proposed algorithm. In the simulated network, we deploy 80 randomly distributed wearable users in an 80 m \times 80 m area. We set 5 sensor nodes and one coordinator on each user's body, and each user is viewed as a whole and represented by a dot in the network topology. Each user moves in any direction at a speed of 1 m/s. The transmission range D_{th} is set as 20 m. The initial energy of each node E_{max} is set as 100 J, and the threshold ξ_{th} is set as 20 J. The thresholds of node congestion degree λ_{th} , the credibility ρ_{th} , and the maximum hop H_{th} are 0.9, 1, and 7, respectively. The maximum length of the buffer queue Q_{max} is set as 2×10^5 . Other parameters are the same as Ref. [12]. This paper compares the proposed algorithm with the traditional AODV algorithm, the RRLS algorithm [12], and our proposed GRCS algorithm.

First, we show a simulated routing path selected by the algorithms, as shown in Figure 2. From node 16 to node 25, the routing paths of AODV, RRLS, and GRCS are 16-22-25, 16-67-27-50-25, and 16-67-27-25, respectively. We can see that the approach of AODV is the shortest, and that of RRLS is with the most hops. The path of GRCS is similar to that of RRLS with a shorter length. The path selected by the proposed QDRS algorithm is the same as that of GRCS because QLLD do not have any effect on the path from node 16 to node 25, which is not marked in Figure 2. However, during the network running process, the advantages of QDRS can be shown according to the following results.

Figure3 illustrates the performance of the packet forwarding rate varying with running time. After the network begins to run, the data packet amount gradually increases. The congestion starts to occur in some intermediate nodes. When the congestion is serious, some packets may be discarded. And along with the mobility of users, the distances among users change varying time so the link state also

```

Initialize node list  $\mathcal{L}$  for  $p^*(s, d)$ ;
Initialize the location of nodes in  $\mathcal{L}$ ;
for  $i = 1; i < |\mathcal{L}|; i++$  do
   $x = \mathcal{L}(i)$ ;
   $y^* = \mathcal{L}(i+1)$ ;
   $s_x(t) \leftarrow \ell_x(t)$ ;
  Search for the neighbor nodes in communication range and put into  $\mathfrak{N}_x(t)$ ;
   $s_x(t) \leftarrow \ell_x(t)$ ;
  if  $\text{rand} < \epsilon$  then
     $a_x(t) \leftarrow \text{argmax}[Q_x(s_t, a_t)]$ ;
  else
     $a_x(t) \leftarrow \text{argrand}[Q_x(s_t, a_t)]$ ;
  endif
  if  $a_x(t) \neq y^*$  then
     $y^* \leftarrow a_x(t)$ ;
  endif
endfor
for  $i = 1; i < |\mathcal{L}|; i++$  do
   $x = \mathcal{L}(i)$ ;
  Update the network state and node state  $s_x(t+1)$ ;
  Update  $\mathfrak{N}_x(t+1)$ ;
  Update  $Q_x(s_t, a_t)$  using (19);
endfor

```

ALGORITHM 2: Q-learning-based local link dynamic algorithm.

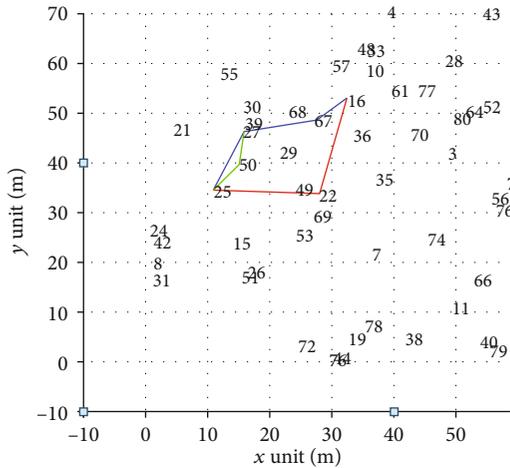


FIGURE 2: The routing path from node 16 to node 25.

becomes uncertain. By selecting the path in the shortest way, the AODV has the lowest packet forwarding rate. The proposed GRCS algorithm and QDRS algorithm obtain a higher packet forwarding rate because of the comprehensive consideration for node congestion, residual energy, credibility, and connection duration. In particular, on account of additional real-time correction for the selected path, the QDRS algorithm can keep the credibility and stability of the path. Therefore, the QDRS algorithm outperforms by 6.8%, 10%, and 64% compared with GRCS, RRLS, and AODV in terms of the packet forwarding rate.

Figure 4 illustrates the performance of the average path delay for each algorithm. The proposed QDRS algorithm updates the optimal path selected by the GRCS algorithm

to reduce the waiting time brought by node congestion and break period caused by link failure. However, it still consumes some nonnegligible time for computing. Therefore, it just performs a little better than the GRCS algorithm for delay. What is more, we can see that QDRS and GRCS provide more stable path delay than AODV and RRLS.

As shown in Figure 5, the network energy consumption of the QDRS algorithm and GRCS algorithm is obviously lower than that of the AODV algorithm and a little more than that of the RRLS algorithm. This is because the residual energy of nodes is not the only consideration and is also not the optimization objective. Additional computing for Q-learning leads to higher energy consumption for the QDRS algorithm than the GRCS algorithm. However, without sacrificing too much energy, we promote the packet forwarding rate greatly.

In order to better verify the advantages of the proposed algorithm, we also simulate and compare the four algorithms changing with communication radius. This is because, in wireless networks, the communication radius is one of the key factors effecting the network performances.

Figure 6 shows the relationship between the packet arrival rate and the communication radius of the four algorithms. It can be seen from Figure 7 that the packet arrival rate of the four algorithms increases before the communication radius reaches 20 m. When the communication radius reaches 20 m, the packet arrival rate of the four algorithms is the maximum. Compared with AODV, RRLS, GRCS, and QDRS consider the residual energy of nodes and other factors, so the link is more stable, so they have better performance in the packet arrival rate. In addition, the performance of the GRCS and QDRS algorithms is better than that of the RRLS algorithm. When the communication radius

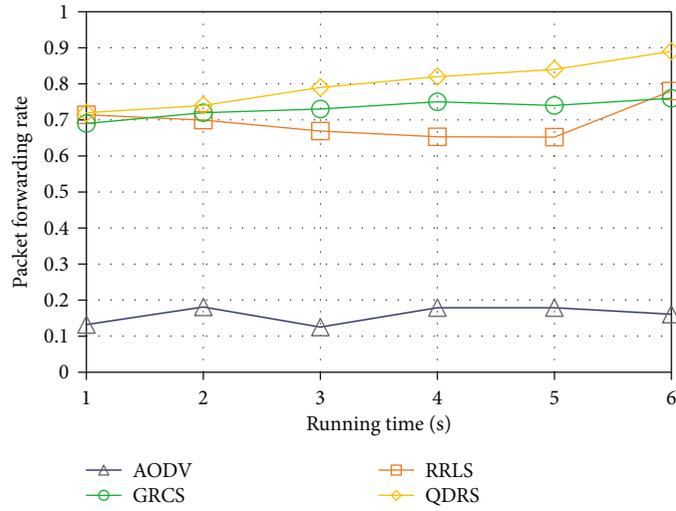


FIGURE 3: The packet forwarding rate.

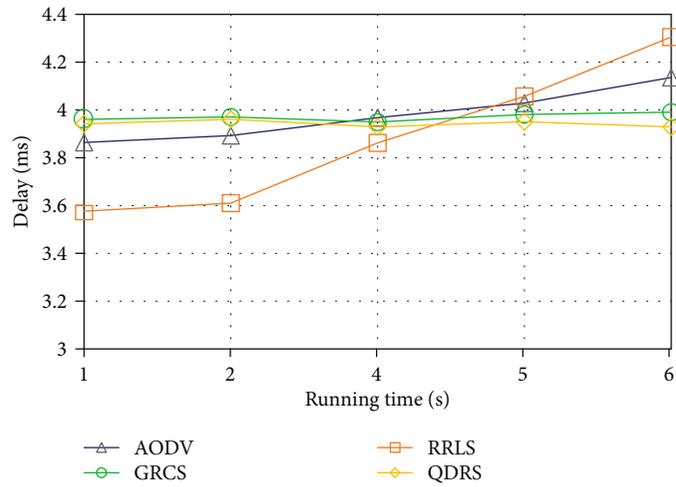


FIGURE 4: Comparison of path delay.

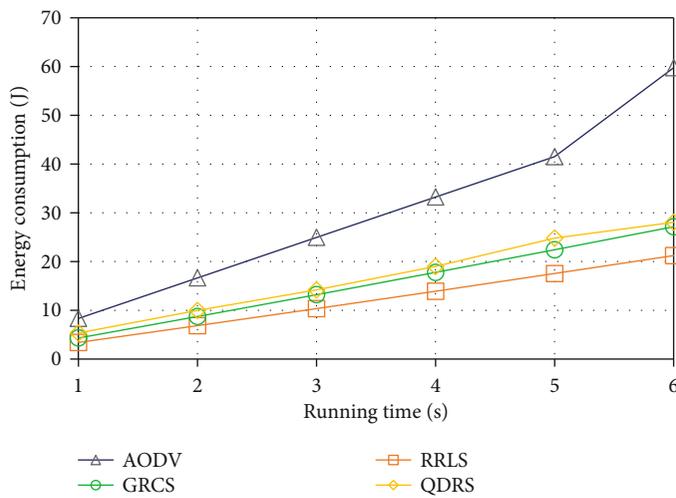


FIGURE 5: Energy consumption.

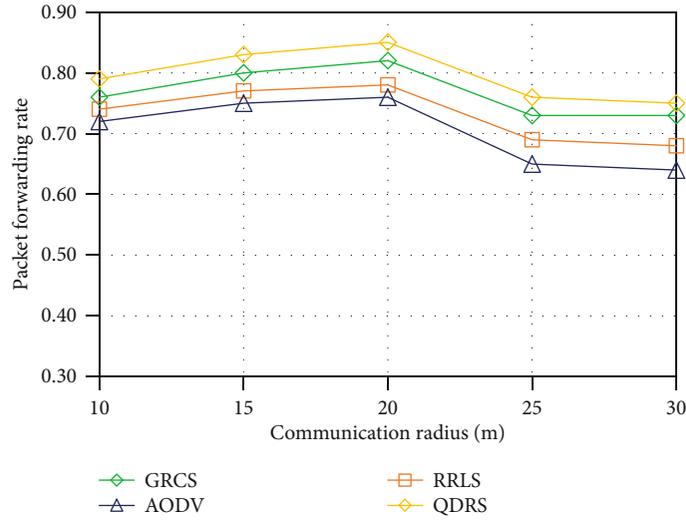


FIGURE 6: Packet forwarding rate varying with communication radius.

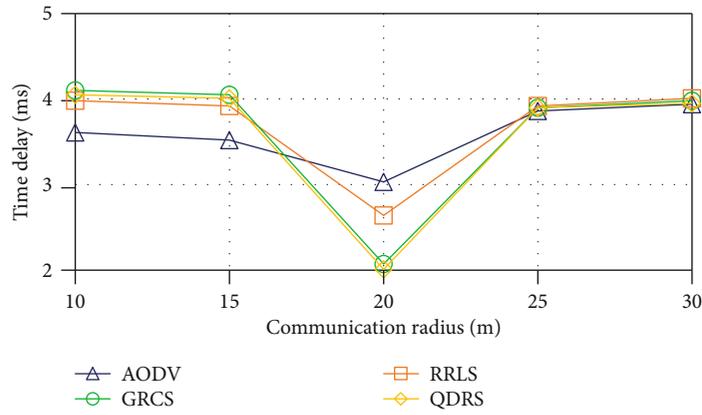


FIGURE 7: Delay varying with communication radius.

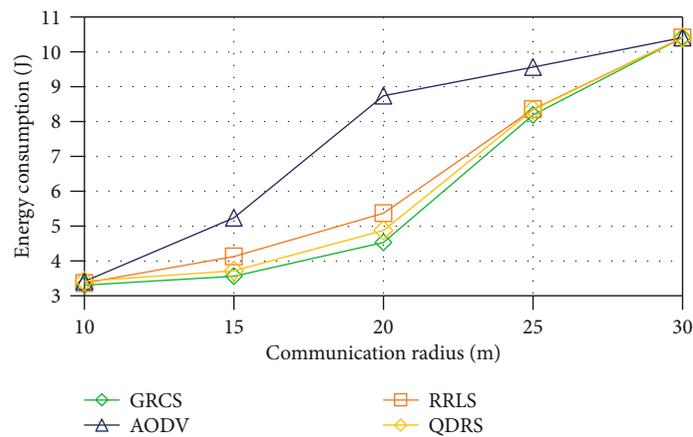


FIGURE 8: Energy consumption varying with communication radius.

continues to increase, it can be seen that the packet arrival rate of the three algorithms has decreased. The main reason is that with the increase of the communication radius, the number of nodes in the communication range increases, which makes the communication traffic of each node rise

and cannot guarantee accurate transmission. Therefore, the packet arrival rates of the algorithms begin to decline, while the decline range is not very large. In general, our proposed QDRS still performs best among the four algorithms due to the path adjustment.

Figure 7 shows the relationship between the delay and the communication radius of the four algorithms. We can see that AODV has better delay performance when the communication radius is small because of the shortest path. With the increase of communication radius, the delays of four algorithms decrease before the communication radius reaches 20 m. When the communication radius is between 10 m and 15 m, the delays decline and the descent speeds of four algorithms are relatively gentle; while the communication radius reaches 15–20 m, the delays of the four algorithms decrease largely. When the communication radius reaches about 20 m, four algorithms achieve the optimal delay performance. Among them, the delays of GRCS and QDRS algorithms are almost the same and lower than the other two algorithms. When the communication radius continues to increase, the number of nodes in the communication range increases, so the traffic loads of nodes increase and then bring the increasing delay. In summary, the best communication radius for the four simulated algorithms is about 20 m.

Figure 8 shows the performance of energy consumption of four algorithms varying with the communication radius. It can be seen that with the increase of the communication radius, the energy consumptions of the four algorithms increase due to the increase of transmission power of nodes. Compared with the AODV algorithm, the growth trend of energy consumption of the other three algorithms is more gentle, where QDRS and GRCS have better performance than RRLS. Due to the modification of the path, some nodes need to establish a connection with the newly selected next hop, so the energy consumption of QDRS is slightly higher than that of GRCS. When the communication radius reaches 20 m, the energy consumption gap between the three algorithms and the AODV algorithm reaches the maximum. When the communication radius reaches 30 m, the energy consumption of the three algorithms tends to be consistent. This is because with the increase of the communication radius and no variation of node density, there is no big difference in the selection of the next-hop node among the four algorithms. Meanwhile, in order to maintain the communication range, more energy consumption is needed, so the energy consumption increases. In addition, due to the large communication range, the link breakage between nodes decreases; QDRS and GRCS have similar performance in energy consumption.

6. Conclusion

This paper studied the routing algorithm for IoMT and proposes a two-step solution to ensure the reliability and stability of network transmission. First, the credibility and stability of the path is the optimization goal, and the communication distance, node congestion, node residual energy rate, inter-node credibility, and internode hops are constrained to construct a mathematical optimization model, and the GRCS algorithm is proposed to find the optimal path. On this basis, the QLLD algorithm based on Q-learning is used to find the optimal next-hop node for the intermediate node to update the optimal path in time. Hence, it prevents the deterioration of the link status during the packet transmission and ensures the credibility and stability of the path. We use MATLAB to

simulate the proposed algorithm, and the simulation results show the effectiveness of the proposed algorithm.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the National Natural Science Foundation of China, Grant/Award Number: 61501308, and the Postdoctoral Research Station project of Shenyang Ligong University.

References

- [1] Z. Ning, P. Dong, X. Wang et al., "Mobile edge computing enabled 5G health monitoring for Internet of medical things: a decentralized game theoretic approach," *IEEE Journal on Selected Areas in Communications*, pp. 1–16, 2020.
- [2] S. Huang, B. Guo, and Y. Liu, "5G-oriented optical underlay network slicing technology and challenges," *IEEE Communications Magazine*, vol. 58, no. 2, pp. 13–19, 2020.
- [3] S. Durga, R. Nag, and E. Daniel, "Survey on machine learning and deep learning algorithms used in internet of things (IoT) healthcare," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1018–1022, Erode, India, 2019.
- [4] T. Gao, X. Li, Y. Wu et al., "Cost-efficient VNF placement and scheduling in public cloud networks," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4946–4959, 2020.
- [5] B. P. Santos, O. Goussevskaia, L. F. M. Vieira, M. A. M. Vieira, and A. A. F. Loureiro, "Mobile matrix: routing under mobility in IoT, IoMT, and social IoT," *Ad Hoc Networks*, vol. 78, pp. 84–98, 2018.
- [6] W. Almobaideen, R. Krayshan, M. Allan, and M. Saadeh, "Internet of things: geographical routing based on healthcare centers vicinity for mobile smart tourism destination," *Technological Forecasting and Social Change*, vol. 123, pp. 342–350, 2017.
- [7] S. Huang, C. Yang, S. Yin, Z. Zhang, and Y. Chu, "Latency-aware task peer offloading on overloaded server in multi-access edge computing system interconnected by metro optical networks," *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5949–5961.
- [8] A. M. Fathollahi-Fard, A. Ahmadi, F. Goodarziyan, and N. Cheikhrouhou, "A bi-objective home healthcare routing and scheduling problem considering patients' satisfaction in a fuzzy environment," *Applied Soft Computing*, vol. 93, article 106385, 2020.
- [9] S. Yin, S. Huang, B. Guo et al., "Shared-protection survivable multipath scheme in flexible-grid optical networks against multiple failures," *IEEE/OSA Journal of Lightwave Technology*, vol. 35, no. 2, pp. 201–211, 2017.

- [10] A. Serhani, N. Naja, and A. Jamali, "AQ-routing: mobility-, stability-aware adaptive routing protocol for data routing in MANET-IoT systems," *Cluster Computing*, vol. 23, no. 1, pp. 13–27, 2020.
- [11] Y. Liu, *Research on the Key Technology for Wireless Body Area Networks*, Beijing University of Posts and Telecommunications, 2017.
- [12] L. Zhang, X. Chen, K. Wei, W. Zhang, and Y. Feng, "Body-to-body network routing algorithm based on link comprehensive stability," in *2019 28th Wireless and Optical Communications Conference (WOCC)*, pp. 1–5, Beijing, China, 2019.
- [13] Z. Ning, P. Dong, X. Wang et al., "Partial computation offloading and adaptive task scheduling for 5G-enabled vehicular networks," *IEEE Transactions on Mobile Computing*, p. 1, 2020.
- [14] X. Wang, Z. Ning, and S. Guo, "Multi-agent imitation learning for pervasive edge computing: a decentralized computation offloading algorithm," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 2, pp. 411–425, 2021.
- [15] F. Khan and S. K. Nguang, "Location-based data delivery between vehicles and infrastructure," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 288–296, 2020.
- [16] X. Wang, Z. Ning, S. Guo, and L. Wang, "Imitation learning enabled task scheduling for online vehicular edge computing [J]," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2020.
- [17] Z. Ning, K. Zhang, X. Wang et al., "Joint computing and caching in 5G-envisioned internet of vehicles: a deep reinforcement learning-based traffic control system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–12, 2020.
- [18] Z. Ning, K. Zhang, X. Wang et al., "Intelligent edge computing in internet of vehicles: a joint computation offloading and caching solution," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–14, 2020.
- [19] B. Guo, S. Huang, Y. Shang et al., "Timeslot switching-based optical bypass in data center for intra-rack elephant flow with an ultrafast DPDK-enabled timeslot allocator," *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 10, pp. 2253–2260, 2019.
- [20] V. Vashishth, A. Chhabra, and D. K. Sharma, "GMMR: a Gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic IoT networks," *Computer Communications*, vol. 134, pp. 138–148, 2019.
- [21] D. K. Sharma, J. J. P. C. Rodrigues, V. Vashishth, A. Khanna, and A. Chhabra, "RLProph: a dynamic programming based reinforcement learning approach for optimal routing in opportunistic IoT networks," *Wireless Networks*, vol. 26, no. 6, pp. 4319–4338, 2020.
- [22] D. K. Sharma, S. K. Dhurandher, I. Woungang, R. K. Srivastava, A. Mohananey, and J. J. P. C. Rodrigues, "A machine learning-based protocol for efficient routing in opportunistic networks," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2207–2213, 2016.
- [23] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, "kROp: k-means clustering based routing protocol for opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 4, pp. 1289–1306, 2019.
- [24] G. Künzel, L. S. Indrusiak, and C. E. Pereira, "Latency and lifetime enhancements in industrial wireless sensor networks: a Q-learning approach for graph routing," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5617–5625, 2019.
- [25] Z. Zhuang, J. Wang, Q. Qi, H. Sun, and J. Liao, "Toward greater intelligence in route planning: a graph-aware deep learning approach," *IEEE Systems Journal*, vol. 14, no. 2, pp. 1658–1669, 2019.
- [26] W. Jin, R. Gu, and Y. Ji, "Reward function learning for q-learning-based geographic routing protocol," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1236–1239, 2019.