Topology Control in Emerging Mobile Networks

Lead Guest Editor: Naixue Xiong
Guest Editors: Sajid Hussain and Jaime Lloret

Topology Control in Emerging Mobile Networks

Topology Control in Emerging Mobile Networks

Lead Guest Editor: Naixue Xiong
Guest Editors: Sajid Hussain and Jaime Lloret

Editorial Board

Contents

Research Article

An Enhanced PEGASIS Algorithm with Mobile Sink Support for Wireless Sensor Networks

Jin Wang [ID],[1] **Yu Gao,**[2] **Xiang Yin,**[2] **Feng Li,**[1] **and Hye-Jin Kim** [ID][3]

[1]*School of Computer & Communication Engineering, Changsha University of Science & Technology, China*
[2]*School of Information Engineering, Yangzhou University, China*
[3]*Business Administration Research Institute, Sungshin W. University, Republic of Korea*

Correspondence should be addressed to Hye-Jin Kim; hye-jinkim@hotmail.com

Energy efficiency has been a hot research topic for many years and many routing algorithms have been proposed to improve energy efficiency and to prolong lifetime for wireless sensor networks (WSNs). Since nodes close to the sink usually need to consume more energy to forward data of its neighbours to sink, they will exhaust energy more quickly. These nodes are called hot spot nodes and we call this phenomenon hot spot problem. In this paper, an Enhanced Power Efficient Gathering in Sensor Information Systems (EPEGASIS) algorithm is proposed to alleviate the hot spots problem from four aspects. Firstly, optimal communication distance is determined to reduce the energy consumption during transmission. Then threshold value is set to protect the dying nodes and mobile sink technology is used to balance the energy consumption among nodes. Next, the node can adjust its communication range according to its distance to the sink node. Finally, extensive experiments have been performed to show that our proposed EPEGASIS performs better in terms of lifetime, energy consumption, and network latency.

1. Introduction

Topology control is a very important research issue for the emerging mobile networks (EMN). Wireless sensor networks (WSNs), as part of the EMN, are usually composed of a large collection of tiny sensors nodes and they are developing very rapidly in recently years due to their wide applications [1]. These nodes are usually deployed in a random way and they can collect information from surroundings, and then transfer the data package to sink node using single or multiple hops communication to form WSNs. The outstanding performance of WSNs like fault tolerance, rapid deployment, self-organizing, timely response, etc. makes WSNs widely used in harsh environment such as military surveillance, industrial product line monitoring, medical health care, and smart homes [2, 3].

Energy efficiency and energy balancing have been a very hot and challenging research issue for WSNs for many years. Since large number of sensor nodes is deployed under very harsh environment, it is unrealistic to change the sensor

batteries for them. Besides, the nodes close to the sink not only need to collect data, but also need to forward its neighbours' data. Thus, those nodes will exhaust their energy more quickly than other nodes far away from sink. This phenomenon is known as hot spots problem. When the first node dies because of energy exhausting, the performance of network in terms of connectivity, coverage, lifetime, etc. will decrease sharply [4, 5].

In order to solve the hot spots problem, sink mobility technology is introduced to many routing protocols [6–9]. By adopting sink mobility, the following advantages can be achieved. Firstly, the mobile sink moves along a certain trajectory so that the nodes close to the sink could take turns to be the forwarder, which will largely alleviate hot spots problem [10]. Secondly, the energy consumption of the whole network can be reduced because of the shorter average transmission distance between sensor nodes and mobile sink, if the mobile sink has a proper trajectory [11]. Thirdly, the performance of the network in terms of transmission latency and network throughput can be greatly improved under

suitable moving strategy. Finally, mobile sink can enhance the network connectivity in sparse network [12–14].

The main contribution in this paper includes the following four aspects. First, we find the optimal communication distance for data transmission. Second, we set threshold value for each node to protect those nodes with low remaining energy. Third, we adjust nodes' communication distance according to its distance to the mobile sink. Finally, mobile sink technology is adopted to balance the energy of different regions. Numerous simulations are conducted to prove our proposed algorithm outperforms than some existing work.

The rest of the paper is organized as follows. In Section 1, we introduce the background of our work. In Section 2, we discuss some classic routing protocols and some latest research achievement. The system model which contains network and energy model is presented in Section 3. In Section 4, we describe our algorithm in detail. Extensive simulations are conducted and the experimental results are discussed and compared in Section 5. Section 6 concludes the paper with some future work.

2. Related Work

Many researchers have paid close attention to energy efficient routing protocols or algorithms in recent years. LEACH (Low-Energy Adaptive Clustering Algorithm) is one of the most famous hierarchical protocols which were proposed about twenty years ago [15]. In LEACH, there are two types of nodes, namely, cluster heads (CHs) and ordinary nodes (ONs). Each ON collects data from area of interest and sends the data package to a closest CH. Each CH take charge of fusing the data it receives and then transmits the fused data to the sink node. The introduction of CHs can avoid long distance communication between ON and sink node; thus much energy is saved. However, the selection of CHs is in a random way as a result of random selection. So, the CH distributes unevenly inside the whole network with degraded performance. Meanwhile, CHs communicate with sink node directly which causes much energy dissipation.

PEGASIS (Power Efficient Gathering in Sensor Information Systems) is a chain-based routing protocol for WSNs [16]. In PEGASIS, each sensor node only needs to transmit data to its neighbour which is closer to the sink node. Several chains could be constructed according to the greedy algorithm and the leader of each chain takes the responsibility to transfer the data to the sink node. Due to the heavy burden of the leaders of chains, each node takes turns to be the leader to balance the energy consumption. Because of multihop propagation, long distance communication between sensor node and sink node is avoided and much energy is saved. Meanwhile, due to multihop propagation, the delay of the network is very serious and it is not suitable for delay sensitive applications.

In [17], the authors propose an energy efficient routing protocol using mobile sink based on clustering and it is suitable for WSNs with obstruction. Mobile sink moves along the CHs and collects data by single hop communication. In this protocol, an efficient scheduling mechanism based on spanning graphs is proposed to find a shortest path for

mobile sink to avoid obstacle. Simulation result shows that the lifetime of the network is prolonged and complexity of network is reduced.

In [18], the authors proposed a data collection algorithm using mobile sink based on tree clustering. This algorithm contains three phases, namely, tree construction, tree decomposition and subrendezvous points (SRP) selection, and data collection phases. Mobile sink moves towards rendezvous points (RP) and SRP to collect data via single hop communication. After data collection in each round, sensor nodes will reselect RP and SRP. Simulation result proves that the proposed algorithm performs better in terms of network lifetime and the moving path of the mobile sink is short.

In [19], the authors study the event-driven application for WSNs with mobile sinks. In this paper, each sensor node has two statuses: monitoring and transmission status. When an event is caught, the status of node changes from monitoring to transmission and the group of active sensor nodes (ASN) is constructed to forward the data to the mobile sink. In large-scale WSNs, the area of interest is divided into several subareas due to the limit of the speed of the mobile sink. Each subarea contains a mobile sink and ASNs are selected to collect and transmit data. In the meantime, the continuous and optimal trajectory (COT) can be calculated for the mobile sink to achieve better performance.

In [20], the authors propose an energy efficient routing protocol for mobile sensor networks using a path-constrained mobile sink. This protocol is suitable for WSNs which limits on the moving path of the mobile sink, such as mobile sinks deployed in bus. Source node sends the data package to the target node which is closest to the location where the mobile sink will arrive next time and ensure the shortest path to transmit data. This kind of protocol is very suitable for delay-tolerant network and it possesses higher robustness and lower energy consumption.

In addition, some heuristic algorithms are used to improve the performance of the network [21–26]. In [21], the authors combine the Particle Swarm Optimization (PSO) algorithms with cluster technology to enhance the network lifetime. In [22], the authors propose an Ant Colony Optimization (ACO) based on clustering algorithm to find an optimal moving trajectory for mobile sink. In [23], techniques like Glow-worm Swarm Optimization (GSO), clustering and mobile sink are combined together to improve energy efficiency as well as to prolong the lifetime of wireless sensor network.

3. System Model

3.1. Basic Assumptions. In this paper, we make some basic assumptions as follows:

(1) All the sensor nodes are randomly deployed and keep static after deployment.

(2) Each sensor node has a unique ID to differ from each other.

(3) All the sensor nodes have the same initial energy and batteries of them cannot be changed.
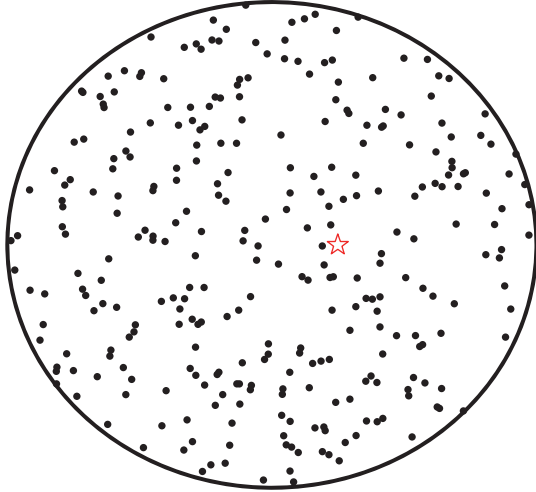
☆ mobile sink
• sensor node

FIGURE 1: Network model.

(4) Mobile sink can move freely and owns unconstrained energy.

(5) Transmission power of sensors can be adjusted based on communication distance [27].

3.2. Network Model. In this paper, N nodes are deployed randomly in a circular field with radius R. These sensor nodes are denoted as $\{n_1, n_2, n_3 \cdots n_n\}$, respectively, as is shown in Figure 1. In each round, every sensor node should transmit one package to the sink using single hop or multihop communication based on the relatively distance.

3.3. Energy Model. The first ratio model is adopted here to calculate the energy consumption [28, 29]. As is shown in Figure 2, the energy consumption mainly contains two parts which are transmission consumption and reception energy consumption.

Transmission consumption mainly includes the energy consumed for transmission circuit and the power amplifier expenditure. The reception consumption mainly includes the energy of reception circuit expenditure. Transmission consumption can be calculated by using the following formula:

$$E_{Tx}(l, d) = \begin{cases} l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2 & if, d < d_0 \\ l \cdot E_{elec} + l\varepsilon_{mp} \cdot d^4 & if, d \geq d_0 \end{cases} \quad (1)$$

where E_{elec} is the energy cost to use transmitter or receiver circuit to process one-bit signal. Symbols ε_{fs} and ε_{mp} denote the amplification coefficient for the free space model and the multipath fading model separately. Metric d_0 is the threshold value and can be calculated as $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$.

The energy consumption for reception can be calculated according to formula (2):

$$E_{Rx}(l) = l \cdot E_{elec} \quad (2)$$

4. Our Proposed Algorithm

4.1. Overview of EPEGASIS. In this section, we will illustrate our proposed algorithm in detail. Our proposed algorithm is called Enhanced PEGASIS (EPEGASIS), which belongs to the chain-based routing algorithms. In EPEGASIS, each node uses optimal communication distance to choose the relay node from its neighbours for data transmission. In order to avoid excessive energy consumption for some nodes with special location, we set a protection mechanism according to the average residual energy of its neighbours. A mobile sink is utilized to gather data for different region energy consumption balancing. The workflow of each node in the network is shown as Figure 3.

4.2. Study on Optimal Communication Distance. According to the energy consumption model, the energy consumption raises rapidly with the increasing of the communication distance. In order to conserve energy, multihop transmission is adopted to avoid long distance communication. However, too much hops may increase the burden of forwarding and cause increased end-to-end delay or latency.

We assume that the source node is m meters away from the mobile sink and it sends a one-bit data package to the mobile sink by k hops, and the total energy consumption can be calculated using formula (3):

$$E_{total}$$
$$= \begin{cases} k \left[E_{elec} + E_{fs} \left(\dfrac{m}{k} \right)^2 \right] + (k-1) E_{elec} & if \ \dfrac{m}{k} < d_0 \\ k \left[E_{elec} + E_{mp} \left(\dfrac{m}{k} \right)^4 \right] + (k-1) E_{elec} & if \ \dfrac{m}{k} \geq d_0 \end{cases} \quad (3)$$

The relationship between total energy consumption and hop counts can be shown as Figure 4. From Figure 4, we can clearly see that, as the hop counts increases, the total energy consumption drops first to reach its lowest point and then rises. The triangle in each line denotes the energy consumption when $m/k = d_0$ and it also denotes the lowest point of each line. Therefore, the optimal communication distance and optimal hop counts is shown as formulas (4) and (5).

$$d_{optimal} = d_0 \quad (4)$$

$$optimal\_hop\_counts = \left\lceil \dfrac{m}{d_0} \right\rceil \quad (5)$$

4.3. Initial Phase of the Network. After nodes deployment, the network enters into initial phase. During the initial phase, the main object is to exchange information to prepare for data transmission. At the very beginning, the mobile sink broadcast a NETWORK_INIT package to remind all the nodes to initial the network. Each node maintains a route table called "NBR_TABLE" to record its neighbours' information. Then each node broadcast a NODE_INFO package with its maximal transmission distance. Neighbours who receive NODE_INFO package will write relevant information into route table. After the initial phase, each node will have

Package type	Content
NETWORK_INIT	The initial location of the mobile sink, the moving speed and direction of the mobile sink, time of duration of each round, TDMA schedule of nodes.
NODE_INFO	The location, the residual energy, ID of source node.
ENERGY_INFO	The residual energy and ID of source node.
SINK_INFO	System time, the current location of mobile sink, the speed and direction of mobile sink.
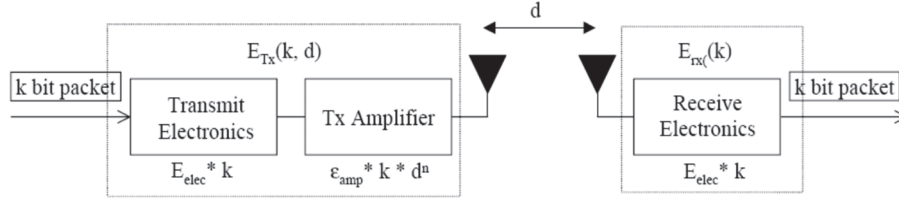
FIGURE 2: Energy model.

the knowledge of its neighbours. In order to update the energy information of sensors, an ENERGY_INFO package is broadcasted by each node during each round. With the running of the network, the system time and the location of the mobile sink may go awry, so the mobile sink will broadcast a SINK_INFO package in the whole network at fixed periods. The detailed information of packages is illustrated in Table 1.

4.4. Network Topology Generation. In PEGASIS, each node selects its closest neighbour as relay node for data forwarding and that causes heavy burden of transmission and long network delay. In our proposed algorithm, we mainly decrease the times of data forwarding and achieve an optimal energy consuming. We use optimal communication and optimal hop counts for data transmission. Node selects a forwarder among its routing table and the forwarder satisfies the following two formulas.

$$d\left(n_{forwarder}, \text{mobile\_sink}\right) < d\left(n_{source}, \text{mobile\_sink}\right) \quad (6)$$

$$
\begin{aligned}
&d\left(n_{source}, n_{forwarder}\right) \\
&\quad = \min\left|d\left(n_{source}, n_{forwarder}\right) - d_{optimal}\right|
\end{aligned}
\quad (7)
$$

where $d(n_i, n_j)$ denotes the distance between node i and node j.

When the mobile sink is in the optimal communication of sensors, they will transmit data to the mobile sink directly. After relay node selection, nodes far away from the mobile sink transmit data to it by several relays and the topology of the network is generated. The topology of the network is shown as Figure 5.

4.5. Protection Mechanism for Dying Node. Nodes close to the sensor field centre take on a heavy burden for data forwarding; however, nodes close to the edge of the sensor field only need to send their own data and that cause the hot

spots problem. Figure 6 describes the hot spots phenomenon after the network running for several rounds.

In order to protect nodes from dying too early, we set a threshold value for nodes according to average neighbour energy. When node's residual energy is less than a threshold value, it will not take the role of forwarder and only sends its own generated data. The threshold value is calculated using formula (8).

$$E_{threshold} = \frac{\sum_{n \in C_i} E_{residual}}{N} \quad (8)$$

where C_i denotes the set of the neighbours of node i, N is the number of its neighbours, and $E_{residual}$ is the current energy of node n.

By means of setting threshold value, nodes in the central area are protected and time of first node die is greatly delayed. During the initial round, all nodes' residual energy reach threshold value and all of them can be chosen as relay node. With round increasing, nodes close to the central area own lower residual energy than threshold value; therefore, nodes close to the edge will gradually become relay nodes and cause ring routing. The ring routing is illustrated as Figure 7.

4.6. Communication Distance Adjustment for Edge Node. Nodes in the edge area hardly need to be the forwarder and it causes the unbalanced energy consumption among different region. In order to take full advantage of edge nodes, communication distance adjustment is introduced to further balance the energy consumption. Namely, the communication distance of nodes is adjusted according to the distance to the mobile sink. Nodes far away from the mobile sink use longer communication distance for fully energy utilizing and nodes close to the mobile sink use shorter transmission distance to reduce energy consumption.
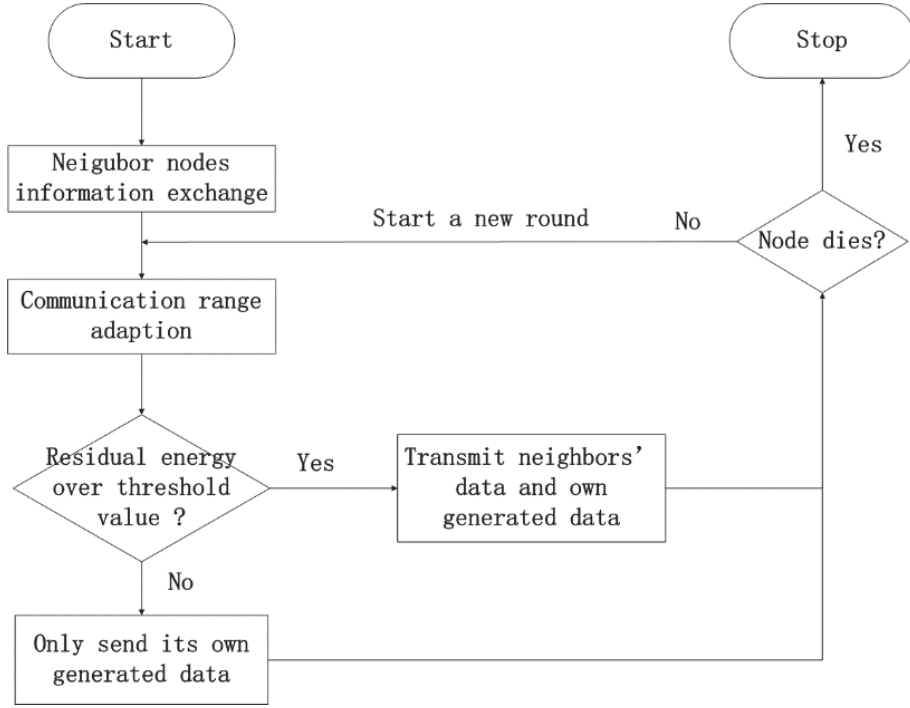
FIGURE 3: Workflow of nodes.
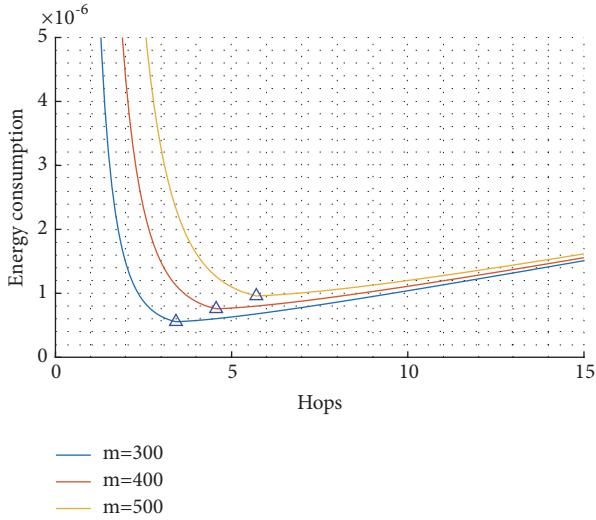
FIGURE 4: Energy consumption of different hop counts.

FIGURE 5: Network topology.

The communication distance of node i can be calculated according to formula (9):

$$d_{adjust\_optimal} = d_{optimal}$$
$$+ \left(d\left(n_{source}, \text{mobile\_sink}\right) - \frac{R}{2} \right) \cdot \alpha \tag{9}$$

where R is the radius of the network and α (as is discussed in next section) is the adjusting parameter. The chain structure after communication distance adjustment is shown as Figure 8.

4.7. Mobile Sink Moving Schema. In our proposed algorithm, the mobile sink moves with a predetermined direction and speed. The mobile sink moves around the centre of the circle with a constant radius (as is discussed in next section). Therefore, the mobile sink only needs to broadcast its position in initial phase. After $\triangle t$ time interval, the mobile sink moves to a new position as is shown in Figure 9.

5. Performance Evaluation

5.1. Simulation Environment. In order to evaluate the performance of our proposed EPEGASIS algorithm, we use Matlab
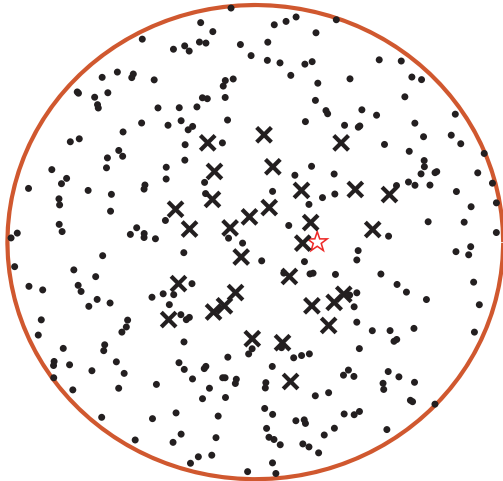
FIGURE 6: Network topology.
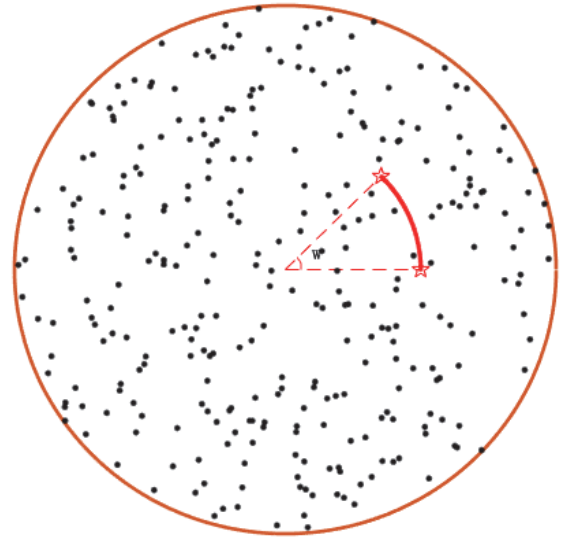
FIGURE 9: Moving schema of mobile sink.
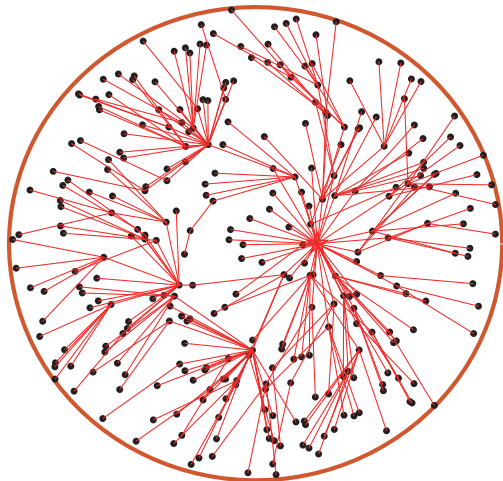
FIGURE 7: Ring routing.

FIGURE 10: Network lifetime comparison under different routing protocols.

FIGURE 8: Chain structure after communication distance adjustment.

simulator to conduct the experiment. We will compare our proposed algorithm with typical PEGASIS. And the simulation parameters are listed in Table 2.

5.2. Network Lifetime Analysis. As is mentioned above, PEGASIS is a classic chain-based routing protocol and we will compare our algorithm with it. The protocol we proposed without communication range adaption is called EPEGASIS1, and the protocol with further improvement which introduces communication distance adjustment is called EPEGASIS2. The simulation results are shown in Figure 10.
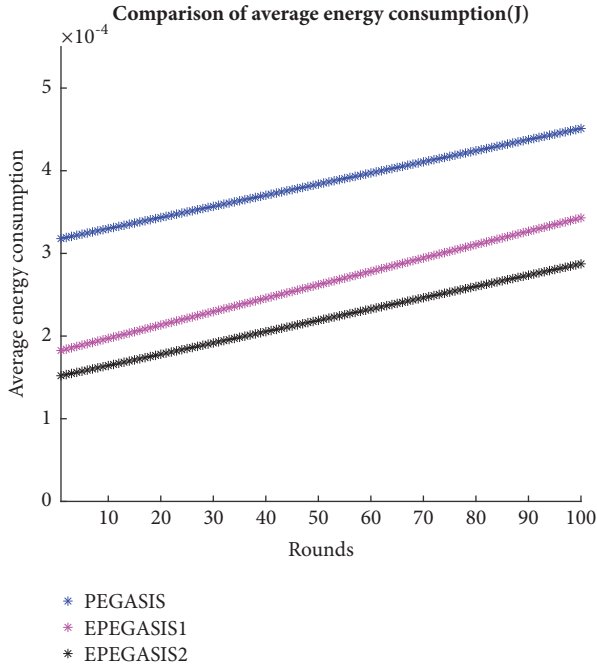
Figure 11: Average energy consumption comparison under different routing protocols.

Table 2: Simulation parameters.

Parameter Name	Parameter Value
Network radius(R)	300m
Mobile sink radius(r)	[0, 75, 150, 225, 300]m
Mobile sink speed(w)	pi/5
Number of nodes (N)	500
Packet length (l)	500 bits
Initial energy (E_0)	0.05 J
Energy consumption on circuit (E_{elec})	50nJ/bit
Free-space channel parameter (ε_{fs})	10pJ/bit/m2
Multi-path channel parameter (ε_{mp})	0.0013pJ/bit/m4
Adjusting parameter (α)	[0.1, 0.2, 0.3, 0.4]

Figure 10 describes that the lifetime of the network is greatly improved in EPEGASIS1 compared to PEGASIS. Due to the heavy burden to transmit the data packages of neighbours, nodes close to the sink begin to die at about 60 rounds. However, the first dies are at about 90 rounds in EPEGASIS2 because of the communication distance adjustment, nodes protection, and sink mobility.

5.3. Energy Consumption Analysis. We compare the average energy consumption per round of different routing protocols, and the result is shown as Figure 11. We can obviously see that the average energy consumption per round in EPEGASIS1 reduces about one-third that of PEGASIS due to the using of optimal communication distance.
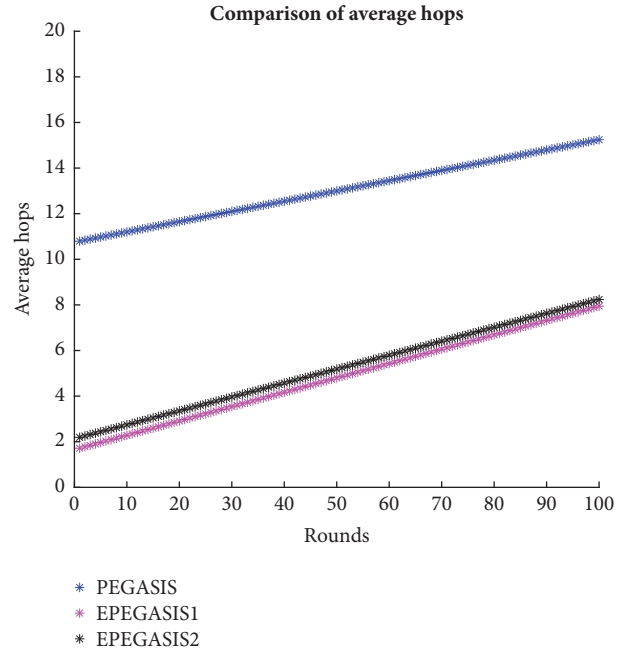
Figure 12: Average hops under different routing protocols.

5.4. Study on Network Latency. We also study the network latency between different routing protocols, and the result is shown as Figure 12. We can clearly see that the network latency has a great improvement in both EPEGASIS1 and EPEGASIS2. In EPEGASIS2, due to the adoption of communication distance adjustment, some nodes extend their communication distance and some nodes short their communication distance. Therefore, EPEGASIS1 and EPEGASIS2 almost have the same performance in terms of network latency.

5.5. Study on Communication Range Adaption. As we discussed above, each sensor node adjusts its communication distance in accordance with the distance to the mobile sink, and the adjusting parameter α determines the detailed changes. From Figure 13 we can obviously see that when α=0.2, the network has a better performance in terms of network lifetime.

5.6. Study on Mobile Sink Moving Trajectory. The moving trajectory of the mobile sink has a significant influence on the lifetime of the network. In our proposed PEGASIS2, the mobile sink moves around the centre of the circle, and we change the radius of the mobile sink to enhance the performance of the network. Simulation result is shown as Figure 14, and it demonstrates that when the mobile sink travels around the sensor filed with 0.25R radius, the network has a better performance in terms of network lifetime.

6. Conclusions

In this paper, we proposed an Enhanced PEGASIS algorithm with mobile sink support to improve the performance of
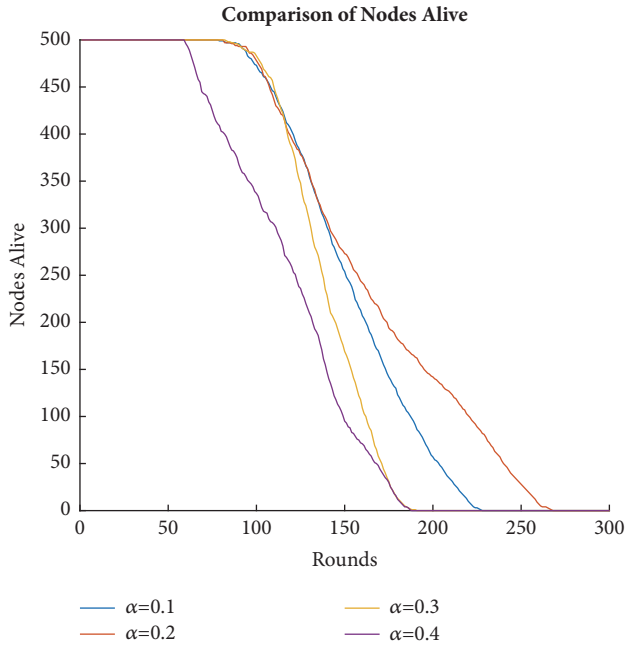
Figure 13: Network lifetime under different adjusting parameter.
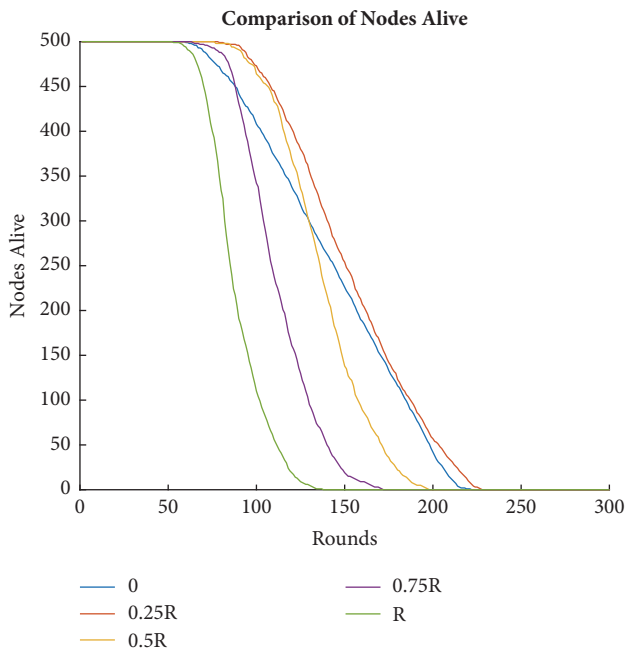
Figure 14: Network lifetime under different mobile sink moving trajectory.

WSNs. Nodes use optimal communication distance to select the forwarder and threshold value is set to protect nodes from dying too early. In order to further improve the performance of the network, we adjust the communication distance of nodes based on their distance to the mobile sink. Therefore, nodes close to the mobile sink will have short communication distance to decrease its energy consumption and the time of first node dies prolongs greatly. Extensive simulation results validate the performance of our proposed algorithms than traditional PEGASIS. However, the hot spots problem is not completely eliminated. Our future work mainly includes proper design of sink moving trajectory, as well as joint optimization of both routing algorithm and sink trajectory design.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

This paper is a revised and expanded version of a paper entitled "An Enhanced PEGASIS Algorithm with Mobile Sink Support for Wireless Sensor Networks" presented at ICCCS 2018 China, June 08-10.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in *Proceedings of the 20th IEEE International Symposium on Intelligent Control and the 13th Mediterranean Conference on Control and Automation (ISIC '05 and MED '05)*, pp. 719–724, Limassol, Cyprus, June 2005.

[2] T. Bokareva, W. Hu, and S. Kanhere, *Wireless Sensor Networks for Battlefield Surveillance*, 2006.

[3] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2688–2710, 2010.

[4] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications Magazine*, vol. 11, no. 6, pp. 6–28, 2004.

[5] N. A. Pantazis and D. D. Vergados, "A survey on power control issues in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 4, pp. 86–107, 2007.

[6] C. Wang, S. Guo, and Y. Yang, "An optimization framework for mobile data collection in energy-harvesting wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 2969–2986, 2016.

[7] C.-F. Cheng and C.-F. Yu, "Data Gathering in Wireless Sensor Networks: A Combine-TSP-Reduce Approach," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2309–2324, 2016.

[8] J.-Y. Chang and T.-H. Shen, "An efficient tree-based power saving scheme for wireless sensor networks with mobile sink," *IEEE Sensors Journal*, vol. 16, no. 20, pp. 7545–7557, 2016.

[9] J. Wang, L. Zuo, J. Shen, B. Li, and S. Lee, "Multiple mobile sink-based routing algorithm for data dissemination in wireless sensor networks," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 10, pp. 2656–2667, 2015.

[10] S. Sasirekha and S. Swamynathan, "Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network," *Journal of Communications and Networks*, vol. 19, no. 4, pp. 392–401, 2017.

[11] V. Soni and D. K. Mallick, "Fuzzy logic based multihop topology control routing protocol in wireless sensor networks," *Microsystem Technologies*, vol. 7, pp. 1–13, 2018.

[12] J. Wang, J. Cao, R. S. Sherratt, and J. H. Park, "An improved ant colony optimization-based approach with mobile sink for wireless sensor networks," *The Journal of Supercomputing*, vol. 1, no. 8, pp. 1–13, 2017.

[13] C. Wu, E. Zapevalova, Y. Chen, and F. Li, "Time Optimization of Multiple Knowledge Transfers in the Big Data Environment," *Computers, Materials & Continua*, vol. 54, no. 3, pp. 269–285, 2018.

[14] J. Wang, C. Ju, H.-J. Kim, R. S. Sherratt, and S. Lee, "A mobile assisted coverage hole patching scheme based on particle swarm optimization for WSNs," *Cluster Computing*, pp. 1–9, 2017.

[15] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *IEEE Computer Society*, vol. 18, p. 8020, 2000.

[16] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference*, vol. 3, 2003.

[17] G. Xie and F. Pan, "Cluster-Based Routing for the Mobile Sink in Wireless Sensor Networks with Obstacles," *IEEE Access*, vol. 4, pp. 2019–2028, 2016.

[18] C. Zhu, S. Wu, G. Han, L. Shu, and H. Wu, "A tree-cluster-based data-gathering algorithm for industrial WSNs with a mobile sink," *IEEE Access*, vol. 3, no. 4, pp. 381–396, 2015.

[19] F. Tashtarian, M. H. Yaghmaee Moghaddam, K. Sohraby, and S. Effati, "On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3177–3189, 2015.

[20] M. T. Nuruzzaman and H.-W. Ferng, "A low energy consumption routing protocol for mobile sensor networks with a path-constrained mobile sink," in *Proceedings of the 2016 IEEE International Conference on Communications, ICC 2016*, pp. 1–6, Malaysia, May 2016.

[21] J. Wang, Y. Cao, B. Li, H.-J. Kim, and S. Lee, "Particle swarm optimization based clustering algorithm with mobile sink for WSNs," *Future Generation Computer Systems*, vol. 76, pp. 452–457, 2017.

[22] J. Wang, J. Cao, B. Li, S. Lee, and R. S. Sherratt, "Bio-inspired ant colony optimization based clustering algorithm with mobile sinks for applications in consumer home automation networks," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 4, pp. 438–444, 2015.

[23] J. Wang, Y.-Q. Cao, B. Li, S.-Y. Lee, and J.-U. Kim, "A glowworm swarm optimization based clustering algorithm with mobile sink support for wireless sensor networks," *Journal of Internet Technology*, vol. 16, no. 5, pp. 825–832, 2015.

[24] W. Wang, H. Shi, D. Wu et al., "VD-PSO: An efficient mobile sink routing algorithm in wireless sensor networks," *Peer-to-Peer Networking and Applications*, vol. 10, no. 3, pp. 1–10, 2016.

[25] M. Azharuddin and P. K. Jana, "PSO-based approach for energy-efficient and energy-balanced routing and clustering in wireless sensor networks," *Soft Computing*, vol. 21, no. 22, pp. 6825–6839, 2017.

[26] S. Arjunan and P. Sujatha, "Lifetime maximization of wireless sensor network using fuzzy based unequal clustering and ACO based routing hybrid protocol," *Applied Intelligence*, pp. 1–18, 2017.

[27] W. Liu, X. Luo, Y. Liu, J. Liu, M. Liu, and Q. Yun, "Localization Algorithm of Indoor Wi-Fi Access Points Based on Signal Strength Relative Relationship and Region Division," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 71–93, 2018.

[28] J. Wang, J. Cao, S. Ji, and J. H. Park, "Energy-efficient cluster-based dynamic routes adjustment approach for wireless sensor networks with mobile sinks," *The Journal of Supercomputing*, vol. 73, no. 7, pp. 3277–3290, 2017.

[29] R. Meng, S. G. Rice, J. Wang, and X. Sun, "A Fusion Steganographic Algorithm Based on Faster R-CNN," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1–16, 2018.

Research Article

Social-Aware Cooperative Video Distribution via SVC Streaming Multicast

Lindong Zhao [iD],[1,2] **Lei Wang** [iD],[1,2] **Xuguang Zhang** [iD],[1,2] **and Bin Kang**[3]

[1]*National Engineering Research Center for Communications and Network Technology,*
Nanjing University of Posts and Telecommunications, China
[2]*National Mobile Communications Research Laboratory, Southeast University, China*
[3]*School of Internet of Things, Nanjing University of Posts and Telecommunications, China*

Correspondence should be addressed to Lei Wang; wanglei@njupt.edu.cn

Scalable Video Coding (SVC) streaming multicast is considered as a promising solution to cope with video traffic overload and multicast channel differences. To solve the challenge of delivering high-definition SVC streaming over burst-loss prone channels, we propose a social-aware cooperative SVC streaming multicast scheme. The proposed scheme is the first attempt to enable D2D cooperation for SVC streaming multicast to conquer the burst-loss, and one salient feature of it is that it takes fully into account the hierarchical encoding structure of SVC in scheduling cooperation. By using our scheme, users form groups to share video packets among each other to restore incomplete enhancement layers. Specifically, a cooperative group formation method is designed to stimulate effective cooperation, based on coalitional game theory; and an optimal D2D links scheduling scheme is devised to maximize the total decoded enhancement layers, based on potential game theory. Extensive simulations using real video traces corroborate that the proposed scheme leads to a significant gain on the received video quality.

1. Introduction

With the rapid development of mobile intelligent terminals, we have witnessed the explosive growth of users' demands for wireless live video services [1]. Such growth in the bandwidth demanding will bring unprecedented challenges to current wireless network in the near future [2, 3]. As communication technologies designed to cope with video traffic overload, video multicast can exploit the broadcasting nature of wireless communication to transmit video to multiple users simultaneously [4]. Due to its effectiveness, video multicast has been applied successfully in many scenarios, e.g., breaking news video, live sport videos, mobile TV programs, and so on.

In mobile video multicast, how to solve varied fading channels from the base station (BS) to multicast users is a key problem. Here, we give an illustrative paragraph. The video transmission rate of a wireless link is determined by the channel quality feedback. For a specific link, high transmission rate with a bad channel quality will render

high bit error rate and degrade the goodput. In a multicast scenario, BS can only select one transmission rate while the channel conditions of multicast users vary from each other. When BS multicast video to users according to the user with good channel quality, users with worse channel quality may fail to play back the video smoothly. Typically, BS will choose transmission rate according to the worst channel, and thus each multicast user's received video quality is limited by the worst one.

Since Scalable Video Coding streaming can be hierarchically divided into one base layer and multiple enhancement layers, the SVC based video multicast can flexibly adjust the quality of live video services according to channel feedback [5]. Thus, SVC is a suitable video coding strategy to adapt to the diversity of multicast channels [6]. One major challenge of employing SVC in video multicast is how to cope with the unavoidable packet loss over the burst-loss prone channels [7], which makes the traditional channel protection method such as FEC insufficient [8]. And due to the stringent playback deadline, the retransmission is also

not preferred especially in multicast scenarios [9]. Although some related works have studied how to reduce the interference and errors in SVC streaming [10–12], most of them need heavy modification for video codec or introduce other network components to the cellular system, which is actually impractical.

Since the cooperative local recovery via D2D communication can make up the packet loss without introducing any redundancy or asking the base station to retransmit the lost packets, it is then a good option for SVC streaming multicast countering the burst-loss. There are some works that studied how to utilize D2D communication or other network interfaces (like Wi-Fi) for video recovery [13]. However, previous cooperative recovery schemes did not take the hierarchical encoding structure of SVC into account, which may not lead to a significant gain to SVC-based video multicast. In other words, from the perspective of user's Quality-of-Experience (QoE) [14], simply boosting the transmission rate does not necessarily result in improvements in the user perception of the video quality [15].

SVC streaming has a hierarchical encoding structure, which means that each enhancement layer can be decoded only after the correct decoding of lower layers. In such case, either bit error or packet loss would lead to the loss of synchronization information and error propagation. The hierarchical encoding structure indicates that packets belonging to different enhancement layers should have different weight for improving video quality. And the SVC based video distribution scheme should also be structured and dealing with missing packets depends on the actual reception. Existing video distribution schemes often treat all the missing packets equally; hence those distribution schemes can not give a good video quality for SVC streaming multicast. In a word, how to coordinate users' cooperation according to the hierarchical encoding structure of SVC is a critical issue for cooperative SVC streaming multicast.

In this paper, we propose a D2D-based cooperative SVC streaming multicast scheme to improve the received video quality, which is based on the consideration of the hierarchical encoding structure of SVC. More specifically, users in the same cooperative group share enhancement layers using D2D communication under the scheduling of the BS. To stimulate effective cooperation among multicast users, the cooperative groups are constructed by using a coalitional game. And to maximize the total decoded enhancement layers, an optimal D2D links scheduling scheme is devised based on potential game theory. To the best of our knowledge, this paper is the first attempt to enable D2D cooperation for SVC streaming multicast to conquer the burst-loss. The main contributions of this paper are summarized as follows.

(i) We propose a social-aware cooperative SVC streaming multicast scheme to improve received video quality by stimulating users to form cooperative groups and share enhancement layers. In particular, this scheme takes full account of the hierarchical encoding structure of SVC to coordinate cooperation.

(ii) To stimulate effective cooperation among users, we propose a cooperative group formation method based

on coalitional game theory, in which the logical relationship among users is developed according to social attributes, interest similarity, and geographical location.

(iii) To maximize the total received enhancement layers of the cooperative group, we devise a D2D links scheduling scheme for the base station. The optimization problem for D2D links scheduling is cast as a potential game and solved by local search algorithm. Extensive simulations with real video traces show that the proposed scheme can achieve up to 65.9% gain of video PSNR (Peak Signal-to-Noise Ratio) and valuable video frame ratio over the traditional scheme.

The rest of this paper is organized as follows. Section 2 gives a review of the related works in the field of wireless video multicast and cooperative local recovery. The system model and workflow of the proposed scheme are discussed in Section 3. In Section 4, we study the problems of cooperative group formation and the optimal links scheduling in detail. The extensive numerical results are shown in Section 5 and conclusions are drawn in Section 6.

2. Related Works

2.1. Wireless Video Multicast. Wireless video multicast has recently become a hot research topic, which is promising to help improve spectrum efficiency. R. Chandra et al. proposed a multicast system DirCast based on Wi-Fi, targeting to minimize the total transmission delay [16]. In order to improve the performance of video multicast, a gateway was introduced between the base station and the cloud servers to determine the order of frames [17], and S. Aditya et al. integrated a joint source-channel coding into the MPEG-4 codec [18]. Deb et al. proposed a greedy algorithm for resource allocation in SVC streaming multicast scenario via WiMAX [11]. S. Jakubczak et al. proposed a cross-layer design for multicasting scalable video, with the goal of improving robustness to wireless interference and errors [12].

Some of the above works with nonscalable encoding suffer from "the limit of the worst one" caused by multicast channel differences, as mentioned earlier. To counter wireless interference and errors, the others with scalable encoding introduce heavy redundancy to video codec or add new network interface to the system, but present insufficient for burst-loss. Different from these works, this paper effectively conquers the burst-loss for SVC streaming multicast via D2D cooperation and is compatible with existing cellular networks.

2.2. Cooperative Local Recovery. There have been some works on the cooperative local recovery, i.e., how to enable users to share missing packets or act as relay nodes. Reference [19] proposed a cooperative peer-to-peer recovery scheme optimally scheduling the priority of packets. Reference [9] recovered the video frames by constructing D2D assisted multipath transmission. Reference [20] realized peer-to-peer
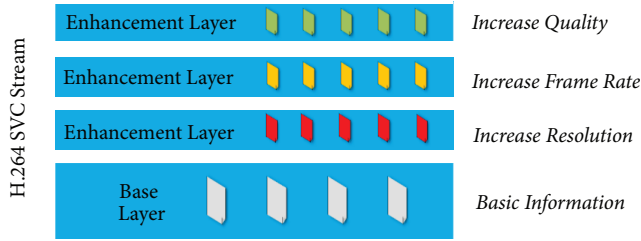
FIGURE 1: Scalable Video Coding structure.

repair by combining batch processing and network coding. Reference [21] deployed D2D communication inside the multicast group, and users who successfully decoded the message retransmit it to those in need. Reference [13] stimulated users' cooperation for video multicast by utilizing two types of social ties, social reciprocity and social trust. The above video distribution schemes often treat all the missing packets equally; hence they cannot give a good video quality for SVC streaming multicast. Different from them, we coordinate users' cooperation based on the unique hierarchical encoding structure of SVC.

3. System Model

3.1. Preliminaries. To facilitate our exposition, it is necessary to first introduce the scalable video encoding structure and live video playback mechanism.

The core idea of SVC is to divide the video streaming into multiple layers according to different requirements. As illustrated in Figure 1, there are one base layer and multiple enhancement layers, each of which is transmitted separately. Devices can reconstruct the basic quality video image as long as it receives the base layer, but the video reconstructed from the base layer is of poor quality. The enhancement layers contain details of the video streaming, which means users can get a better video quality by receiving more enhancement layers within the range of the total video bit rate. In addition, an enhancement layer can be decoded after the correct decoding of the base layer and lower enhancement layers, while the base layer can be decoded separately.

In general, SVC streaming is composed of Group-of-Picture (GoP) sequences, and a GoP consists of a base layer and one or more enhancement layers with RLC encoding. We denote these layers by $\mathscr{L} = \{1, 2, \cdots, L\}$, each encoded layer $l \in \mathscr{L}$ is processed into P^l packets, and P^l is also the decoding threshold for layer l. The condition for successfully receiving the layer l can be expressed as

$$R_i^l \geqslant P^l, \quad l \in \mathscr{L}. \tag{1}$$

However, receiving the full information of the layer l does not mean that layer l can be reconstructed; the condition for successfully decoding the layer l is given by

$$R_i^{l'} \geqslant P^{l'}, \quad \forall l' \leqslant l,\ l \in \mathscr{L},\ \forall l' \in \mathscr{L}, \tag{2}$$

where $R_i^{l'}$ denotes the packets which user i received for layer l'.

Equation (2) means that, to restore video quality to layer l, not only should the received packets of layer l be more than its threshold, but also the lower layers are decoded correctly.

The live video playback mechanism is illustrated in Figure 2. Unlike data transmissions such as browsing a Web page or downloading a file, live video multicast is a delay sensitive application with real-time requirements. Generally, the smooth video playback is tolerable with a maximum delay according to Persistence of Vision. That is, each GoP has a fixed transmission deadline. Once a packet in the GoP misses this deadline, it will be dropped. For SVC streaming, packets on each layer are transmitted independently. Users reconstruct each video layer with the received packets and the layer which does not receive enough packets before the deadline cannot be reconstructed.

3.2. System Model and Workflow. From the previous discussion, we can see that affected by hierarchical encoding structure of SVC and live video delay constraint, users may not be able to reconstruct a certain layer due to their failure to receive the lower layer correctly. Referring to SVC streaming, once a user receives enough packets for the corresponding layer, it can be reconstructed, regardless of whether the packets are coming from the base station or other users. Due to the diversity of multicast channels, when some users lose a certain layer, other users may receive the layer successfully. First, we propose to enable users to form "cooperative groups" taking into account social attributes, preference similarity, and geographical location, according to the strategy proposed in Section 4. Then, we develop collaboration among users via D2D communication inside each group and missing packets are shared among users under the base station's scheduling to rebuild lost layers.

As shown in Figure 3, one base station and N users are in a cell, and partial users are divided into different multicast groups receiving different videos via the cellular downlink, called multicast users. Others who have no intention of participating in multicast are referred to as ordinary users. Since the ordinary user is not covered by our discussion, we may assume that the N users are all multicast users. And all users are equipped with both D2D and cellular interfaces, so that they receive packets by D2D communication if necessary.

The workflow of our proposed intragroup recovery is shown in Figure 4; we divide one GoP delivery process into two parts: multicast round and cooperation round. At the beginning of one GoP transmission, each user receives packets from the base station during multicast phase. In process of following cooperation round, users would report to the base station which layers are missing. The D2D link will be established by the unified scheduling of the base station to realize the complementation of the data received among users. If the time slots allocated to cooperation round are sufficient, we can encapsulate the above reporting, scheduling, and sharing processes into a subround, allowing users to switch multiple times to obtain their missing layers.

Application Layer	GoP i-1 Playback	GoP i Playback	GoP i+1 Playback	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
Transport Layer	GoP i Transmission	GoP i+1 Transmission	GoP i+2 Transmission	$\cdots$
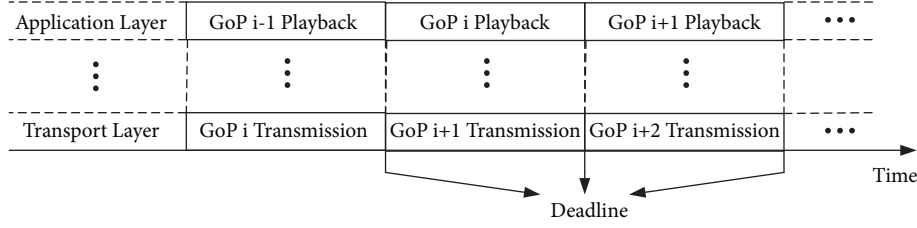
FIGURE 2: Live video transmission and playback.
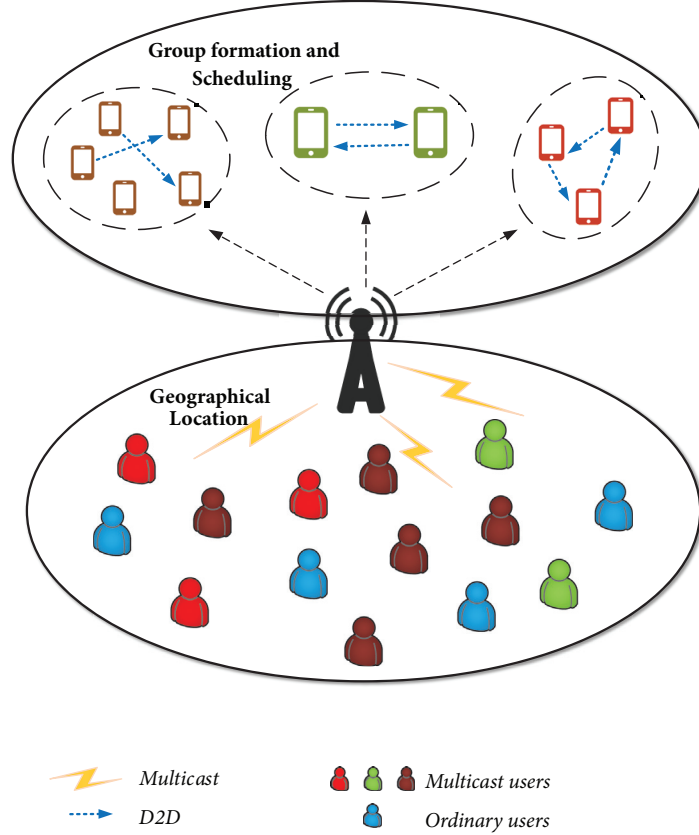
FIGURE 3: An illustration of cooperative SVC streaming multicast system.

4. Cooperative Group Formulation and Optimal Links Scheduling

There are two technical challenges for our proposed scheme: (1) *how to effectively screen users into multicast group to stimulate cooperation* and (2) *how to schedule D2D links to maximize the overall reconstructed SVC layers*. Now we will look at these two subproblems.

4.1. Cooperative Multicast Group Formation. Now we formulate the cooperative multicast group formation for SVC streaming recovery as a coalitional game.

4.1.1. Utility Function. When a user is ready to invite another user to join the multicast group, what are the main considerations?

First, consistency of user preferences within group is a prerequisite for multicast group formation, in order to improve the efficiency of spectrum resources in cellular system. We use collaborative filtering method to obtain the similarity of users' interest. Assume that the number of multicast video types available is X, $K(i) = \{a_1^i, a_2^i, \cdots, a_X^i\}$ denoting user i's interest set, obtained from historical data analysis, where a_x^i is a binary variable, which is equal to 1 if user i is interested in video type x; otherwise $a_x^i = 0$. The interest similarity between users i and j can be obtained by the Jaccard coefficient [22], expressed by

$$S(i, j) = \frac{|K(i) \cap K(j)|}{|K(i) \cup K(j)|}, \tag{3}$$

where $0 \le S(i, j) \le 1$, the larger the value is, the more likely the users are to choose the same video content.
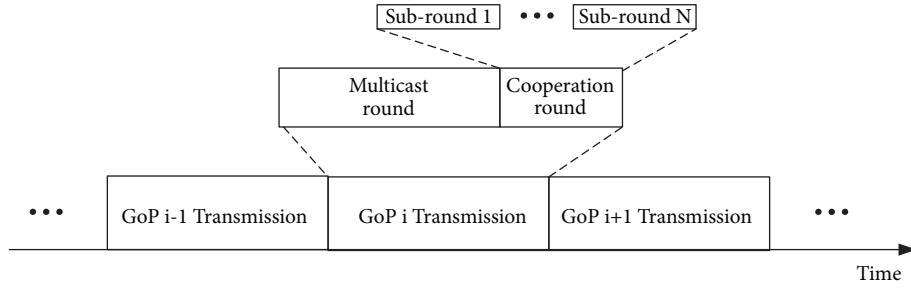
FIGURE 4: Intragroup recovery workflow.

Since social attributes represent long-term contact between users, links maintained with them tend to be stable [23]. With the help of social-related parameters to filter users, we can maximize the use of temporary links to achieve QoE promotion. We use $Q(i, j)$ to denote the social affinity of users i and j, which can be obtained by analyzing their interactions with social software. Note that $0 \leq Q(i, j) \leq 1$, the larger the value, the closer the relationship between users. Besides, since the next collaboration will take place with D2D communication, users have to consider location information when inviting other members, lest they become isolated nodes. We characterize this property in terms of the ratio of distance between users to diameter of the cell, expressed by $G(i, j) = d_{i,j}/D$, where D denotes the diameter and $d_{i,j}$ denotes the distance between users i and j, $0 < G(i, j) \leq 1$.

After the above discussion, we can see that there are three main factors that affect the preference of user invitation, and their relationship to utility functions can be expressed as

$$\frac{\partial U}{\partial S} \geq 0,$$
$$\frac{\partial U}{\partial Q} \geq 0, \qquad (4)$$
$$\frac{\partial U}{\partial G} \leq 0$$

And we will use the method of microeconomics to design utility function to quantify this preference. In microeconomics, utility is the quantitative expression of preference, and utility is divided into ordinal utility and cardinality utility [24]. Cardinal utility is measured by absolute value, while ordinal utility is concerned with the relative relationship of utility. Here, the users are concerned only with the relative relationship of the utility brought about by inviting different users; hence we use ordinal utility to quantify preferences. Because there is no alternative or complementary relationship between interest similarity, social attributes, and geographical location, we use Cobb-Douglas preference model to design our utility function, expressed as

$$U(i, j) = \frac{S(i, j)^{\alpha} Q(i, j)^{\beta}}{G(i, j)^{\gamma}}, \qquad (5)$$

where α, β, and γ are the corresponding coefficients to adjust the priorities and all positive. Note that the value of ordinal utility itself has no physical meaning; we just use it to quantify preference; the forms of utility function are not unique. Besides, without considering the preference similarity, the proposed scheme can also be used for users in the same MBMS service area [25].

4.1.2. Coalitional Game Formulation. Then, we cast the multicast group formation as a coalitional game $\Omega = \{\mathbb{N}, \Phi_{\mathbb{N}}, W, (\succ_i)_{i \in \mathbb{N}}\}$ as follows:

(i) The set of players $\mathbb{N}$ is the set of multicast users.

(ii) The set of cooperation strategies is $\Phi_{\mathbb{N}} = \{(f_i)_{i \in \mathbb{N}} : f_i \in \mathbb{N}^t, \forall i \in \mathbb{N}\}$, which denotes the set of current investible users for all users.

(iii) The characteristic function is $W(\Delta) = \{(f_i)_{i \in \mathbb{N}} \in \Phi_{\mathbb{N}} : (f_i)_{i \in \Delta} = (j)_{j \in \Delta} \text{ and } f_k = k, k \in \mathbb{N} \setminus \Delta\}$ for each multicast group $\Delta \in \mathbb{N}$. There are only two cases for the players, and the condition $(f_i)_{i \in \Delta} = (j)_{j \in \Delta}$ denotes that user i chooses to invite user j to join a multicast group Δ (a coalition). The condition $f_k = k, k \in \mathbb{N} \setminus \Delta$ represents that user k out of the coalition Δ is not in any group.

(iv) The preference order $(\succ_i)_{i \in \mathbb{N}}$ is defined as $(f_k)_{k \in \mathbb{N}} \succ_i (f'_k)_{k \in \mathbb{N}}$ if and only if $U(i, f_i) \geq U(i, f'_i)$. In detail, user i prefers the group formation strategy $(f_k)_{k \in \mathbb{N}}$ to $(f'_k)_{k \in \mathbb{N}}$ if and only if user i invites assigned user f_i in strategy $(f_k)_{k \in \mathbb{N}}$ to have more utility than inviting assigned user f'_i in strategy $(f'_k)_{k \in \mathbb{N}}$ according to above-mentioned utility function.

Definition 1. The core is the set of $(f_i^*)_{i \in \mathbb{N}} \in W(\mathbb{N})$ for which a coalition Δ does not exist and $(f_i)_{i \in \mathbb{N}} \in W(\Delta)$ such that $(f_i)_{i \in \mathbb{N}} \succ_i (f_i^*)_{i \in \mathbb{N}}$.

That is, no set of participants will actively take a joint action that is better for all of them to deviate from current coalition, which is in the same spirit as Nash equilibrium in a noncooperative game. By obtaining a core solution of the coalition game, the formation of multicast groups is completed.

4.1.3. Coalitional Game Core Solution. Now, we study the existence of the core solution.

Definition 2. A game $\mho = \{\mathbb{Z}, \Phi_{\mathbb{Z}}, W, (\succ_i)_{i \in \mathbb{Z}}\}$ is a subgame of the game Ω if and only if $\mathbb{Z} \subseteq \mathbb{N}$ and $\mathbb{Z} \neq \emptyset$.

Definition 3. A nonempty subset $\Delta \in \mathbb{Z}$ is a top coalition for a subgame $\mho = \{\mathbb{Z}, \Phi_{\mathbb{Z}}, W, (\succ_i)_{i \in \mathbb{Z}}\}$ if and only if, for strategy set $(f_z^*)_{z \in \mathbb{Z}} \in W(\Delta)$ and any strategy $(u_z)_{z \in \Delta} \in W(\mathbb{C})_{\mathbb{C} \subseteq \mathbb{Z}}$, we have $U(z, f_z^*) \geqslant U(z, u_z)$ for all $z \in \mathbb{Z}$.

In fact, a subgame is a subset of the players of the original game. Note that the strategy set to build a top coalition is the core of the subgame made up of members in the top coalition, since the group is collectively best for all its members.

Definition 4. The game $\Omega = \{\mathbb{N}, \Phi_{\mathbb{N}}, W, (\succ_i)_{i \in \mathbb{N}}\}$ owns only one core solution if and only if all its subgames have a top coalition and any player in any subgame cannot belong to more than one top coalition.

That is, the top coalitions set up by adopting the core strategy are the multicast groups we need.

Definition 5. For a coalitional subgame $\mho = \{\mathbb{Z}, \Phi_{\mathbb{Z}}, W, (\succ_i)_{i \in \mathbb{Z}}\}$, we define a user sequence $(n_1, \cdots, n_C)$ as a multicast friend cycle if and only if $U(z_c, z_{c+1}) \geqslant U(z_c, z_k) \; \forall z_k \in \mathbb{Z}$ for all $z = 1, \cdots, C - 1$ and $U(z_C, z_1) \geqslant U(z_C, z_k) \; \forall z_k \in \mathbb{Z}$.

Lemma 6. *For any subgame $\mho = \{\mathbb{Z}, \Phi_{\mathbb{Z}}, W, (\succ_i)_{i \in \mathbb{Z}}\}$, we can obtain at least one multicast friends cycle. Any multicast friend cycle is a top coalition of corresponding subgame $\mho$.*

Proof. At the initial stage of any subgame, we randomly choose user z_1 from the players subset $\Phi_{\mathbb{Z}}$, then we can find the second user z_2 who can maximize user z_1's utility. Repeat the steps above and we can find a multicast friend cycle. Using proofs by contradiction, if such cycle does not exist, the above-mentioned user sequence will extend infinitely. Since any two users are different and the total number of users in the cell is limited, this assumption is clearly not true. And according to Definition 5, each user takes the decisions that are best for themselves; no participant can take a joint action that is better for all of them to deviate from current coalition. Any multicast friend cycle generated through the above steps is a top coalition. □

Theorem 7. *For game $\Omega = \{\mathbb{N}, \Phi_{\mathbb{N}}, W, (\succ_i)_{i \in \mathbb{N}}\}$, the strategy set for generating multicast friend groups is a core solution, and the set of coalitions resulting from a decision initiated by any one of the multicast users is unique if no strategies for one player can bring about the same utility.*

Proof. According to Lemma 6, at the beginning of game Ω, we randomly choose a user to invite other users and obtain a top coalition C_1. By Definition 3, players in this top coalition have achieved the maximum utility currently available; these players will not continue to participate in current game. After removing the players mentioned above, the original game becomes its corresponding subgame $\mho = \{\mathbb{Z}, \Phi_{\mathbb{Z}}, W, (\succ_i)_{i \in \mathbb{Z}}\}$, where $\mathbb{Z} = \mathbb{N} \backslash C_1$. With the generation of one multicast friend cycle after another using Lemma 6, we have $\mathbb{Z} = \mathbb{N} \backslash C_t$ for corresponding subgame to each iteration $t = 1, \cdots, T$. According to Definition 4, this completes the proof. For the second part of the theorem, using proofs by contradiction, if one player can join two different top coalitions, it must have more than one optimal strategy, which contradicts Definition 5.

Based on Lemma 6 and Theorem 7, we can find the core solution for cooperative group formation. □

4.2. User Cooperation Scheduling. As we mentioned earlier, with a view to the hierarchical encoding structure of SVC streaming, the same layer of a GoP has different values according to the actual reception. For example, layer 4 enables better video quality for user who owns layers $\{1, 2, 3\}$, but it does not work for user who only has layers $\{1, 2\}$. The proposed scheme targets to improve video quality of the multicast group as much as possible in a subphase, which is directly related to layers that can be reconstructed. Considering the upper limit to the perceived video quality of users and marginal utility on the improvement of QoE [26, 27], the optimization problem can be formulated as

$$\max \sum_{i=1}^{N} \frac{1}{1 + \exp\left(-\omega l_i\right)} \tag{6}$$

where ω is the marginal utility coefficient and l_i is the number of layers that user i has successfully reconstructed from this collaboration. Now we analyze this problem with an exact potential game.

Definition 8. A game $\Gamma = \{\mathbb{C}, (Y^i)_{i \in \mathbb{C}}, (U_i)_{i \in \mathbb{C}}\}$ is an exact potential game, if there exists a function $H : Y \longrightarrow R$, such that, $\forall i \in \mathbb{C}, \forall y^{-i} \in Y^i, \forall x, z \in Y^i$, we have $U_i(x, y^{-i}) - U_i(z, y^{-i}) = H(x, y^{-i}) - H(z, y^{-i})$. $\mathbb{C}, H, Y^i, U_i$ are, respectively, player set, potential function, player i's strategy set, and utility function.

A potential game must have equilibrium with pure strategy and a discrete potential game has finite improvement properties [28]. If the objective function is a potential function of a game with discrete strategy space, it must have an optimal solution and can be solved by local search. Now we formulate a potential game Γ as follows: the set of players $\mathbb{C}$ is the set of users in a multicast group, player i's available strategy Y^i is to choose which player in the group to request video layer, utility function is $F_i = 1/(1 + \exp(-\omega l_i))$, and potential function is $H = \sum_{i=1}^{N} 1/(1 + \exp(-\omega l_i)) = \sum_{i=1}^{N} F_i$.

According to [28], we can prove the proposed objective function is a potential function of game Γ. Since the number of members in multicast group is finite, game Γ's strategy space is discrete. The problem (6) can be solved by local search algorithm such as Tabu Search and Simulated Annealing.

5. Performance Analysis

In this section, we evaluate the performance of the proposed cooperative SVC streaming multicast scheme through simulations by using MATLAB.

Without loss of generality, we suppose that the wireless multicast channel conditions are random but quasi-static and remain unchanged from one GoP that is sent out until the deadline to be dropped. Since channel capacity described by

the Shannon equation does not reflect the wireless channel conditions preciously, where the wireless channels are burst-loss prone at upper layer, we use the Gilbert-Elliot channel model to capture burst-loss features in the proposed system without handling the details of the datalink layer and physical layer. Gilbert-Elliot model can be denoted by a two-state stationary continuous time Markov chain; there are only two cases per link: Good or Bad. If the state is Bad, one frame will be lost; otherwise, it can be successfully decoded.

For video streaming, we consider H.264/SVC video codec with YUV420; i.e., there are 8 layers per GoP and each layer consists of 16 encoded video frames. We implement the experiments with the real video traces Foreman, where the average PSNR (Peak Signal-to-Noise) of the Foreman is 43.32 dB, the resolution of the Foreman is 352×288, and default video encoding bitrate is 450Kbps.

Besides, in order to measure the performance of the proposed scheme, we consider the following three measurement metrics:

Peak Signal-to-Noise: PSNR (Peak Signal-to-Noise) is an objective evaluation of video quality and is a function of the mean square between the original and the received video sequences, which can be formulated as

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \qquad (7)$$

where MAX_I is the maximum value of frame pixel and MSE is the mean square error between the original and the received video sequences.

Valuable Video Frame Ratio: the valuable video frame ratio is defined as the ratio between the number of valuable video frames received by a user and the number of all transmitted frames before playback.

Mean Opinion Score: MOS (Mean Opinion Score) is a subjective measurement of user's view on the video quality. When the PSNR range is about >37, 31~37, 25~31, 20~25, <20, user's subjective experience quality is "Excellent", "Good", "Fair", "Poor", and "Bad", respectively [13].

The performance of the proposed scheme is compared with direct multicast in our simulations. According to multicast channel conditions, users with direct SVC multicast discard the enhancement layers that could not be decoded due to burst-loss. N users are scattered into a round cell, and the interest similarity and social affinity between each other are also randomly obtained, and coefficients α, β, γ, ω are equal to 1. The base station determines the average bit rate that can transmit all layers of one GoP in multicast round, and D2D cooperation round is set to transmit an enhancement layer.

First, we evaluate the performance of our proposed cooperative group formation scheme; the experiences are implemented with different number of multicast users in the cell, $N = 100, 200, 300, 400, 500, 600, 700, 800$. To ensure the generality of the result, for each given N, we average over 1000 runs. Figure 5 shows the average size of cooperative multicast groups. We can see the size of groups has not skyrocketed with the number of users but slows down the growth rate gradually. This is because users select clusters based on the
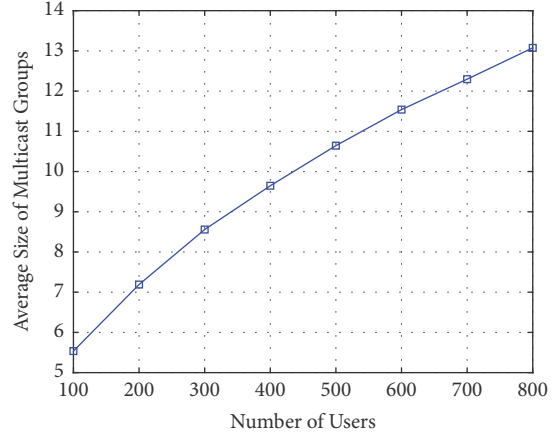
Figure 5: Average size of cooperative groups under various number of users.

coalitional game. A player can only select one player as the optimal strategy, and each player can only be selected once, thus limiting the size of cooperative groups.

Next, we employ the above evaluation metrics to compare the received video quality of the proposed scheme and traditional direct multicast. The default multicast group size is set to 10. Figure 6 shows the relationship between the received video quality in terms of PSNR and the probability of burst-loss during one GoP transmission. Obviously, the growing probability of burst-loss signifies worse channel qualities. With the deterioration of channel quality, the gain of the proposed scheme over direct multicast shows a gradual increase and then remains stable. This is because the restoration of video quality in the proposed scheme is based on the complementarity among users. With the degradation of channel conditions, the probability of complementarity between users increases gradually. But when the channel quality deteriorates to a certain extent, most users have only few of enhancement layers, and the possibility of cooperation is reduced.

In Figure 7, curves of valuable video frame ratios with the probability of burst-loss during one GoP transmission are depicted. We can observe that the proposed scheme far outperforms direct multicast, especially when the channel quality is quite poor. The valuable video frame ratio of the proposed scheme is 65.9% higher than direct multicast. This is because users in the direct SVC streaming multicast system have to discard the corresponding enhancement layers, and the system resources are severely wasted.

To better understand the performance of the proposed scheme, Figure 8 shows the experience qualities in terms of MOS under various burst-loss probability and the SVC streaming recovery with the real video traces Foreman is shown in Figure 9. As we can see, the distortion of the video sequences with the proposed scheme is slight. According to Figures 6 and 9, the average PSNR of the proposed scheme is higher than 35dB and users' MOS is "Excellent" or "Good". But the PSNR of direct multicast is between 24dB and 34dB, which signifies the MOS is "Fair". So the scheme proposed in this paper greatly improves the user's experience.
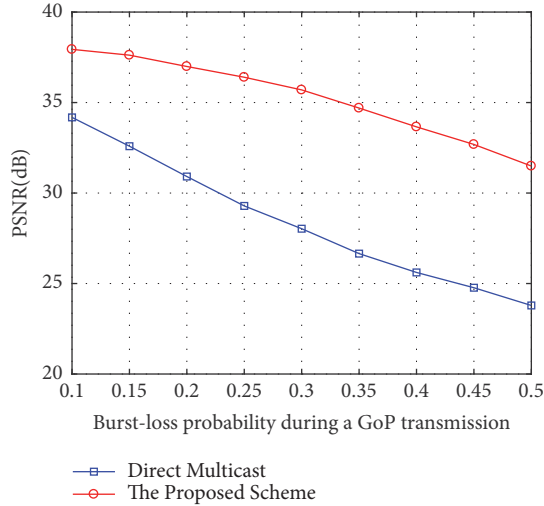
FIGURE 6: Relationship between the video quality and burst-loss probability.
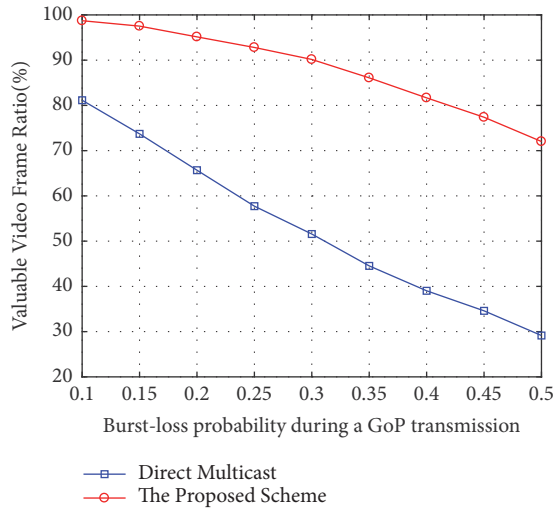
FIGURE 7: Valuable video frame ratio under various burst-loss probability.

FIGURE 8: Mean opinion score under various burst-loss probability.
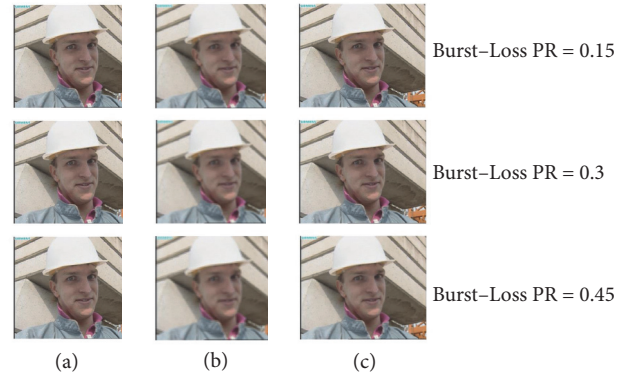
FIGURE 9: Video recovery performance under various burst-loss probability: (a) Foreman frame, (b) direct multicast, and (c) the proposed scheme.

6. Conclusions

This paper studied a cooperative SVC streaming multicast scheme based on D2D communication to improve the received video quality over burst-loss prone channels. The proposed scheme motivates multicast users to form cooperative groups based on coalitional game theory to share enhancement layers. The proposed optimal D2D links scheduling based on potential game helps to improve the received video quality of multicast users, while taking into account the hierarchical encoding structure of SVC. Moreover, the proposed scheme can make up the burst-loss without introducing heavy redundancy to video codec or add new network interface. Extensive numerical analysis studies with real video traces have corroborated the significant gain using the proposed scheme.

In future, it is of great interest to design a distributed optimal link scheduling algorithm to reduce the burden of centralized scheduling on the BS.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Rodrigo Muñoz, D. Mora, and R. Barriga, "Teachers Cisco certification and their impact in Cisco Networking Academy program," *Revista Científica de la UCSA*, vol. 4, no. 1, pp. 50–56, 2017.

[2] L. Zhou, D. Wu, J. Chen, and Z. Dong, "Greening the Smart Cities: Energy-Efficient Massive Content Delivery via D2D Communications," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1626–1634, 2017.

[3] L. Zhou, D. Wu, Z. Dong, and X. Li, "When Collaboration Hugs Intelligence: Content Delivery over Ultra-Dense Networks," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 91–95, 2017.

[4] L. Zhou, D. Wu, J. Chen, and Z. Dong, "When Computation Hugs Intelligence: Content-Aware Data Processing for Industrial IoT," *IEEE Internet of Things Journal*, 2017.

[5] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1204–1217, 2007.

[6] G. Feng, Y. Zhang, J. Lin, H. Wang, and L. Cai, "Joint optimization of downlink and D2D transmissions for SVC streaming in cooperative cellular networks," *Neurocomputing*, vol. 270, pp. 178–187, 2017.

[7] S. Tang and P. R. Alface, "Impact of random and burst packet losses on H.264 scalable video coding," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2256–2269, 2014.

[8] I.-H. Hou and P. R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *Proceedings of the IEEE INFOCOM*, pp. 1–9, San Diego, CA, USA, 2010.

[9] Z. Liu, M. Dong, H. Zhou, X. Wang, Y. Ji, and Y. Tanaka, "Device-to-device assisted video frame recovery for picocell edge users in heterogeneous networks," in *Proceedings of the ICC 2016 - 2016 IEEE International Conference on Communications*, pp. 1–6, Kuala Lumpur, Malaysia, May 2016.

[10] Y.-J. Yu, P.-C. Hsiu, and A.-C. Pang, "Energy-efficient video multicast in 4G wireless systems," *IEEE Transactions on Mobile Computing*, vol. 11, no. 10, pp. 1508–1522, 2012.

[11] S. Deb, S. Jaiswal, and K. Nagaraj, "Real-time video multicast in WiMAX networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 2252–2260, April 2008.

[12] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proceedings of the ACM 17th Annual International Conference*, pp. 289–300, Las Vegas, Nevada, USA, September 2011.

[13] Y. Cao, T. Jiang, X. Chen, and J. Zhang, "Social-aware video multicast based on device-to-device communications," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1528–1539, 2016.

[14] A. Cánovas, M. Taha, J. Lloret, and J. Tomás, "Smart resource allocation for improving QoE in IP Multimedia Subsystems," *Journal of Network and Computer Applications*, vol. 104, pp. 107–116, 2018.

[15] M. Taha, J. Lloret, A. Canovas, and L. Garcia, "Survey of Transportation of Adaptive Multimedia Streaming service in Internet," *Network Protocols and Algorithms*, vol. 9, no. 1-2, p. 85, 2018.

[16] R. Chandra, S. Karanth, T. Moscibroda et al., "DirCast: A practical and efficient Wi-Fi multicast system," in *Proceedings of the 2009 17th IEEE International Conference on Network Protocols (ICNP)*, pp. 161–170, Plainsboro, NJ, USA, October 2009.

[17] J. Yoon, H. Zhang, S. Banerjee, and S. Rangarajan, "Video multicast with joint resource allocation and adaptive modulation and coding in 4G networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 5, pp. 1531–1544, 2014.

[18] S. Aditya and S. Katti, "FlexCast: Graceful wireless video streaming," in *Proceedings of the ACM 17th Annual International Conference*, pp. 277–288, Las Vegas, Nevada, USA, September 2011.

[19] S. Raza, D. Li, C.-N. Chuah, and G. Cheung, "Cooperative peer-to-peer repair for wireless multimedia broadcast," in *Proceedings of the IEEE International Conference onMultimedia and Expo, ICME 2007*, pp. 1075–1078, China, July 2007.

[20] X. Liu, S. Raza, C. N. Chuah, and G. Cheung, "Network Coding Based Cooperative Peer-to-Peer Repair in Wireless Ad-Hoc Networks," in *Proceedings of the 2008 IEEE International Conference on Communications*, pp. 2153–2158, Beijing, 2008.

[21] B. Zhou, H. Hu, S.-Q. Huang, and H.-H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2315–2326, 2013.

[22] X.-F. Zhang and D.-Q. Dai, "A framework for incorporating functional interrelationships into protein function prediction algorithms," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 3, pp. 740–753, 2012.

[23] Y. Li, T. Wu, P. Hui, D. Jin, and S. Chen, "Social-aware D2D communications: qualitative insights and quantitative analysis," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 150–158, 2014.

[24] Z. Weiqing and L. Zhenyu, "Information Measurement Model Based on Ordinal Utility," in *Proceedings of the 2011 International Conference on Future Computer Sciences and Application*, pp. 154–157, Hong Kong, China, 2011.

[25] M. Taha, J. M. Jimenez, A. Canovas, and J. Lloret, "Intelligent Algorithm for Enhancing MPEG-DASH QoE in eMBMS," *Network Protocols and Algorithms*, vol. 9, no. 3-4, p. 94, 2018.

[26] J. P. Hansen, S. Hissam, and L. Wrage, "QoS optimization in ad hoc wireless networks through adaptive control of marginal utility," in *Proceedings of the 2013 IEEE Wireless Communications and Networking Conference, WCNC 2013*, pp. 1192–1197, Shanghai, China, 2013.

[27] L. Badia, M. Lindström, J. Zander, and M. Zorzi, "Demand and Pricing Effects on the Radio Resource Allocation of Multimedia Communication Systems," in *Proceedings of the Global Telecommunications Conference, 2003. GLOBECOM '03*, vol. 7, pp. 4116–4121, IEEE, 2003.

[28] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

Research Article

Securely Outsourcing ID3 Decision Tree in Cloud Computing

Ye Li,[1] Zoe L. Jiang [iD],[1] Xuan Wang [iD],[1] Junbin Fang,[2] En Zhang [iD],[3] and Xianmin Wang [iD][4]

[1]*Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China*
[2]*Jinan University, Guangzhou, China*
[3]*Henan Normal University, Henan, China*
[4]*School of Computer Science, Guangzhou University, China*

Correspondence should be addressed to Zoe L. Jiang; zoeljiang@hit.edu.cn

With the wide application of Internet of Things (IoT), a huge number of data are collected from IoT networks and are required to be processed, such as data mining. Although it is popular to outsource storage and computation to cloud, it may invade privacy of participants' information. Cryptography-based privacy-preserving data mining has been proposed to protect the privacy of participating parties' data for this process. However, it is still an open problem to handle with multiparticipant's ciphertext computation and analysis. And these algorithms rely on the semihonest security model which requires all parties to follow the protocol rules. In this paper, we address the challenge of outsourcing ID3 decision tree algorithm in the malicious model. Particularly, to securely store and compute private data, the two-participant symmetric homomorphic encryption supporting addition and multiplication is proposed. To keep from malicious behaviors of cloud computing server, the secure garbled circuits are adopted to propose the privacy-preserving weight average protocol. Security and performance are analyzed.

1. Introduction

In the modern Internet of Things (IoT), huge data are collected from sensor-networks and need to be provided for analysis by high-effective techniques, such as data mining. This process requires enormous computation and storage to support; cloud computing technology can provide the corresponding support. However, this process may leak the privacy of participants' information. The privacy-preserving data mining (PPDM) based on encryption method has emerged as a solution to this problem.

Privacy-Preserving Data Mining Framework. Considering different frameworks and theories, PPDM was originated by Lindell et al. [1] and Agrawal et al. [2] in 2002, respectively. Lindell's framework is essentially a secure cryptography-based two-participant computation protocol without outsourcing. In other words, two parties can interactively compute $(x_1 + x_2)\ln(x_1 + x_2)$ on their private input x_1 and x_2. Agrawal's framework is essentially a single-participant disturbance-based data storage and computation outsourcing algorithm. In particular, one party can upload disturbed data

to server for private computation. With the development of cloud computation and IoT, a multiparty storage and computation outsourcing framework is preferred.

Cryptography-based privacy-preserving data mining supporting one-party outsourcing has been studied [3, 4], with homomorphic encryption. However, multiple-key homomorphic encryption is an open problem when multiple parties are involved in the outsourcing framework. For example, how to execute ciphertext addition and multiplication on ciphertexts encrypted by different public keys?

Security Models. We usually consider two different security models, including the semihonest and malicious security model. The definition in the semihonest model requires that all the users need to follow the rules of protocol. But we allow the dishonest users to obtain internal states of the other users. In the malicious model, different from the first security model, the corrupted users are allowed to deviate from the specified protocol. The success of the adversary means that the adversary can get the results of these protocols.

Data Distribution. Three types of distributed datasets are defined in related works, including the horizontally distributed datasets, vertically distributed datasets, and arbitrarily distributed datasets. The users in the horizontally distributed dataset can keep divided parts for the same attributes. However, in the vertical datasets, users are allowed to keep different attributes. In the last one, the datasets can be arbitrarily divided and stored by the users.

Due to the existence of malicious participants in the real environment, malicious participants may not follow the protocol. For example, they can intentionally tamper with the data, suspend the protocol anytime during the execution of the protocol, and so on. To solve this problem, this paper combines the noncontact commitment and confusion circuit mechanism, studies the average computing protocol based on confusion circuit, and then proposes the framework of a secure cryptography-based two-participant protocol with data storage and computation outsourcing. The framework consists of two data owners and two cloud servers (cloud storage server (CSS) and cloud computing server (CCS)). Each data owner has a horizontally distributed private database that is encrypted before being outsourced to the cloud for storage and computation.

1.1. Our Contribution.

We decompose the key function of distributed ID3 decision tree, $Gain(S, A_i)$, into counting, $(x_1 + x_2)/(a_1 + a_2)$, sum, multiplication, and comparison.

In counting, we propose the Secure Equivalent Testing (SET) protocol to calculate the number of items for each attribute value based on the encrypted data.

To calculate the value of $(x_1 + x_2)/(a_1 + a_2)$ in malicious model, we implement the Outsourcing Secure Circuit (OSC) protocol.

To perform the sum and multiplication operations over ciphertext, we adopt the Paillier encryption system and implement the Secure Multiplication (SM) protocol.

To execute comparison over ciphertext, we adopt the Secure Minimum out of 2 Numbers (SMIN2) protocol.

1.2. Related Work

Distributed PPDM without Outsourcing. Distributed PPDM without outsourcing is mainly for data stored and calculated locally by the participant, based on distributed data based on various data mining methods, which can be decomposed to different operations, such as average calculation, calculation, and calculation of logarithmic vector inner product. Then the cryptography-based technology is used to design various privacy-preserving computing protocols. In 2002, Lindell and Pinkas [1] proposed a secure ID3 decision tree algorithm over horizontally partitioned data. They decompose the distributed ID3 algorithm to multilogarithmic calculation, polynomial evaluation calculation, and data comparison, and then designed the security log protocol, polynomial evaluation protocol, and secure comparison protocol, so as to achieve privacy-preserving in distributed ID3 algorithm. In 2007, Emekci et al. [5] implemented a secure addition computational protocol based on the secret sharing algorithm and extended the secure logarithmic computing protocol from

two parties to multiple parties; thus realizing the multiparty participation of the privacy protection ID3 method. However, the complexity of the algorithm increases exponentially when the participant data are more numerous. In 2012, Lory et al. [6] used Chebyshev polynomial expansion to replace Taylor expansion in [1], thus further improving the computational efficiency of secure logarithmic computing protocols. However, their agreements still have limited efficiency in the implementation of privacy protection protocols.

Different from above, in 2003, Vaidya et al. [7, 8] designed a multiparty privacy-preserving ID3 algorithm of vertically distributed data sets. They vectorized all attribute value information by constructing constraint sets and then computed it by using the method of secure intersection protocol, thus designing privacy-preserving ID3 for vertically distributed data sets.

In 2007, Han and Ng [9] proposed a multiparty distributed privacy-preserving ID3 method based on arbitrary distributed data sets. Firstly, each participant's data set is vectorized, and then the attribute value information is computed by using security intersection protocol and so on. Then, the entropy value of each attribute is computed by using security logarithm computation protocol and so on. Thus, the ID3 decision tree classification method of privacy protection based on arbitrary distributed data set is obtained. However, with the increase of the number of participants, the computing volume of the client increases exponentially.

Li et al. [10] and Gao et al. [11] addressed the Naive Bayes Learning for aggregated arbitrary distributed databases.

PPDM with Computation Outsourcing. Cryptography-based privacy-preserving data mining has a lot of encryption and decryption operations in the computation process. Therefore, it is difficult for large-scale data processing. As a measure for solving resource-restricted problems, the outsourcing technique has been widely used in cloud computing applications, such as data sharing [12, 13], data storing [14, 15], data updating [16, 17], and social network analysis [18, 19]. In this context, we need to rely on security outsourcing technology to outsource computing or storage tasks of all participants to the cloud to process, thus greatly reducing the computing/storage load of each participant. In 2014, Liu et al. [3] adopted a new encryption scheme that supports both addition and multiplication over cipher texts. In this scheme, most of the computations are performed on the cloud, which reduces the computation workload of the data owner. However, the scheme is limited to a single party's data mining operation. Chen et al. [20] designed new algorithms for secure outsourcing of modular exponentiations. In 2015, Bost et al. [21] proposed the privacy-preserving hyperplane decision, Naive Bayesian, and decision tree classification algorithms, and through the semihonest model, secure two-party computation model to prove that the above scheme can satisfy the semantic security (Semantic Security); and the related protocol makes it possible to design an adaptive enhancement algorithm (Adaptive Boosting) combine to further enhance the accuracy of the algorithm; building a classifier can be used to construct the privacy protection of the library, the further development of the classification

algorithms for privacy-preserving technology in the future lays a solid foundation.

PPDM with Multiparticipant Data Storage and Computation Outsourcing. In 2013, Peter et al. [22] proposed a new solution for the outsourcing of multiparty computation. Such a technique can be used in our setting. But as the security analysis in the previous works, they can only achieve security in the semihonest model. In [23], a new protocol was proposed to achieve data mining for two parties. In [24], association rule mining was addressed in the malicious model. In [25], the privacy-preserving *KNN* classification was addressed. In [26], the deep learning task was addressed. Besides the above related work, several fundamental secure algorithms, such as dynamic homomorphic encryption [27, 28], authentication [29, 30], and light-weight multiparty computation [31], which have also been considered in the malicious model, have been proposed. However, to the best of our knowledge, no existing study has considered a method for outsourcing computation in the malicious model.

In this study, the secure outsourcing of ID3 data mining is considered in the malicious model for the cloud environment. We show how to solve the outsourcing problem for ID3 protocol over horizontally partitioned data.

2. Preliminaries

In this section, we present a brief overview of the preliminaries used in this paper, including the ID3 decision tree algorithm, Paillier's homomorphic encryption scheme, and the other related protocols.

2.1. Distribute ID3 Decision Tree Algorithm. The ID3 algorithm description is given as follows. It builds a decision tree in a top-down manner with the information of samples. Starting at the root, the best object classification will be obtained. The best prediction is computed with the *information gain*. The *information gain* of an attribute A_t is defined as

$$Gain\left(S, A_t\right) = Entropy\left(S\right)$$
$$- \sum_{a_{tj} \in A_t} \frac{\left|S_{a_{tj}}\right|}{|S|} Entropy\left(S_{a_{tj}}\right) \tag{1}$$

Assume that there are 2 parties, $\mathscr{P} = \{P_i \mid i = 1, 2\}$, with 2 databases, $\mathscr{S} = \{S_i \mid i = 1, 2\}$. Each party P_i has one database S_i. All databases share the same general attribute (column) set $\mathscr{A} = \{A_t \mid t = 1, 2, .., p\}$ and each attribute A_t has several general discrete attribute values, denoted by $A_t = \{a_{tj} \mid j = 1, 2, \ldots, m_t\}$, and one class attribute $C = \{c_j \mid j = 1, 2, ..., m\}$.

Without considering privacy, each party P_i shares his own $|S_{a_{tj}c_j}|_i, |S_{a_{tj}}|_i$ and $|S_{c_j}|_i$ to all other parties. As a result, any party can calculate $Entropy(S)$ and $Entropy(S_{a_{tj}})$.

$$Entropy\left(S\right) = \sum_{j=1}^{m} - \frac{\left|S_{c_j}\right|}{|S|} \log_2 \frac{\left|S_{c_j}\right|}{|S|}$$
$$= \sum_{j=1}^{m} - \frac{\sum_{i=1}^{n} \left|S_{c_j}\right|_i}{|S|} \log_2 \frac{\sum_{i=1}^{n} \left|S_{c_j}\right|_i}{|S|} \tag{2}$$

where S_{c_j} is the subset of S with tuples that have value c_j for class attribute C. $|S_{c_j}|_i$ equals the set of transactions with class attribute C set to c_j in database S_i.

Then the value of $Entropy(S_{a_{tj}})$ can be calculated as

$$Entropy\left(S_{a_{tj}}\right) = \sum_{j=1}^{m_t} - \frac{\sum_{i=1}^{n} \left|S_{a_{tj}c_j}\right|_i}{\sum_{i=1}^{n} \left|S_{a_{tj}}\right|_i} \cdot \log_2 \frac{\sum_{i=1}^{n} \left|S_{a_{tj}c_j}\right|_i}{\sum_{i=1}^{n} \left|S_{a_{tj}}\right|_i} \tag{3}$$

where $S_{a_{tj}c_j}$ is the subset of S with tuples that have values a_j for attribute A_t and c_j for class attribute C.

Therefore, (3) can be easily computed by party P_i and parties P_j ($j \neq i$) all of the values $|S_{a_{tj}}|_i$ and $|S_{a_{tj}c_j}|_i$ from its database. Each party P_j then sums these together with the values $|S_{a_{tj}}|_j$ and $|S_{a_{tj}c_j}|_j$ from its database and completes the computation.

Then each party can calculate $Gain(S, A_t)$ value at its own side.

2.2. Paillier's Homomorphic Encryption Scheme. Homomorphic encryption is a special type of encryption in which the result of applying a special algebraic operation to plain texts can be obtained by applying another algebraic operation (which may be different or the same) to the corresponding ciphertexts. Thus, even when the user does not know the plain texts, he/she can still obtain the results of applying that algebraic operation to the plain texts.

Let m_1 and m_2 be two plain texts with encryptions $E(m_1)$ and $E(m_2)$, respectively.

The Paillier encryption scheme [32] is described as follows:

$$E\left(m_1\right) \oplus E\left(m_2\right) = E\left(m_1 + m_2\right) \tag{4}$$

2.3. Li's Symmetric Homomorphic Encryption Scheme. The description of symmetric homomorphic encryption scheme proposed by Li et al. [33] is as follows.

  (i) **KeyGen**():

$$\left(s, q, p\right) \longleftarrow KeyGen\left(\lambda\right) \tag{5}$$

    KeyGen() is used to generate key for users as $SK = (s, q)$. p and q are primes with the condition that $p \gg q$. s is chosen from $\mathbb{Z}_N^*$.

  (ii) **Enc$_{sk}$**():

$$Enc_{sk}\left(m, d\right) = s^d\left(rq + m\right) \mod p. \tag{6}$$

d is a small positive integer, which is denoted as **ciphertext degree** in this paper.

1. The computation at Data owner:
   Compute $\tau = Enc_{sk}(1)$ for the cloud.
2. The computation at cloud:
   u, v are chosen such that
   $$u \gg v$$
   $$v \gg \max(\alpha, \beta)$$
   $$\frac{(q-1)}{2} \gg \beta \times u + v$$
   $$-\alpha \times u + v \gg -\frac{(q-1)}{2}$$
3. The cloud compute the following value for the data owner:
   $\Phi = cu + \tau v \bmod p$, and sends Φ.
4. Data Owner computes the following values:
   $\varphi = Dec_{sk}(SK, \Phi, d) = (m \times u + v) \bmod q$,
   and compares φ with $(q-1)/2$.
   If $\varphi < (q-1)/2, m \geq 0$. Otherwise, $m < 0$.

ALGORITHM 1: Secure outsourcing comparison (OSCP).

(iii) **Dec$_{sk}$():**

$$Dec_{sk}(c, d) = \left(c \times s^{-d} \bmod p\right) \quad \bmod q. \tag{7}$$

2.3.1. Properties of the Proposed Homomorphic Encryption

Homomorphic Multiplication. Let c_1, c_2 be the ciphertexts of two plaintexts m_1, m_2. Then we have $c_1 = s^{d_1}(r_1q + m_1) \bmod p$ and $c_2 = s^{d_2}(r_2q + m_2) \bmod p$ for some random ingredients r_1 and r_2. And we can obtain that

$$
\begin{aligned}
&(c_1 \times c_2) \bmod p \\
&= s^{d_1}(r_1q + m_1) \bmod p \times s^{d_2}(r_2q + m_2) \bmod p \\
&= s^{d_1+d_2}\left((r_1r2q + r_1m_2 + r_2m_1)q + m_1 \times m_2\right) \bmod p \\
&= Enc_{sk}(m_1 \times m_2, d_1 + d_2).
\end{aligned}
\tag{8}
$$

Homomorphic Addition

$$
\begin{aligned}
(c_1 + c_2) \bmod p &= s^d(r_1q + m_1) \bmod p \\
&\quad + s^d(r_2q + m_2) \bmod p \\
&= s^d((r_1 + r2)q + m_1 + m_2) \bmod p \\
&= Enc_{sk}(m_1 + m_2, d)
\end{aligned}
\tag{9}
$$

Readers may refer to [18] for details on the scheme.

2.4. Garbling Scheme.
A garbling scheme [34] consists of four algorithms, which is denoted by $G = (Gb, En, De, Ev)$. f can be transformed by Gb into (F, e, d). Note that F is the garbled circuit. The encoding and decoding information algorithms are denoted by e, d. The output of garbled Y can be encrypted and get the result $y = De(d, Y)$.

2.5. Noninteractive Commitment.
A noninteractive commitment scheme [35] is also required in our paper, denoted by

(Com_{crs}, Chk_{crs}). The distribution of $Com_{crs}(x; r)$ is determined by the value of r as $Com_{crs}(x)$.

2.6. Basic Cryptographic Subprotocols.
In this section, we present a set of cryptographic subprotocols that will be used as subroutines when constructing the proposed protocol.

2.6.1. Outsourcing Secure Comparison Protocol (OSCP).
The value of m is kept secure from the cloud and users. The value of $c = E(m \bmod q)$ is computed. SK is kept by the data owner (Algorithm 1).

2.6.2. Secure Equivalent Testing Protocol (SET).
With two ciphertexts $c_1 = Enc_{sk}(m_1)$ and $c_2 = Enc_{sk}(m_2)$, **SET** is to compute f and decide if the plaintexts are identical ($m_1 = m_2$) (Algorithm 2).

2.6.3. Secure Multiplication Protocol (SMP).
The algorithm is described as in Algorithm 3.

2.6.4. Secure Minimum out of 2 Numbers Protocol (SMIN2).
The algorithm is described as in Algorithm 4.

2.7. Secure Circuit Protocol (SCP).
We denote the three parties of the protocol by CSS_1, CSS_2, and CCS and their respective inputs by x_1, x_2, or x_3^*. Their goal is to securely compute the function $y = f(x_1, x_2, x_3^*) = x_1 \cdot x_2/x_3^*$ [34] (Algorithm 5).

For simplicity, we assume that $|x_i| = |y| = m$. All communication between parties is via private point-to-point channels. Next, we assume that CSS_1 and CSS_2 can learn the same output y, while CCS can get the garbled values for the portion of the output wires corresponding to its own outputs only. CCS cannot get the output y with these garbled values. This protocol uses a garbling scheme, a four-tuple algorithm $\delta = (Gb, En, De, Ev)$, as the underlying algorithm. Gb is a randomized garbling algorithm that transforms a function of a triple. En and De are encoding and decoding algorithms, respectively. Ev is an algorithm that produces a garbled output

Two ciphertext are computed by the cloud as $c_1 = Enc_{sk}(m_1)$ and $c_2 = Enc_{sk}(m_2)$.
1. The cloud computes $c_{00} = Enc_{sk}(m_{00}) = c_1 - c_2$ and $c_{01} = Enc_{sk}(m_{01}) = c_2 - c_1$.
2. Check if $m_{00} \geq 0$ or $m_{01} \geq 0$ and computes
if $m_{00} \geq 0 \wedge m_{01} < 0$, $f_0 = 1$, $f_1 = 0$. else if $m_{00} < 0 \wedge m_{01} \geq 0$, $f_0 = 0$, $f_1 = 1$.
3. The value of f is computed as follows:
$$f = f_0 \oplus f_1.$$
If $f = 0$, set $m_1 = m_2$.

ALGORITHM 2: Secure equivalent testing protocol (SET).

The values are computed $E_{pk}(x)$ and $E_{pk}(y)$; P keeps (pk, sk).
1. C:
   (a) Choose $r_x, r_y \in Z_N$
   (b) $x' \longleftarrow E_{pk}(x)E_{pk}(r_x)$, $y' \longleftarrow E_{pk}(y)E_{pk}(r_y)$
   (c) computes and sends x', y' to P
2. P computes:
   (a) $h_x \longleftarrow D_{sk}(x')$, $h_y \longleftarrow D_{sk}(y')$, $h \longleftarrow h_x h_y \bmod N$, $h' \longleftarrow E_{pk}(h)$
   (b) The value of h' is sent to C
3. C computes:
   (a) $s \longleftarrow h' E_{pk}(x)^{N-r_y}$, $s' \longleftarrow s E_{pk}(y)^{N-r_x}$
   (b) $E_{pk}(xy) \longleftarrow s' E_{pk}(r_x r_y)^{N-1}$

ALGORITHM 3: $SM(E_{pk}(x), E_{pk}(y)) \longrightarrow E_{pk}(xy)$.

1. C:
   (a) The function is chosen F
   (b)
   for $i = 1$ *to* λ **do**
     $E_{pk}(u_i v_i) \longleftarrow SM(E_{pk}(u_i), E_{pk}(v_i))$.
     if $F: u > v$ **then**
       $W_i \longleftarrow E_{pk}(u_i) E_{pk}(u_i v_i)^{N-1}$,
     end
     else
       $W_i \longleftarrow E_{pk}(v_i) E_{pk}(u_i v_i)^{N-1}$
     end
     $G_i \longleftarrow E_{pk}(u_i \oplus v_i)$, $H_i \longleftarrow H_{i-1}^{r_i} G_i$; $r_i \in_R$
     Z_N and $H_0 = E(0)$
     $\Phi_i \longleftarrow E_{pk}(-1)H_i$,
     $L_i \longleftarrow W_i \Phi_i^{r_i'}$; $r_i' \in Z_N$
   end
   (c) Sends L to P
2. P:
   (a) $M_i \longleftarrow D(L_i)$, for $1 \leq i \leq \lambda$
   (b)
   if $\exists j$ *such that* $M_j = 1$ **then**
     $\alpha \longleftarrow 1$ (which means $u > v$)
   end
   else
     $\alpha \longleftarrow 0$ (which means $u < v$)
   end

ALGORITHM 4: Secure Minimum out of 2 Numbers Protocol (SMIN2).

Input: CSS_1 has x_1, CSS_2 has x_2
Output: $(x_1 + x_2)/(a_1 + a_2)$
1. CCS:
 (a) Randomly selects crs for commitment, and sends crs to CSS_1 and CSS_2.
 (b) Generates the random number λ and share it secretly as $\lambda = \lambda_1 \oplus \lambda_2$, sends $lambda_1$ to CSS_1, and sends $lambda_2$ to CSS_2.
2. CSS_1: Select seed $r \longleftarrow 0, 1^k$ for pseudo random function PRF and send r to CSS_2.
3. CSS_1 and CSS_2: (a) Generate corresponding circuit $Gb(1^\kappa, f) \longrightarrow (F, e, d)$ based on function $f = (x_1 + x_2) * \lambda$. (b) Random
selection of $b_1, b_2 \longleftarrow \{0, 1\}^{4m}$ and generate the following commitments for all $j \in [4m]$ and $t \in \{0, 1\}$:
    $(C_{1,j}^t, \sigma_{1,j}^t) \longleftarrow Com_{crs}(e[j, b_1[j] \oplus t])$.
    $(C_{2,j}^t, \sigma_{2,j}^t) \longleftarrow Com_{crs}(e[j, b_2[j] \oplus t])$.
 (c) CSS_1 and CSS_2 send the following information to CCS:
    $(b_1[2m+1 \cdots 4m], F, \{C_{1,j}^t\}_{j,t})$.
    $(b_2[2m+1 \cdots 4m], F, \{C_{2,j}^t\}_{j,t})$.
4. CCS: Abort if CSS_1 and CSS_2 report different values for these items.
5. CSS_1 and CSS_2:
 (a) CSS_1 sends decommitment
$\sigma_{1,j}^{x_1[j]\oplus b_1[j]}, \sigma_{1,2m+j}^{\lambda_1[j]\oplus b_1[2m+j]}, \sigma_{2,j}^{x_1[j]\oplus b_2[j]}$ and $\sigma_{2,2m+j}^{\lambda_1[j]\oplus b_2[2m+j]}$ to CCS
 (b) CSS_2 sends decommitment $\sigma_{1,m+j}^{x_2[j]\oplus b_1[m+j]}, \sigma_{1,3m+j}^{\lambda_2[j]\oplus b_1[3m+j]}, \sigma_{2,m+j}^{x_2[j]\oplus b_2[m+j]}$ and $\sigma_{2,3m+j}^{\lambda_2[j]\oplus b_2[3m+j]}$ to CCS.
6. CCS: (a) For $j \in [4m]$, compute $X[j] = Chk_{crs}(C_{1,j}^{o[j]}, \sigma_{1,j}^{o[j]})$, $X'[j] = Chk_{crs}(C_{2,j}^{o[j]}, \sigma_{2,j}^{o[j]})$, for the appropriate $o[j]$. If any call to
    Chk returns $\perp$, then abort. Similarly, CCS knows the values $b_1[2m+1 \cdots 4m]$ and $b_2[2m+1 \cdots 4m]$, and aborts if CSS_1 or CSS_2
    can not open the corresponding commitments of λ_1 and λ_2: $C_{1,2m+j}^{\lambda_1[j]\oplus b_1[2m+j]}, C_{1,3m+j}^{\lambda_2[j]\oplus b_1[3m+j]}, C_{2,2m+j}^{\lambda_1[j]\oplus b_2[2m+j]}$ and $C_{2,3m+j}^{\lambda_2[j]\oplus b_2[3m+j]}$.
 (b) Run $Y \longleftarrow Ev(F, X)$ and $Y' \longleftarrow Ev(F, X')$, then broadcasts Y and Y' to CSS_1 and CSS_2.
7. CSS_1 and CSS_2: Compute $(x_1 + x_2)/(a_1 + a_2) = De(d, Y)/De(d, Y')$.

ALGORITHM 5: Secure circuit protocol (SCP).

for a garbled input and garbled circuit. Further, Chk is an algorithm that can verify commitments.

3. Outsourcing Privacy-Preserving ID3 Decision Tree Algorithm in Malicious Model

In this section, we present our secure outsourcing ID3 decision tree in cloud computing using the homomorphic encryption scheme and subprotocols proposed in Section 2 as building blocks.

3.1. Main Concept. The aim is to privately compute ID3 over encrypted databases, and the key is to find privately the attribute A for which Gain is maximum. From the above description, the key value which needs to be calculated with other parties is $Entropy(S_{a_i})$.

Since all the data was encrypted and sent to the cloud, the cloud server can count the number of $|S(c_k)|_t$, $|S|_{it}$ using the **SET** protocol described in Section 2. Now, (3) can be executed as $(x_1 + x_2)/(a_1 + a_2)\log_2(x_1 + x_2)/(a_1 + a_2)$, and the calculation of the logarithmic operation can be performed in CSS. The value to be calculated is the value of $c_1 = (x_1 + x_2)/(a_1 + a_2)$, which can be easily determined using our **SCP** protocol as explained in Section 2. Then, all the parties can calculate the value of $Entropy(S)$ independently.

3.2. System Model. The system model is shown in Figure 1, which includes two data owners and cloud servers (cloud storage server $\{CSS_1, CSS_2\}$, and cloud computing server CCS). Each data owner owns a private data set that is encrypted and outsourced to cloud server storage. Data

owners can request cloud server to process ID3 data on encrypted data. At the same time, CSS and CCS servers participate in supporting the outsourcing privacy protection ID3 data mining algorithm steps; after the implementation of the algorithm, the final results are sent to the data owner. Assuming that the data owner and the CSS server are semihonest participants, CCS is a malicious participant.

3.3. Details of the Proposed Algorithm. Our securely outsourcing ID3 decision tree (SOID3) algorithm is detailed as follows:

(1) P_1 and P_2 run KeyGen(λ) to generate the secret key $SK_i, i = 1, 2$ and a public parameter p of Li's homomorphic encryption scheme. Further, each party shares p with the other party and the cloud but shares SK_i only with itself.

(2) Each party uses its key SK_i to encrypt every attribute value of its database, and then outsources the encrypted database to the CSS (CSS_1 and CSS_2).

(3) The CSS_1 and CSS_2 use the **SET** protocol to calculate the value of $|S_{a_j}|_i$ and $|S_{a_j}(c_k)|_i$ for each attribute with each party P_i.

(4) Each party generates its Paillier public and private keys (pk_i, sk_i), $i = 1, 2$, and sends the public keys to the CSS_1 and CSS_2.

(5) CCS, CSS_1, and CSS_2 jointly use the **SCP** protocol to compute $(x_1 + x_2)/(a_1 + a_2)$. Here, CSS_1 has (x_1, a_1), and CSS_2 has (x_2, a_2).

(6) Each party decrypts the received information, calculates it with the logarithmic operation of $(x_1 + x_2/a_1 +$
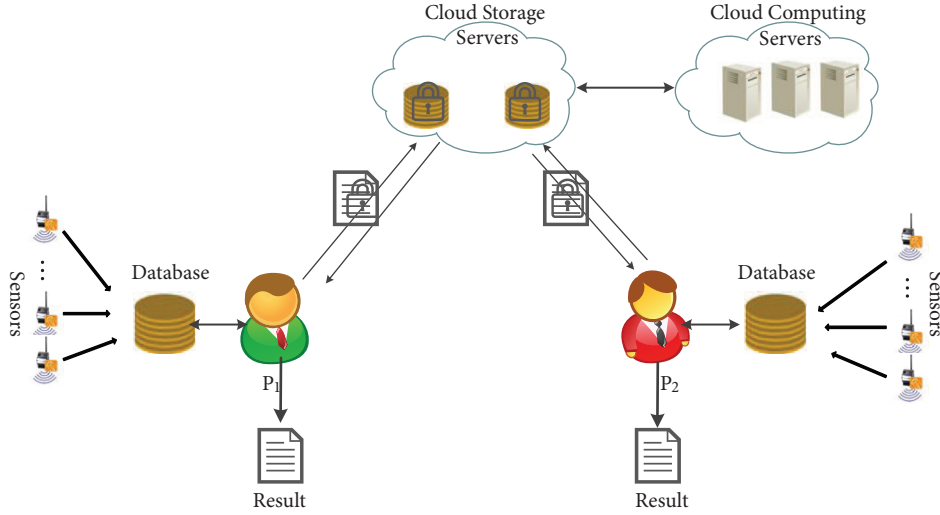
FIGURE 1: System model under consideration.

$a_2)\log_2(x_1 + x_2/a_1 + a_2)$, and then encrypts it with its public key. Then, it sends it back to the cloud.

(7) After getting the result, CSS_1 and CSS_2 use the **SMIN2** protocol to select the ciphertext data with the minimum value and then further select the attribute label with the maximum information gain and return it to each participant.

(8) The participants divide the data sets and build tree nodes. Then, go to Step (3) until termination.

4. Security Analysis

In this section, we prove that the secure outsourcing ID3 decision tree (SOID3) algorithm can offer protection against the malicious cloud server.

Theorem 1. *The SOID3 algorithm can achieve privacy for each party and the semihonest cloud storage server.*

Proof. We mainly consider the security model under the noncollusive semihonest model and the semihonest cloud server. Suppose there are two parties, P_1 and P_2, and cloud storage server CSS.

Let $P = (P_1, P_2, CSS)$ be the participants of all protocols. Consider three types of attackers (A_{P_1}, A_{P_2}, and A_{CSS}) that can invade P_1, P_2, and CSS. In the real model, P_1 and P_2 have data sets D_x and D_y, respectively, and CSS has encrypted data sets $Enc(D_x)$ and $Enc(D_y)$. Make $H \subset P$ a collection of honest participants. For all $P_i \epsilon H$, out_{P_i} indicates the output of P_i. If P_i is invaded, out_{P_i} represents all views of participant P_i in running protocol Π.

For each $P^* \in P$, the attacker $A = (A_{P_1}, A_{P_2}, A_{CSS})$ view in the runtime protocol Π can be defined as

$$\text{REAL}_{\Pi,A,H}^{P^*}\left(D_x, D_y\right) = out_{P_i} : P_i \epsilon H \cup out_{P^*} \quad (10)$$

In the ideal model, there exists an ideal model F for function f, and all participants can interact with the model F. That is, Challenger DP_a and participant P_i can send data x and y to F. If D_x or D_y is $\perp$, F returns $\perp$. Finally, F can return

$f(D_x, D_y)$ to challenger DP_a. As mentioned earlier, $H \subset P$ is a collection of honest participants. For each participant $P_i \epsilon H$ in the collection, return the out_{P_i} as F output to P_i. If P_i is intruded on by a semihonest attacker, out_{P_i} is still consistent with the output of P_i in previous realistic models.

For all $P^* \in P$, in the ideal model, in the presence of independent simulators $Sim = (Sim_{P_1}, Sim_{P_2}, Sim_{CSS})$, the P^* view is

$$\text{IDEAL}_{F,Sim,H}^{P^*}\left(D_x, D_y\right) = out_{P_i} : P_i \epsilon H \cup out_{P^*} \quad (11)$$

Therefore, it is considered that the protocol Π is secure in the presence of noncolluded semitruthful attackers.

Definition 2. Let f be a deterministic functionality among parties in P. Let $H \subset P$ be the subset of honest parties in P. We say that Π securely realizes f if there exists a set $Sim = \{Sim_{P_1}, Sim_{P_2}, Sim_S\}$ of PPT transformations (where $Sim_{D_a} = Sim_{P_1}(A_{P_1})$ and so on) such that for all semihonest PPT adversaries $A = \{A_{P_1}, A_{P_2}, A_S\}$, for all inputs D_x, D_y and auxiliary inputs z, and for all parties $P \in P$ the following holds:

$$\text{REAL}_{\Pi,A,H,z}^{P^*}\left(\lambda, x, y\right)_{\lambda \in N} \stackrel{\cong}{\equiv}$$
$$\text{IDEAL}_{F,Sim,H,z}^{P^*}\left(\lambda, x, y\right)_{\lambda \in N}, \quad (12)$$

where $\stackrel{\cong}{\equiv}$ denotes computational indistinguishability. $\square$

Theorem 2. *The SOID3 algorithm is secure with the semihonest cloud storage server and the malicious cloud computing server.*

Proof. First consider the case where CSS_1 or CSS_2 is corrupted. It is necessary to prove that, in the **SCP** protocol, the ideal model and the realistic model are not distinguishable. That is, in the following interactions, it is impossible to distinguish between the various types of interaction information and outputs of the participants in the ideal model and the real model.

(1) In the real model, assume that there is an emulator that can simulate various behaviors of a semihonest participant CSS_1 (or CSS_2), and receive inputs (x_1, a_2) and (x_2, a_2) from the execution environment of the protocol. At the same time, the simulator can simulate the function F_f, which sends all inputs (x_1, a_1) and (x_2, a_2) to the simulated F_f. Since the simulator does not do anything computed by F_f, there is no difference between the real F_f and the simulated F_f from the execution environment point of view.

(2) Because in Step 2, CSS_1 and CSS_2 uniformly select the seed r of Pseudo-Random Function (PRF), the PRF security shows that the real model in Step 2 is indistinguishable from the ideal model.

(3) In Step 3, we modify the simulator, which knows in advance what promises will be opened when the simulator generates commitment C. First, the simulator selects the random numbers o_1, o_2 that can be marked which promises to be opened and calculates the values of $b_1 = o_1 \oplus x_1 \parallel x_2$, $b_2 = o_2 \oplus a_1 \parallel a_2$. At this point, the simulator has obtained the values of $x_1, x_2,$ and a_1, a_2. Then, the simulator can submit the markings that promise not to be opened. In this process, due to the concealment of commitment, the realistic model and the ideal model are equally indistinguishable.

(4) In Step 6a, the simulated CSS_1 and CSS_2 stop executing when $De(D, \widetilde{Y}) = \perp$. Change the emulator to make $\widetilde{Y} \neq Ev(F, X)$. By obfuscating the authenticity of the circuit, CCS has only negligible probability to obtain $\widetilde{Y} \neq Ev(F, X)$ in $De(d, \widetilde{Y}) = \perp$. Therefore, in this step, the realistic model and the ideal model are equally indistinguishable.

(5) In Step 6b, the correctness of the obfuscation circuit guarantees that both CSS_1 and CSS_2 of the analog can be output. Therefore, if there is no pause in 6a, we can modify the simulator to an analog obfuscation circuit that generates (F, X, d). We can simulate the output of CSS_1 and CSS_2 by simulating the instructions of F_f. According to the security of the confusing circuit, the real model is also indistinguishable from the ideal model in this step.

Therefore, in this protocol, the execution environment can not distinguish between the realistic model and the ideal model. And the protocol is secure when CCS is a malicious participant. □

5. Performance Analysis

In this paper, we consider that CSS has a strong calculation ability and we ignore its computation time. Each data owner does not need to store the ciphertext but can just use the public key to encrypt the message and the private key to decrypt the ciphertext.

In each iteration, first, each data owner will execute the **SBD** protocol and **SMIN2** protocol with the cloud. There are two interactions in the **SBD** protocol and 2k interactions in the **SMIN2** protocol. Then, CSS_1, CSS_2, and CCS will execute 6 interactions in the **SCP** protocol. Finally, each data owner will execute 1 interaction when it goes to the new iteration. We assume that t is the iteration time, so the communication traffic is at most $O(k * t)$.

In this paper, a secure average computing protocol based on **SCP** is implemented. The server selected in the
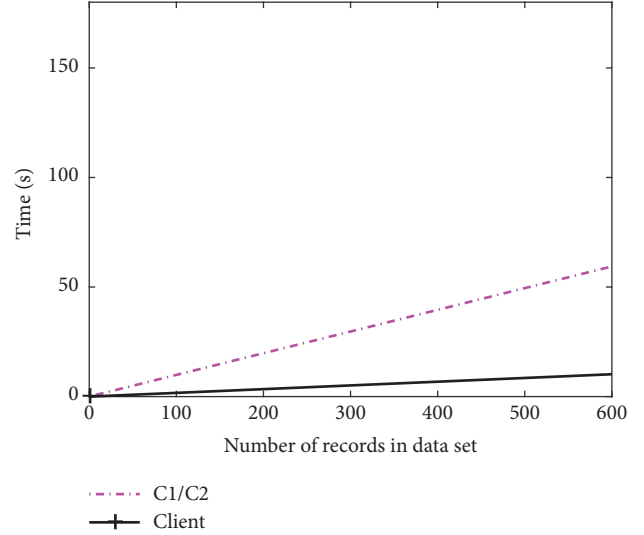
FIGURE 2: Performance measurements for our SOID3 with 2 participants.

experimental environment is CPU: Intel (R) Xeon (R) CPU E5-2620 v3@2.40GHz∗2, memory: 32G, operating system Ubuntu 16.04.4 LTS version. In the experiment, AES-128 is chosen as the basic encryption method of the confusion circuit, and the open source code of JustGarble is changed, and the commitment protocol is implemented based on SHA-256. Finally, the average values obtained from experiments are as follows.

In our secure outsourcing ID3 decision tree (SOID3) algorithm experiment, two participants were tested with different numbers of records. The experimental results are shown in Figure 2.

From Figure 2, since the client is only responsible for encrypting uploaded data, the time consumption is very low. In the cloud, CCS and CSS servers need to run **SCP** protocol, resulting in a lot of time consumption (Table 1). The main reason is that a large number of bit commitment processing is needed in the obfuscation circuit, and the performance improvement will be focused on this issue in the follow-up work.

6. Conclusion

In this paper, we proposed a secure outsourcing ID3 decision tree algorithm for two parties of the malicious model. Our algorithm can preserve the privacy of the users' data as well as that of the data mining scheme for the cloud servers. The parties can get only the result trees and have no knowledge about the data mining scheme. Moreover, the cloud servers cannot get any private information about the parties. In summary, our protocol offers protection against malicious cloud servers.

In the future, we intend to extend our algorithm to vertical and arbitrary partitioning in the malicious model. In addition, we plan to extend our algorithm to a general multiparty privacy-preserving framework suitable for other useful schemes, such as random decision tree, Bayes, SVM,

TABLE 1: Time cost of the SCP protocol.

AND	XOR	OR	Input size (bits)	Output size (bits)	CSS_1/CSS_2 (ms)	CCS (ms)
5090	4034	2016	129	65	26	47

and other data mining methods, and can be extended for use in the wireless sensor-networks [36, 37].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Lindell and B. Pinkas, "Privacy preserving data mining," *Journal of Cryptology*, vol. 15, no. 3, pp. 177–206, 2002.

[2] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 439–450, 2000.

[3] D. Liu, E. Bertino, and X. Yi, "Privacy of outsourced *k*-means clustering," in *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, ASIA CCS 2014*, pp. 123–133, June 2014.

[4] P. Li, J. Li, Z. Huang, C.-Z. Gao, W.-B. Chen, and K. Chen, "Privacy-preserving outsourced classification in cloud computing," *Cluster Computing*, pp. 1–10, 2017.

[5] F. Emekci, O. D. Sahin, D. Agrawal, and A. El Abbadi, "Privacy preserving decision tree learning over multiple parties," *Data & Knowledge Engineering*, vol. 63, no. 2, pp. 348–361, 2007.

[6] P. Lory, "Enhancing the Efficiency in Privacy Preserving Learning of Decision Trees in Partitioned Databases," in *Privacy in Statistical Databases*, vol. 7556 of *Lecture Notes in Computer Science*, pp. 322–335, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[7] J. Vaidya and C. Clifton, "Privacy-preserving *K*-means clustering over vertically partitioned data," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)*, pp. 206–215, Washington, DC , USA, August 2003.

[8] J. Zhan, S. Matwin, and L. W. Chang, "Privacy-Preserving Naive Bayesian Classification over Horizontally Partitioned Data," in *Data Mining: Foundations and Practice*, vol. 118 of *Studies in Computational Intelligence*, pp. 529–538, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[9] S. Han and W. K. Ng, "Multi-party privacy-preserving decision trees for arbitrarily partitioned data," *International Journal of Intelligent Control and Systems*, vol. 12, no. 4, 2007.

[10] T. Li, J. Li, Z. Liu, P. Li, and C. Jia, "Differentially private Naive Bayes learning over multiple data sources," *Information Sciences*, vol. 444, pp. 89–104, 2018.

[11] C. Gao, Q. Cheng, P. He, W. Susilo, and J. Li, "Privacy-preserving Naive Bayes classifiers secure against the substitution-then-comparison attack," *Information Sciences*, vol. 444, pp. 72–88, 2018.

[12] J. Shen, T. Zhou, X. Chen, J. Li, and W. Susilo, "Anonymous and Traceable Group Data Sharing in Cloud Computing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 912–925, 2018.

[13] Z. Cai, H. Yan, P. Li, Z.-A. Huang, and C. Gao, "Towards secure and flexible EHR sharing in mobile health cloud under static assumptions," *Cluster Computing*, vol. 20, no. 3, pp. 2415–2422, 2017.

[14] J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1206–1216, 2015.

[15] Z. Liu, Y. Huang, J. Li, X. Cheng, and C. Shen, "DivORAM: Towards a practical oblivious RAM with variable block size," *Information Sciences*, vol. 447, pp. 1–11, 2018.

[16] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, "Verifiable computation over large database with incremental updates," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 65, no. 10, pp. 3184–3195, 2016.

[17] J. Shen, C. Wang, T. Li, X. Chen, X. Huang, and Z.-H. Zhan, "Secure data uploading scheme for a smart home system," *Information Sciences*, vol. 453, pp. 186–197, 2018.

[18] S. Chen, G. Wang, G. Yan, and D. Xie, "Multi-dimensional fuzzy trust evaluation for mobile social networks based on dynamic community structures," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 7, p. e3901, 2017.

[19] E. Luo, Q. Liu, J. H. Abawajy, and G. Wang, "Privacy-preserving multi-hop profile-matching protocol for proximity mobile social networks," *Future Generation Computer Systems*, vol. 68, pp. 222–233, 2017.

[20] X. F. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2386–2396, 2014.

[21] R. Bost, R. A. Popa, S. Tu, and S. Goldwasser, "Machine Learning Classification over Encrypted Data," in *Proceedings of the Network and Distributed System Security Symposium*, San Diego, CA.

[22] A. Peter, E. Tews, and S. Katzenbeisser, "Efficiently outsourcing multiparty computation under multiple keys," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2046–2058, 2013.

[23] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, "A Practical Differentially Private Random Decision Tree Classifier," in

Proceedings of the 2009 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 114–121, Miami, FL, USA, December 2009.

[24] B. Gilburd, A. Schuster, and R. Wolff, "Privacy-preserving data mining on data grids in the presence of malicious participants," in *Proceedings of the Proceedings. 13th IEEE International Symposium on High performance Distributed Computing, 2004.*, pp. 225–234, Honolulu, HI, USA.

[25] D. Shah and S. Zhong, "Two methods for privacy preserving data mining with malicious participants," *Information Sciences*, vol. 177, no. 23, pp. 5468–5483, 2007.

[26] P. Li, J. Li, Z. Huang et al., "Multi-key privacy-preserving deep learning in cloud computing," *Future Generation Computer Systems*, vol. 74, pp. 76–85, 2017.

[27] Q. Lin, H. Yan, Z. Huang, W. Chen, J. Shen, and Y. Tang, "An ID-based linearly homomorphic signature scheme and its application in blockchain," *IEEE Access*, vol. PP, no. 99, pp. 1-1, 2018.

[28] J. Xu, L. Wei, Y. Zhang, A. Wang, F. Zhou, and C. Gao, "Dynamic Fully Homomorphic encryption-based Merkle Tree for lightweight streaming authenticated data structures," *Journal of Network and Computer Applications*, vol. 107, pp. 113–124, 2018.

[29] J. Shen, Z. Gui, S. Ji, J. Shen, H. Tan, and Y. Tang, "Cloud-aided lightweight certificateless authentication protocol with anonymity for wireless body area networks," *Journal of Network and Computer Applications*, vol. 106, pp. 117–123, 2018.

[30] X. Zhang, Y. Tan, C. Liang, Y. Li, and J. Li, "A Covert Channel Over VoLTE via Adjusting Silence Periods," *IEEE Access*, vol. 6, pp. 9292–9302, 2018.

[31] Y. Wang, T. Li, H. Qin et al., "A brief survey on secure multi-party computing in the presence of rational parties," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 6, pp. 807–824, 2015.

[32] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology—EURO-CRYPT '99*, vol. 1592, pp. 223–238, Springer, 1999.

[33] L. Li, R. Lu, K.-K. R. Choo, A. Datta, and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1547–1861, 2016.

[34] P. Mohassel, M. Rosulek, and Y. Zhang, "Fast and Secure Three-party Computation," in *Proceedings of the the 22nd ACM SIGSAC Conference*, pp. 591–602, Denver, Colorado, USA, October 2015.

[35] M. Naor, "Bit commitment using pseudorandomness," *Journal of Cryptology*, vol. 4, no. 2, pp. 151–158, 1991.

[36] H. Cheng, Z. Su, N. Xiong, and Y. Xiao, "Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model," *Information Sciences*, vol. 329, pp. 461–477, 2016.

[37] H. Cheng, N. Xiong, A. V. Vasilakos, L. Tianruo Yang, G. Chen, and X. Zhuang, "Nodes organization for channel assignment with topology preservation in multi-radio wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 5, pp. 760–773, 2012.

Research Article

A Heuristic Evolutionary Algorithm of UAV Path Planning

Zhangjie Fu [iD],[1,2] Jingnan Yu,[1] Guowu Xie,[3] Yiming Chen,[1] and Yuanhang Mao[1]

[1]*School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China*
[2]*College of Information Science and Technology, Jinan University, Guangzhou 510632, China*
[3]*Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA*

Correspondence should be addressed to Zhangjie Fu; wwwfzj@126.com

With the rapid development of the network and the informatization of society, how to improve the accuracy of information is an urgent problem to be solved. The existing method is to use an intelligent robot to carry sensors to collect data and transmit the data to the server in real time. Many intelligent robots have emerged in life; the UAV (unmanned aerial vehicle) is one of them. With the popularization of UAV applications, the security of UAV has also been exposed. In addition to some human factors, there is a major factor in the UAV's endurance. UAVs will face a problem of short battery life when performing flying missions. In order to solve this problem, the existing method is to plan the path of UAV flight. In order to find the optimal path for a UAV flight, we propose three cost functions: path security cost, length cost, and smoothness cost. The path security cost is used to determine whether the path is feasible; the length cost and smoothness cost of the path directly affect the cost of the energy consumption of the UAV flight. We proposed a heuristic evolutionary algorithm that designed several evolutionary operations: substitution operations, crossover operations, mutation operations, length operations, and smoothness operations. Through these operations to enhance our build path effect. Under the analysis of experimental results, we proved that our solution is feasible.

1. Introduction

With the rapid development of the network and the informatization of society, how to improve the accuracy of information is an urgent problem to be solved. The existing method is to use an intelligent robot to carry sensors to collect data and transmit the data to the server in real time [1]. Many intelligent robots have emerged in life; the UAV (unmanned aerial vehicle) is one of them. UAV is the unmanned aircraft operated by radio remote control devices and self-contained program; it is operated entirely or intermittently by on-board computers. Compared to the manned aircraft, it has many advantages, such as small size, low cost, easy to use, and portable camera or other sensors. UAVs are used in various tasks in industries such as commerce, agriculture, and military applications, including environmental monitoring, target identification, border patrols, and search and rescue assistance. However, UAVs also exposed a lot of problems, the most important issue is the endurance of UAVs. Energy scheduling problem has always been a thorny issue in wireless

network applications [2–4]; UAVs are the same. Due to the limited payload of UAVs, it is not possible to add more batteries to the UAVs. Only when the UAVs perform their tasks, can they fly as effective paths as possible. So path planning is the primary issue we want to solve when we study UAVs.

UAV path planning refers to the optimal path planning problem of UAVs. The main purpose of the optimization path is to find a safe flight path with minimum energy consumption on the premise of completing the UAV mission. That is to say, the essence of path planning is to find one in the workspace according to one or some optimization criteria (e.g., minimum working cost, shortest walking path, and shortest walking time). To find the optimal path for UAV, recently, a lot of work has been proposed [5, 6]. These proposed algorithms mainly are divided into two categories in terms of the algorithm form, including optimal algorithm and the heuristic algorithm. The optimal algorithms mainly include mathematical programming algorithm, parameter optimization algorithm and exhaustion

method, while the heuristic algorithms include A∗ algorithm, ant colony algorithm, genetic algorithm, and simulated annealing algorithm. There are many differences between two kinds of algorithms. First, the optimal algorithm will have high computational complexity when the scale is large, while the heuristic algorithm will have certain advantages in efficiency. Second, the heuristic algorithm exists in a random probability, while the optimal algorithm can solve this problem. Moreover, algorithms can be divided into coverage paths and optimal path planning algorithms according to the tasks performed by UAVs. Coverage path planning is mainly used in environmental monitoring and regional surveys. The main algorithms designed for environmental monitoring and regional surveys are raster scanning [7] and regional segmentation [8–13]. The purpose of the optimal path planning algorithm is to find the minimum flight cost path; it means that UAV fly in the effective path within the effective time. Both of these algorithms is aimed at minimizing the cost of UAV flight. Although existing algorithms have made some achievements on the UAV path planning, these methods only take the UAV's flight length or the number of turns of UAV flight paths (the number of UAV's corners) into consideration when referring to the UAV's flight cost. The classical evolutionary algorithm extends from one point to the surrounding adjacent points while traversing the path and cannot skip over the adjacent points.

In this paper, there are two contributions. (1) The first one is analysis of UAV flight energy consumption factors and feasibility of flight paths. We propose three cost functions: path security cost, path length cost, and path smoothness cost. The path security cost is used to determine whether the path is feasible; the length cost and smoothness cost of the path directly affect the cost of the energy consumption of the UAV's flight. (2) A heuristic evolutionary algorithm is proposed and several evolutionary operations are designed: substitution operation, cross operation, mutation operation, length operation, and smoothness operation, through these operations to enhance our build path effect.

The structure of this paper is as follows. In the second section, the current research status of intelligent robot path planning algorithms is introduced. Section 3 mainly introduces the construction of cost functions. Section 4 mainly introduces the design of path planning algorithms. Section 5 mainly introduces the experimental results. It mainly introduces some conclusions and further discussion.

2. Related Work

The core of path planning is the design of algorithms. Currently, path planning algorithms have attracted widespread attention. Whether they are global path planning or local path planning, the essence of the algorithm is to solve the travelling salesman problem [15]; the algorithms can be roughly classified into the following categories: traditional algorithms, heuristic algorithms, and intelligent bionic algorithms.

2.1. Traditional Path Planning Algorithms. The traditional path planning algorithms mainly include the visibility graph, artificial potential field, simulate anneal arithmetic method, and fuzzy logic algorithm.

Visibility Graph. The viability graph is to regard the robot as a point and connect the vertices of the robot, the target point, and the polygonal obstacle. These lines cannot cross the obstacle. So a visibility graph is formed [16]. Since the vertices of any two lines are visible, for the robot, all paths from the starting point along these lines to the target point are collision-free paths. Taking advantage of an optimized search algorithm to search for the visibility graph, a shortest path can eventually be found. The advantage is that the shortest path can be found. The disadvantage is the lack of flexibility. Once the starting point and target change, the visibility graph can be reconstructed.

Artificial Potential Field. The artificial potential field is a virtual force method proposed by Khatib [17]. The basic idea of this method is to consider the movement of robot in the environment as a movement of the robot in a virtual force field, where the obstacle generates a repulsive force on the robot and the target point generates a gravitational force on the robot. The combined force of gravity and repulsion acts as the robot control force to control the robot to avoid the obstacles and reach the target point. The advantage of the artificial potential field method is that it has good real-time performance and is convenient for robot bottom control. Nevertheless, the traditional artificial potential field method has the following disadvantages: local minimum points and target inaccessibility [18].

Simulate Anneal Arithmetic. Simulate anneal arithmetic is an effective approximation algorithm that is suitable for large-scale combinatorial optimization problems. It imitates the annealing process of solid materials. By setting the initial temperature, initial state, and cooling rate, the temperature is controlled to continuously decrease. Combining the probabilistic jump characteristics, make use of the domain of solution space for random search to avoid falling into a local optimum. It has the advantage of simple and efficient operation, etc. However, there are shortcomings such as slow convergence and randomness. The setting of its parameters is a key part to the realization of the algorithm [19].

Fuzzy Logic Algorithm. The fuzzy logic algorithm simulates the driver's driving experience and combines the physiological perception and action. According to the real-time sensor information of the system, the data elements are converted into fuzzy sets, and the output results are determined according to the membership function, and then the table information is obtained from the table [20]. The advantage of fuzzy logic algorithm is its robustness, which does not require the establishment of complex mathematical models, and avoids the disadvantages of mobile robots in other algorithms that are strongly dependent on the environment. The disadvantage is that it is more difficult to establish fuzzy rules [21].

2.2. Heuristic Algorithms. The heuristic algorithm has a strong path search capability and can be used in a discrete path topology. Common heuristic search algorithms include the A∗ algorithm, the Dijkstra algorithm and the Floyd algorithm.

(1) A∗ Algorithm. By setting appropriate heuristic functions, the A∗ algorithm comprehensively evaluates the value of each extended node and expands by comparing the nodes of each extended node to select the lowest cost node until the target point is found. The advantage of the A∗ algorithm is that there are few extension nodes and it has good real-time performance. The disadvantage is that the size of the robot itself is ignored in the actual motion process.

(2) Dijkstra's Algorithm. Dijkstra's algorithm is a typical shortest path algorithm. It extends from the starting point to the center layer until it reaches the end point and then obtains the shortest path by comparing the nodes' forward traversal [22]. The advantage is that the shortest path has a high success rate and good robustness. The disadvantage is that all nodes need to be traversed to obtain the shortest path. Compared to the A∗ algorithm, the efficiency is poor, and it is less effective for complex topological networks. In addition, the algorithm cannot handle the problem with negative edges.

(3) Floyd Algorithm. Floyd algorithm is an algorithm for finding the shortest path between vertices in a given weighted path topology network. It first converts the path network into a weight matrix and then finds the shortest distance between any two points in the weight matrix [23]. It has higher efficiency than Dijkstra's algorithm, but it also has disadvantages such as high time complexity and unsuitable for big data calculation.

2.3. Intelligent Bionic Path Planning Algorithms. Intelligent bionic algorithm is an algorithm discovered by people through bionics research in a series of natural phenomena, including ant colony optimization, particle swarm optimization algorithm, genetic algorithm, and neural network algorithm.

(1) Ant Colony Optimization. The ant colony optimization is an algorithm inspired by the phenomenon in which ants search for food. Each ant leaves a certain concentration of pheromone from the starting point to the foraging road. The pheromone will slowly evaporate with the passage of time, and the ants will use the pheromone concentration as the basis for path selection. The higher the pheromone concentration, the greater the probability that the path will be selected. As time goes on, the higher the concentration of pheromone left by the shorter path due to more ants' traversing, the higher the probability of ant selection on the higher concentration path and the more pheromones would be left. The shortest path will be discovered quickly after repeated iterations, which achieves the goal of path planning. Ant colony algorithm is essentially a parallel algorithm that is easy to implement by computers and has good global optimization ability. However, as the environment expands, the computational volume will increase exponentially, and it is easy to fall into the local optimum [24–26].

(2) Particle Swarm Optimization. Particle Swarm Optimization (PSO) is also an iterative algorithm. It simulates bird's flight predation behavior and makes use of the sharing of information among individuals in a group to make the movement of the entire population produce an evolutionary process from disorder to order in the problem solving space, thereby obtaining an optimal solution. The system is initialized as a set of random solutions. The individual updates the position through the individual historical best solution and the global history optimal solution makes it follow the optimal particle and iteratively searches for the optimal value. Compared with genetic algorithms, PSO does not have many parameters that need to be adjusted. The algorithm has the advantages of simplicity, easy implementation, good robustness, and fast convergence, but it is easy to fall into the local optimal solution [27].

(3) Genetic Algorithm. Genetic algorithm is a computational model that simulates the natural selection of Darwin's biological evolution theory and the genetic evolution process of genetic mechanisms. Starting from a population that may have potential solution set, the population is first coded, and the adaptive value of each individual is calculated. According to the principle of survivability of the fittest, selection, crossover, and mutation are performed, the poorer individuals are eliminated, and the remaining individuals are repeatedly iterated again until the best individual is produced. The greatest advantage of the genetic algorithm is that it is easy to combine with other algorithms and give full play to its own iterative advantages. The disadvantage is that because there is no feedback information, the computational efficiency will be greatly reduced after the later period [28].

(4) Neural Network Algorithm. Artificial neural network system appeared after the 1940s. It is made up of a number of connective weights with adjustable neurons. It has features such as large-scale parallel processing, distributed information storage, and good self-organizing self-learning capabilities [29]. The neural network algorithm is used in path planning as follows: the neural network is used to describe the environmental constraints and the collision energy function is calculated. The sum of the collision energy function and the distance function is used as the optimization objective function, and the point set is determined by optimizing the extreme value of the objective function. The equation of motion ultimately makes the iterative path point set tend to be the optimal planning path. Although neural network has excellent learning ability, its poor generalization ability is its fatal weakness, but because of its strong learning ability and good robustness, its combination with other algorithms has become a research hotspot in the field of path planning [30].

3. The Cost Function

In some classical optimization algorithms, the length of the path represents the energy consumption of UAV flight.

Parameter	
$Cost_s$	the security cost of the path
$Cost_l$	the length cost of the path
$Cost_a$	the smoothness cost of the path
Cost	the effective cost of the path
l_i	the length of the i th obstacle covering the path
L	the total length of the path
m	the number of obstacles on the path
n	the number of points traversed by the path
(x_n, y_n)	the coordinate of the nth traversal point
θ_n	the rotation angle of the path at the nth traversal point
W_l	the energy consumption of UAV flying in a straight line
W_a	the energy consumption of drones flying at corners
ω_1, ω_2	two weights

Thinking about the problem of energy consumption and the feasibility of the path of UAV flight, this paper proposes three costs. The cost of path security is to determine whether the path is feasible. The length cost of the path and the smoothness cost of the path represent the energy cost of the path flight of the UAV. Table 1 shows the meaning of the parameters in our paper.

3.1. Path Security. During the flight of a UAV, the most important thing is not to collide with obstacles in the environment. The purpose of the security of the route is to avoid obstacles on the route. Therefore, the number of obstacles on a path is the primary determinant of the feasibility of the path. We define the degree of security of the path to be calculated by the occupancy of obstacles that have passed through the UAV flight path. Equation (1) shows the calculation of the path security $Cost_s$.

$$Cost_s = \sum_{i=1}^{m} \frac{l_i}{L} \tag{1}$$

$Cost_s$ indicates the security cost above the path, m is the number of obstacles on the path, and l_i is the length of the ith obstacle covering the path. L is the total length of the path.

We ultimately find that the theoretical number of path obstructions should be 0. This cost is used as a factor to determine the feasibility of the path.

3.2. The Length of the Path. The length of the path refers to the path from the starting point to the end point of the UAV. The length of the path directly relates to the UAV flight time, which directly affects the energy consumption of the UAV. Therefore, the shorter the UAV's flight path means less energy consumed by UAV flight. In mathematics, we mainly calculate the Euclidean metric distance of the flight path of each UAV. The length cost of the path is shown in

$$Cost_l = \sum_{n=1}^{n-1} \sqrt{(x_{n+1} - x_n)^2 + (y_{n+1} - y_n)^2} \tag{2}$$

$Cost_l$ is the length of the path of which UAV go through the threat area and $Cost_l$ represents the length cost of the path, which is the most important factor affecting the flight energy consumption of the UAV. We can calculate $Cost_l$ by point coordinates on the UAV's flight path. n is the number of points traversed by the flight path and (x_n, y_n) is the coordinate of the nth traversal point.

3.3. The Smoothness of the Path. The principle of the corner of a multirotor UAV is to obtain a body rotation force by changing the speed of the adjacent motor by a UAV. Therefore, the angle of rotation is also a factor that determines energy consumption. Considering this situation, we define a cost function for path smoothness. Equation (3) shows the smoothness cost $Cost_a$.

$$Cost_a = \begin{cases} 0, & n \leq 2 \\ \sum_{n=2}^{n-1} \theta_n, & n > 2 \end{cases} \tag{3}$$

$Cost_a$ represents the smoothness cost of the UAV's flight path. At a fixed angular speed, $Cost_a$ also affects the UAV's flight energy consumption, which is the angle number of the UAV's path rotation. θ_n is the rotation angle of the path at the nth traversal point, which can be calculated by the law of cosines.

3.4. Effective Cost Function. When searching for the best flight path of UAV, we must consider the three cost functions above. Since the three cost function units and the meaning of inclusion are different, the sum of the three costs cannot be calculated directly. Therefore, we set the weight based on the weight of each cost function. Since both the length of the path and the smoothness of the path can affect the energy consumption of the UAV flight, we can determine the weight by the energy loss value of the UAV between the length of the path and the smoothness. Jalil Modares et al. [31] verified that the energy consumption of UAV and the path distance and number of angles are approximately linear. We use them to establish the energy model to determine the weight values of the length cost and the angle cost. Energy equation (4) shows the relationship between UAV energy W_l and flight distance:

$$W_l = \omega_1 Cost_l \tag{4}$$

Equation (5) shows the relationship between the UAV energy W_a and the angle of rotation:

$$W_a = \omega_2 Cost_a \tag{5}$$

The energy cost of the flight path is shown in

$$Cost = \omega_1 Cost_l + \omega_2 Cost_a \tag{6}$$

where Cost represents an effective cost that directly affect the energy consumption of the UAV and is directly used as an important factor in determining the fitness function of the path in the algorithm and the value method of ω_1 and ω_2 is shown in experiment.

4. Algorithm Design

Similar to evolutionary algorithms, the methods our proposed include initializing the population, calculating the individual's cost fitness, evolving operations, and jumping out of the loop. The optimization mode proposed by this paper is different from the classical optimization algorithm. Our method is to minimize the energy consumption of the path and proposes several evolutionary operations: replacing operation, crossover operation, length operation, and smoothness operation. Our method retains the infeasible path, which improves the diversity of paths, so that it is possible to jump out of the local optimization. The classical evolutionary algorithm extends from one point to the surrounding adjacent points while traversing the path and cannot skip over the adjacent points. When traversing a path, our method can skip some adjacent traversal points, which can effectively reduce the length and smoothness of the path. The algorithm flow is shown in Figure 1.

Algorithm 1 is the pseudocode of this algorithm. Step 1 is initialization population; we randomly generate a set of solutions. Steps 2-4 are a replacement operation. During the initialization and evolution of classical optimization algorithms, the solution generated by our method may be an infeasible solution. Therefore, we propose a replacement operation. The replacement operation is mainly to turn the infeasible solutions into feasible solutions. Steps 6-13 are to find the optimal solution by the evolutionary operations. The paper describes in detail the evolutionary operation in the later. Our termination condition is the number of iterations.

4.1. Replacing Operation. During the initialization and evolution of classical optimization algorithms, their paths are feasible paths. Such operations easily fall into local optimization. In our method, the infeasible path is retained, which improves the diversity of paths, so that it is possible to jump out of the local optimization. A replacing operation is proposed to deal with an infeasible path in our method. The replacement operation is mainly to turn the infeasible path into a feasible path and determine the feasibility of the path through the path security cost $Cost_s$ (see (1)). When the path security cost is 0, the path is feasible; otherwise the path is not feasible. When the path is not feasible, finding a new feasible path replaces the infeasible path. The specific steps are as follows: First, the collision path segmentation is determined. Then the distance of the obstacle vertex to the segmented path is calculated, respectively. The point of the maximum distance between the two sides of the piecewise path is taken as the candidate for in terms of passing the obstacle and connecting the endpoints of the segmentation path and the candidate points, respectively, to generate two new paths. Calculating the effective cost Cost (see (5)) of the two paths, the path with small energy consumption cost is selected to replace the original collision. As shown in Figure 2, when the segmented path SE is infeasible, the two points P, Q are found that are farthest from the obstacle vertex on both sides of the collision path SE to the segmented path. The infeasible path SE is replaced by two feasible paths SPE and SQE, and then the effective costs of SPE and SQE are calculated, respectively.

1	Initialization population
2	Calculating the security cost of each individual $Cost_s$
3	If $Cost_s \neq 0$
4	Replacing operation
5	Updating population, go to step 2
6	If $Cost_s = 0$
7	Calculating effective cost of each individual $Cost$
8	If has not terminated
9	Evolutionary operations
10	Updating population, go to step 2
11	If termination
12	Output optimal path
13	End

Algorithm 1

Last, the minimum energy cost path SPE is chosen to take the replace of the path SE.

4.2. Crossover Operation. The crossover operation is to generate a new path through cross replacement. When there are two or more cross points in two paths, two cross points are randomly selected, and the path between the two selected cross points is exchanged to generate new paths. The effective cost of four paths is calculated, and the optimal effective cost path is chosen. Figure 3 is the result of the crossover operation. As shown in Figure 3(a), when the initial split path L1 and L2 have more than two intersections, Figure 3(b) shows that the crossover produces two new paths L3 and L4.

4.3. Mutation Operation. Mutation refers to replacing certain gene values in individual code strings with other gene values based on the probability of mutation to form a new individual. In this paper, the feasibility of the path will change whenever the coordinate node of the path mutates. In order to solve this problem, two different probability mutation operations are proposed to deal with the feasible path and the infeasible path respectively. When the path is feasible, the coordinates of the path node are mutated in a small range with a small probability, and the variation path is still feasible; when the path is not feasible, the coordinates of the path node are mutated in a large range with a large probability. And ensure better individual costs after the mutation. When the path is not feasible, we change the path to a feasible path by replacing the operation.

4.4. Length Operation. The purpose of the path length operation is to reduce the distance of the drone flight path. When the path is feasible, some paths usually contain additional path segments. Considering the length cost of the path and the feasibility of the path, if the path is still feasible after randomly deleting one or more path nodes, the path is deleted from the original path node and a new path is generated. Figure 4 shows an example of a path length operation. As shown in Figure 4(a), ABCD is the initial path. As shown in
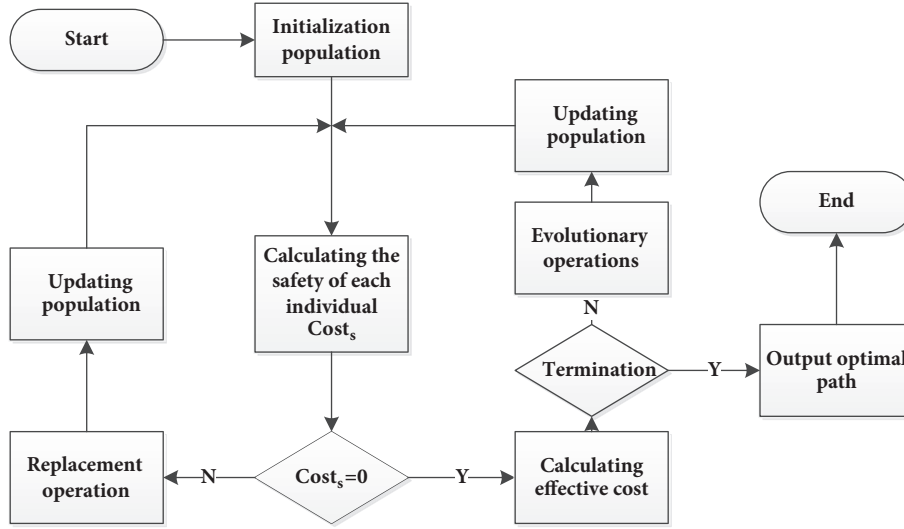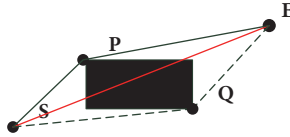
FIGURE 1: Flowchart of our method.

FIGURE 2: An example of an infeasible path replacing operation.

TABLE 2: Specific parameters of the system platform equipment.

Equipment	
UAV	DJI GO M100
Onboard computer	DJI GO Manifold
Ground station	Iphone7 plus+ Lenovo ThinkPad x260
Power sensors	ADS1115 IC

Figure 4(b), if the point B is randomly deleted in the path and the path is still feasible, a new path ACD is generated.

4.5. Smoothness Operation. The purpose of the path smoothness operation is to reduce the number of turns of the UAV and randomly select two adjacent straight-line paths, which are used as input. And then the angle between them is changed. In two adjacent straight paths, the midpoint of a straight path is randomly selected to replace the intersection of the two straight paths, and a new path is generated. Figure 5 shows an example of path smoothness operation. ABC is an adjacent straight-line path, and ADC is a new path through path smoothing.

5. Experiment

5.1. Experimental Model. In the experiment, according to the wireless network data acquisition model proposed by Zheng et al. [32], we built a system platform, as shown in Figure 6. The system platform includes the UAV, on-board computers manifold, ground station, and power sensors. Table 2 shows the specific parameters of the system platform equipment. The main use of UAV in this system platform is DJI GO M100, because DJI GO M100 supports secondary development, which is conducive to the monitoring of data. The use of on-board computer in this system platform is DJI GO manifold. The system of manifold is Linux; the system has Mobile SDK and on-board SDK development interfaces. The system platform uses manifold and data acquisition sensors to build

a detection module. The system platform uses the manifold to send the detected data to the ground station. Through our system platform, the UAV's own data and other inspection data can be transmitted to the ground station in real time, and the flight status of the UAV can be effectively observed in real time.

5.2. Setting of Weights. In this experiment, we analyzed the relationship between flight path distance and rotation angle with the energy consumption of UAV based on the energy model proposed by Jalil Modares et al. [31]. We determine the value of ω_1 and ω_2 by analyzing their relationship. Note that the values ω_1 and ω_2 detected by different UAVs are different. The UAV used in this experiment is the DJI GO M100.

In this experiment, through analyzing the relationship between path distance and energy consumption, the fly speed of UAV is set as 10m/s to conduct 20 times flights in the line distance of 50m, 100m, and 150m, respectively. To study the relations between rotation angle and energy consumption, the UAV is allowed to fly 20 times in the following four situations: (1) straight flight distance 100m, (2) straight distance 50m, 45 degree angle, and straight flight 50m, (3) straight distance 50m, 90 degree rotation, and 50m Straight flight, (4) straight distance 50m, degree angle 45, and straight flight 50m. To analyze the relationship between the angle number and the energy consumption, the measured value is subtracted from the average test value of the line distance. Figure 7 shows the relationship between the drone
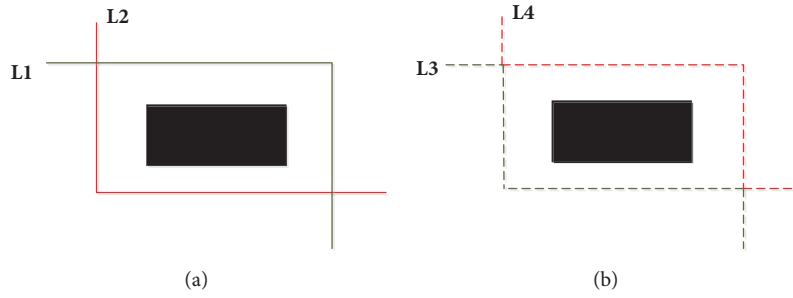
FIGURE 3: An example of a crossover operation.
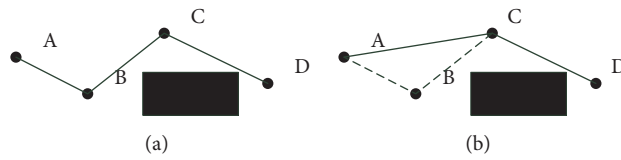
FIGURE 4: An example of a path length calculation.
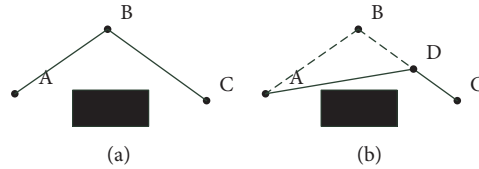
FIGURE 5: An example of a path length calculation.
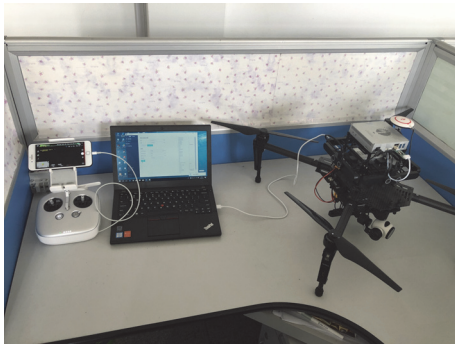
FIGURE 6: System model.
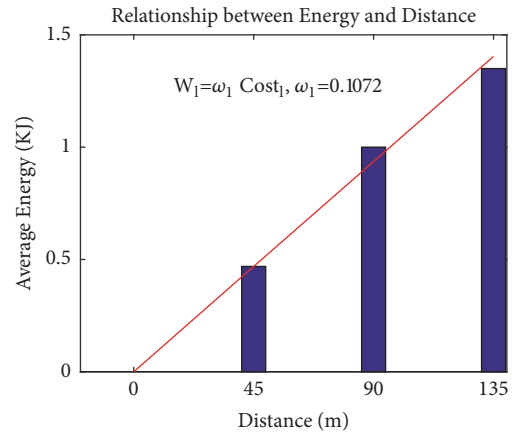
FIGURE 7: Relationship between energy and distance.

energy loss and the length of the flight path; Figure 8 shows the relationship between energy loss and rotational angle of the UAV. Through the experiment, we get the value of ω_1 and ω_2, $\omega_1 = 0.1072$ and $\omega_2 = 0.0104$.

5.3. Experimental Results. This section mainly discusses the effect of our proposed method. The result of our method is compared with other algorithms. To assess the feasibility of our program, we use matlab2014a for simulation experiments. The implementation platform of us is Windows 7 server with Core CPU 2.85GHz.

The method proposed in this paper is based on an evolutionary algorithm of genetic algorithm, so we choose

several classical evolutionary algorithms as the comparison algorithm. We have done some work comparisons in our previous works [33]. In this paper, we make different comparisons between the efficiency of the algorithm and the results generated by the algorithm. We make different comparisons between the efficiency of the algorithm and the results generated by the algorithm. We use the same modeling environment for comparison experiments. I set the lower left corner as the starting point and the upper right corner as the end point to generate an optimal path; the grid ratio is 1:10; the

TABLE 3: Comparison of algorithms.

Algorithm	Time (s)	Path length (10m)	Angle (degree)	Theoretical Cost (KJ)	Actual Cost (KJ)
GA[14]	137.76	13.899	-	14.900	19.910
PSO	201.46	13.899	-	14900	19.816
Our method	181.25	13.291	71.565	14.992	16.071

TABLE 4: The result of using the path length as a cost function.

Algorithm	Time (s)	Path length (10m)	Theoretical Cost (KJ)	Actual Cost (KJ)
GA	137.76	13.899	14.900	19.910
Our method	120.25	13.291	14.248	16.071

TABLE 5: The result of using our proposed cost as a cost function.

Algorithm	Time (s)	Path length (10m)	Angle (degree)	Theoretical Cost (KJ)	Actual Cost (KJ)
GA	210.46	13.899	135	16.304	18.673
Our method	181.25	13.291	71.565	14.992	16.071
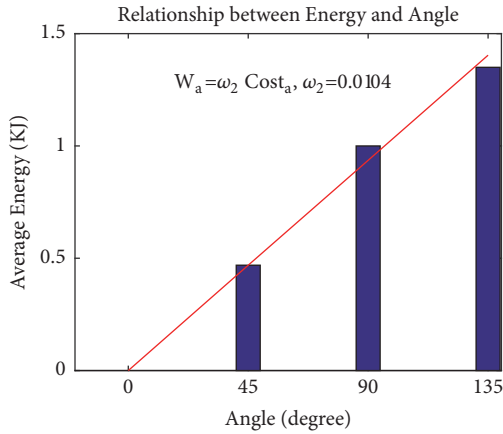
FIGURE 8: Relationship between energy and angle.

number of iterations is set to 500 times. Figure 9(a) shows the path generated by the classical GA [14]. Figure 9(b) shows the path effect generated by PSO. Figure 9(c) is the result of the algorithm generation path proposed by this paper.

In Table 3, GA and PSO methods use the path length as a cost function. Our method uses the path security, smoothness, and length as cost functions. Time represents the time of each method run. The path length is the length of the path generated by each method. The angle is the angle of the path generated by each method. The theoretical cost of our method is given by (6). The theoretical costs of GA and PSO are given by (4). The actual cost is to measure energy consumption through our system platform of the path generated by each method. Because GA and PSO methods use the path length as a cost function, we only discuss their run time and the length of the path. According to the data in Table 3, the running time of GA is better than that of PSO. This conclusion has been verified by Roberge V et al. [34]. In aspect of the running time of the algorithm, because our method needs to calculate the path smoothness cost, the time

cost of our method is larger. However, the time overhead of our method is better than PSO in this case. According to the data in Table 3, our method is superior to PSO in other aspects. Compared with GA, although our method has a large time cost, the path length generated by our method is dominant.

Because the two methods use different cost functions, we cannot judge whether the two schemes are good or bad. So we set up two sets of comparative experiments. One set of experiments used the length of the path as a cost function, and the other set used our proposed cost as a cost function. Figure 10 shows the result of using the path length as a cost function. Figure 10(a) shows the path generated by the classical GA. Figure 10(b) is the result of the algorithm generation path proposed by this paper. Table 4 shows the parameters of the results in Figure 10. Figure 11 shows the result of using our proposed cost as a cost function. Figure 11(a) shows the path generated by the classical GA. Figure 11(b) is the result of the algorithm generation path proposed by this paper. Table 5 shows the parameters of the results in Figure 11.

In Table 4, the theoretical cost is the effective cost (see (4)) of the path generated by each method. According to the data in Table 4, two methods use the path length as a cost function. In aspect of the running time of the algorithm, our method is better than GA. In aspect of path length our method is shorter than GA. In Table 5, the theoretical cost is the effective cost (see (6)) of the path generated by each method. According to the data in Table 5, two methods used our proposed cost as a cost function. In aspect of the running time of the algorithm, our method is better than GA. In aspect of path length our method is shorter than GA. According to the data in Tables 4 and 5, we can know that our method is better than GA in the case of using the same cost function.

The theoretical cost is the effective cost (see (6)) of the path generated by each method. We designed a set of experiments. From the above experiments, we can know that the path effects generated by our method using different cost functions are the same in our envisaged environment. Note
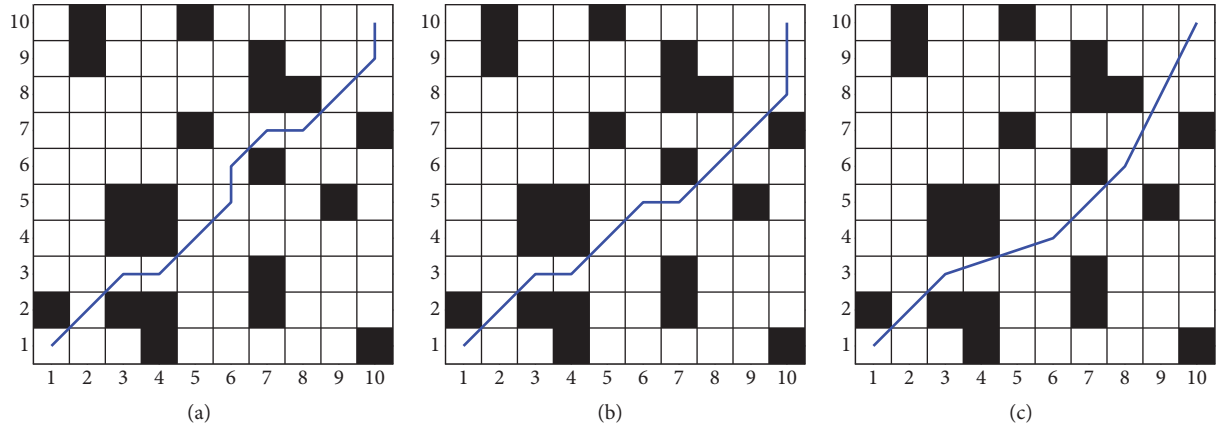
FIGURE 9: Three algorithms running path results.
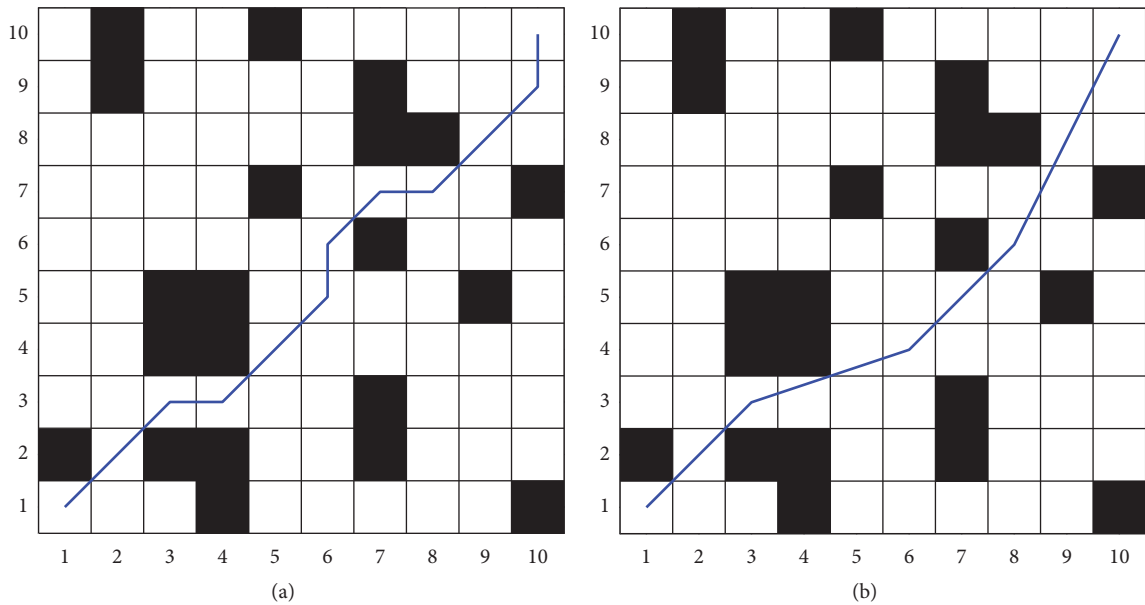
FIGURE 10: The result of using the path length as a cost function.

that the path effects generated by our method using different cost functions are the different in same environment. The experimental results of our experiments are the same because our environment is simple. We use the path generated by our method to actually fly 20 times. Record the energy consumption of each flight. Figure 12 is the result of our experiment. In Figure 12, the x-axis is the number of experiments and the y-axis is the energy consumption of the drone. Each point is the actual energy consumption of each flight. The red line is the average of the actual energy consumption. The blue line is the theoretical value that uses our cost function to calculate energy consumption. The green line is the theoretical energy consumption calculated using length as a function of cost. From the data in Figure 12, we can see that our cost function is close to the actual value. The reason for the deviation of our cost function may be that wind and other factors affect the energy consumption of the drone in the actual flight

environment. In future work, we can continue to study other factors that affect the energy consumption of drones, we continue to refine the design of the cost function.

6. Conclusion and Future Work

This paper analyzes the factors affecting the flight path of UAVs and proposes three cost functions: path security cost, length cost, and smoothness cost. The path security cost is used to determine whether the path is feasible; the length cost and smoothness cost of the path directly affect the cost of the energy consumption of the UAV's flight. Because of the weighting of the two costs, we set the weights for the two costs through real experimental data. In order to improve the effect of generating effective paths, we propose a heuristic evolutionary algorithm that sets several evolutionary operations: substitution operation,
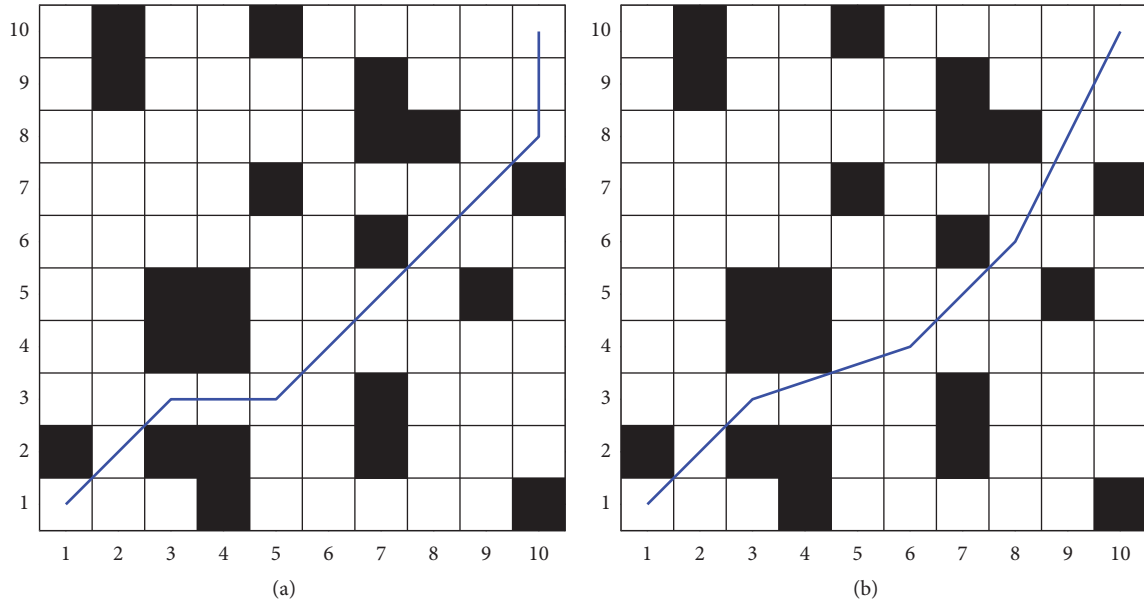
FIGURE 11: The result of using our proposed cost as a cost function.
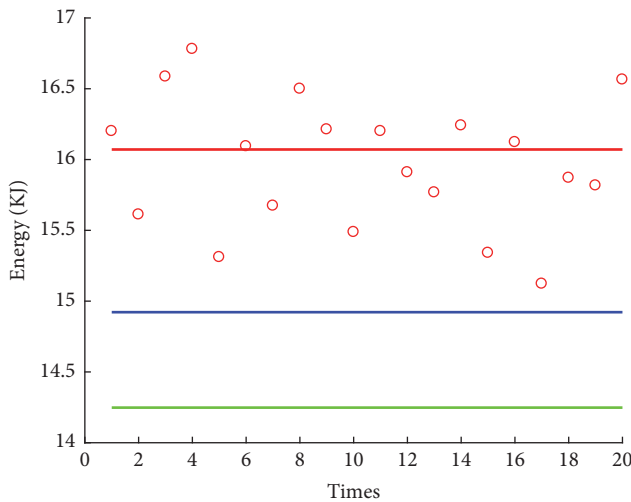
FIGURE 12: The actual energy consumption of the path.

cross operation, mutation operation, length operation, and smoothness operation. Through our combination of real experiments and simulation experiments, we verified that our proposed method is feasible; compared with classical algorithms, our method is better at generating the optimal path.

The three cost functions proposed in this paper influence each other. We must find the best of the three cost functions, that is, a multiobjective optimization problem. The solution proposed in this paper is to convert multiobjective optimization into single-objective optimization by setting weights. Our next work will use a multiobjective optimization algorithm to solve this problem.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. H. Cui, Y. Y. Zhang, and Z. P. Cai, "Securing Display Path for Security-Sensitive Applications on Mobile Devices Computers," *Materials & Continua*, vol. 55, no. 1, pp. 17–35, 2018.

[2] H. Cheng, Z. Su, N. Xiong et al., "Energy-Efficient Node Scheduling Algorithms for Wireless Sensor Networks," *Information Sciences*, vol. 329, no. 2, pp. 461–477, 2016.

[3] H. Cheng, N. Xiong, A. V. Vasilakos, L. Tianruo Yang, G. Chen, and X. Zhuang, "Nodes organization for channel assignment with topology preservation in multi-radio wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 5, pp. 760–773, 2012.

[4] A. Vinel, W.-S. E. Chen, N. N. Xiong, S. Rho, N. Chilamkurti, and A. V. Vasilakos, "Enabling wireless communication and networking technologies for the internet of things [Guest editorial]," *IEEE Wireless Communications Magazine*, vol. 23, no. 5, pp. 8-9, 2016.

[5] B. M. Sathyaraj, L. C. Jain, A. Finn, and S. Drake, "Multiple UAVs path planning algorithms: a comparative study," *Fuzzy Optimization and Decision Making. A Journal of Modeling and Computation Under Uncertainty*, vol. 7, no. 3, pp. 257–267, 2008.

[6] R. Samar and W. A. Kamal, "Optimal Path Computation for Autonomous Aerial Vehicles," *Cognitive Computation*, vol. 4, no. 4, pp. 515–525, 2012.

[7] A. Ahmadzadeh, J. Keller, G. Pappas et al., "An optimization-based approach to time-critical cooperative surveillance and coverage with UAVs," *Experimental Robotics. Springer Berlin/Heidelberg*, pp. 491–500, 2008.

[8] A. Xu, C. Viriyasuthee, and I. Rekleitis, "Efficient complete coverage of a known arbitrary environment with applications to aerial operations," *Autonomous Robots*, vol. 36, no. 4, pp. 365–381, 2014.

[9] E. Santamaria, F. Segor, I. Tchouchenkov et al., "Path planning for rapid aerial mapping with unmanned aircraft systems," in *Proceedings of the Eighth International Conference on Systems*, 2013.

[10] W. H. Huang, "Optimal line-sweep-based decompositions for coverage algorithms," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 27–32, May 2001.

[11] R. Mannadiar and I. Rekleitis, "Optimal coverage of a known arbitrary environment," in *Proceedings of the 2010 IEEE International Conference on Robotics and Automation, ICRA 2010*, pp. 5525–5530, USA, May 2010.

[12] I. Maza and A. Ollero, "Multiple UAV cooperative searching operation using polygon area decomposition and efficient coverage algorithms," *Distributed Autonomous Robotic Systems 6. Springer, Tokyo*, vol. 6, pp. 221–230, 2007.

[13] H. Choset, "Coverage of known spaces: The boustrophedon cellular decomposition," *Autonomous Robots*, vol. 9, no. 3, pp. 247–253, 2000.

[14] W. Zhou, Z. Yi, and Y. Ruimin, "Mobile robot path planning based on genetic algorithm," *Microcomputer Information*, vol. 24, no. 26, pp. 187–189, 2008.

[15] N. Xiong, W. Wu, and C. Wu, "An Improved Routing Optimization Algorithm Based on Travelling Salesman Problem for Social Networks," *Sustainability* , vol. 9, no. 6, p. 985, 2017.

[16] C. Chen, J. Tang, and Z. Jin, "A Path Planning Algorithm for Seeing Eye Robots Based on V-Graph," *Mechanical Science and Technology for Aerospace Engineering*, 2014.

[17] O. Khatib, "Real-time obstacle avoida nce for manipulators and mobile robots," *International Journal of Robotics Research*, vol. 5, no. 1, pp. 90–98, 1986.

[18] P. Vadakkepat, K. C. Tan, and W. Ming-Liang, "Evolutionary artificial potential fields and their application in real time robot path planning," in *Proceedings of the Congress on Evolutionary Computation (CEC 00)*, vol. 1, pp. 256–263, July 2000.

[19] A. E.-S. Ezugwu, A. O. Adewumi, and M. E. Frîncu, "Simulated annealing based symbiotic organisms search optimization algorithm for traveling salesman problem," *Expert Systems with Applications*, vol. 77, pp. 189–210, 2017.

[20] D. Adhikari, E. Kim, and H. Reza, "A fuzzy adaptive differential evolution for multi-objective 3D UAV path optimization," in *Proceedings of the 2017 IEEE Congress on Evolutionary Computation, CEC 2017*, pp. 2258–2265, Spain, June 2017.

[21] O. Castillo, H. Neyoy, J. Soria, M. García, and F. Valdez, "Dynamic Fuzzy Logic Parameter Tuning for ACO and Its Application in the Fuzzy Logic Control of an Autonomous Mobile Robot," *International Journal of Advanced Robotic Systems*, vol. 10, no. 1, 2013.

[22] S. Wang X, X. Wu Z, and X. Wang S XWu Z, "Improved Dijkstra shortest path algorithm and its application," *Computer Science*, vol. 39, no. 5, pp. 223-222, 2012.

[23] Z. He and L. Zhao, "The comparison of four UAV path planning algorithms based on geometry search algorithm," in *Proceedings of the 9th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2017*, pp. 33–36, China, August 2017.

[24] X. F Wan, W. Hu, W. F. Fang et al., "Research on path planning of robot based on improved ant colony algorithm," *Computer Engineering and Applications*, vol. 50, no. 18, pp. 63–66, 2014.

[25] X. Chen, Y. Kong, X. Fang, and Q. Wu, "A fast two-stage ACO algorithm for robotic path planning," *Neural Computing and Applications*, vol. 22, no. 2, pp. 313–319, 2013.

[26] T. Zhao, X. Pan, and Q. He, "Application of dynamic ant colony algorithm in route planning for UAV," in *Proceedings of the 7th International Conference on Information Science and Technology, ICIST 2017*, pp. 433–437, Viet Nam, April 2017.

[27] M. D. Phung, C. H. Quach, T. H. Dinh, and Q. Ha, "Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection," *Automation in Construction*, vol. 81, pp. 25–33, 2017.

[28] G. Ji, "A survey of Genetic Algorithm," *Computer Applications and Software*, vol. 21, no. 2, pp. 69–73, 2004.

[29] S. A. Yildizel and A. U. Öztürk, "A study on the estimation of prefabricated glass fiber reinforced concrete panel strength values with an artificial neural network model," *Computers Materials & Continua*, vol. 52, no. 1, pp. 41–52, 2016.

[30] W. Wang, S. M. Wei, Y. Q. Yang, Y. F. Jiang, and L. I. DuanLing, "Path planning for a mobile robot using neural networks," *Journal of Beijing University of Technology*, vol. 45, no. 10, pp. 221–225, 2009.

[31] J. Modares, F. Ghanei, N. Mastronarde, and K. Dantu, "UB-ANC planner: Energy efficient coverage path planning with multiple drones," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation, ICRA 2017*, pp. 6182–6189, Singapore, June 2017.

[32] H. Zheng, W. Guo, and N. Xiong, "A Kernel-Based Compressive Sensing Approach for Mobile Data Gathering in Wireless Sensor Network Systems," *IEEE Transactions on Systems Man & Cybernetics Systems*, 2017.

[33] J. Yu, Y. Chen, Y. Mao et al., "A Heuristic Evolutionary Algorithm of UAV Path Planning," in *Proceedings of the International Conference on Cloud Computing and Security (ICCCS 2018)*, 2018, http://www.icccsconf.org/%E6%8E%A8%E8%8D%90SCI%E6%9C%9F%E5%88%8A%E8%AE%BA%E6%96%87%E5%88%97%E8%A1%A8.pdf.

[34] V. Roberge, M. Tarbouchi, and G. Labonte, "Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 132–141, 2013.

Research Article

CS-PLM: Compressive Sensing Data Gathering Algorithm Based on Packet Loss Matching in Sensor Networks

Zeyu Sun [1,2] **Rong Tao** [1] **Naixue Xiong** [3] **and Xiaoyan Pan** [1]

[1]*School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, China*
[2]*Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China*
[3]*Department of Mathematics and Computer Science, Northeastern State University, USA*

Correspondence should be addressed to Rong Tao; taorong_123456@163.com

The data transmission process in Wireless Sensor Networks (WSNs) often experiences errors and packet losses due to the environmental interference. In order to address this problem, we propose a Compressive Sensing data gathering algorithm based on Packet Loss Matching (CS-PLM). It is proven that, under tree routing, the packet loss on communication links would severely undermine the data reconstruction accuracy in Compressive Sensing (CS) based data gathering process. It is further pointed out that the packet loss in CS based data gathering exhibits the correlation effect. Meanwhile, we design a sparse observation matrix based on packet loss matching and verify that the designed matrix satisfies the Restricted Isometry Property (RIP) with a probability arbitrarily close to 1. Therefore, reliable transmission of the compressed data can be guaranteed by adopting the multipath backup routing among CS nodes. It is shown in the simulation results that, with a 60% packet loss ratio of the link, the CS-PLM algorithm can still ensure the effective reconstruction of the data gathered by the CS algorithm and the relative reconstruction error is lower than 5%. Therefore, it is verified that the proposed algorithm could effectively alleviate the sensitivity to packet losses for the CS based data gathering algorithm on unreliable links.

1. Introduction

The nodes in the wireless sensor networks (WSNs) are usually densely deployed and a lot of redundancy exists in the data gathered, which leads to the waste of the energy of the nodes. The compressive sensing (CS) algorithm is a new technique which could largely reduce the sampling frequency and execute the sampling process in parallel with the compression process. As a result, this technique has drawn much attention by researchers.

In order to balance and reduce the energy consumption of the nodes as well as prolong the network lifetime, researchers have proposed data gathering algorithms based on compressed sensing. At present, most of these algorithms are focused on how to effectively reduce network energy consumption and extend the network lifetime [1–3]. For example, it was proposed in paper [4] to employ sparse measurement matrices to reduce the communication cost of each measurement. The spatial and temporal correlation of the sensing data

was exploited in [5] to improve the compression ratio and further reduce the number of measurements. A multilevel hierarchical clustering topology was employed in [6] to gather the data in the network as well as reduce the number of sent and received packets at each layer of nodes. As a result, the total number of transmitted packets is reduced in the entire network. It was pointed out in [7] that the block diagonal measurement matrix could guarantee the reconstruction accuracy with a smaller number of measurements and a longer network lifetime. In recent years, with the progress in the theory of CS, researchers have started to investigate CS based data gathering algorithms for practical applications. The fact was considered in [8] that the data sparsity in realistic data sets would vary with time and space. Therefore, it was proposed to employ the autoregressive AR model to predict data changes and adaptively adjust the number of measurements to achieve the optimal reconstruction performance. It was pointed out in [9] that the environmental noise of the wireless links imposes prominent influences on the transmission of

the undersampled CS data in the network. An approximate gradient descending algorithm was therefore proposed to reconstruct the compressed data under the influence of noise. For the TF-packet loss problem of wireless links in practical application scenarios, there are relatively few works studying CS based data gathering algorithms. Due to the dynamic and asymmetry of wireless links, channel interference, improper antenna direction and height, etc., unreliable links are often the key issue faced by data gathering algorithms in practical sensor networks [10].

For the CS based data gathering problem on unreliable links, we propose a compressive sensing data gathering algorithm based on packet loss matching (CS-PLM). In this algorithm, the nodes in the network are divided into two types, i.e., the traditional forwarding (TF) nodes and the compressive sensing (CS) nodes. The packet loss of TF nodes does not exhibit the correlation while the lost packets of CS nodes are strongly correlated. In the process of the Compressive Sensing data gathering, a packet loss will lead to the loss of the data gathered from multiple nodes; the packet loss correlation effect is caused by the superimposed transmission of the collected data from each node of the multihop link in the CS compression sampling process. The closer the packet loss node S_k is to Sink, the greater the effect of packet loss is. In particular, if the packet of the next-hop neighbor node of Sink is lost, the correlation effect will result in the loss of data collected by some nodes of the whole network. While TF nodes only relay data in the traditional way of data gathering, packet loss has no correlation. As a result, we design a sparse measurement matrix based on the packet loss matching to recover the gathered data if packet losses occur at TF nodes. Therefore, the recovery problem for lost packets is transformed into the sparse matching sampling process in CS. However, for the lost packets at CS nodes, we design the multipath backup transmission scheme to guarantee the reliable data transmission and avoid the correlation effect for packet loss. Therefore, the impacts of unreliable links are alleviated for the CS based data gathering process and the reconstruction accuracy is guaranteed.

The main contributions of this paper are as follows.

①By analyzing the routing with tree structure, it is pointed out that the packet loss would seriously undermine the reconstruction accuracy of the CS based data gathering process and the packet loss in the CS based data gathering process exhibits the correlation effect.

②We design a SPLM measurement matrix and further prove that this matrix satisfies the Restricted Isometry Property (RIP) with a probability arbitrarily close to 1.

③We propose a multipath backup routing transmission scheme based on Hybrid CS to guarantee the reliable handover of the CS projection data.

2. Related Works

Due to its feature of simple encoding and complex decoding, compressed sensing theory has been widely applied to the area of data collection in WSNs. At present, the research of CS based data collection algorithms in wireless sensor networks mainly focuses on how to use CS technology to reduce the network energy consumption of the data gathering process in WSNs. Most of these works assume that the network link is an ideal link where the impact of packet loss on the CS based data gathering process is ignored. The number of transmitted data packets was presented in paper [11] under the tree topology with and without CS based data gathering algorithm. Furthermore, a hybrid CS based data gathering method was proposed which combines the conventional relay based data gathering method with the CS based data gathering method. It was shown that the network energy consumption could be further reduced by less transmitted data packets in the proposed protocol.

The application of CS was investigated in [2, 12] under the clustering routing structure. Since single-hop transmission is employed in the clustering topology, the packet loss on the links does not exhibit the correlation. As a result, the CS based data gathering algorithm is insensitive to the packet loss on the links. A CS based data gathering algorithm was proposed in [13] for unreliable links under the cluster topology where the column vectors of the measurement matrix are adjusted according to the packet loss nodes in the cluster. Therefore, the influence of packet loss on CS based data reconstruction can be alleviated. The tree-like multihop network routing topology is often used for large-scale WSNs. The CS based data gathering algorithm with multihop routing is studied in paper [14, 15], where the unreliability of the wireless link was ignored while special attention was paid to the optimal matching between the measurement matrix in CS and the structure of the tree-like routing topology. However, the transmission of CS data packets under the multihop routing requires the weighted superposition of the data from multiple nodes. Multiple original data can be lost once one packet loss occurs. Therefore, the transmission of CS data packets is highly sensitive to the packet loss under the multiple-hop routing.

It was shown via simulations in [16–18] that the data reconstruction accuracy under the tree topology can be seriously affected by the packet losses on the link and the Sparsest Random Scheduling (SRS) was further proposed for CS based data gathering in lossy WSNs. In this protocol, a sparsest measurement matrix is constructed according to the reception condition at the Sink end, which is further employed to reconstruct the original sensing data for all the nodes in the network and alleviate the influence of packet losses on CS data reconstruction [19–22]. However, this algorithm is only limited to the application scenarios where the spatial correlation of the sensing data in the network is relatively strong.

There are other data gathering methods in conventional WSNs such as ARQ, multipath transmission, network coding, etc. However, there are relatively few works studying the reliable CS based data gathering algorithm in WSNs. Furthermore, the CS based data gathering algorithm is much more sensitive to packet losses than conventional methods. Therefore, the study of CS based data gathering algorithm on unreliable links is quite meaningful to the application of CS theory in practical scenarios.

In wireless sensor network, serious packet loss will undermine the communication performance, service quality, and

application effect of sensor network. In recent years, the research premise of the CS based data gathering theory is the ideal link, and, because of the dynamic characteristics of the wireless link, channel interference and asymmetry of conflict, the wrong direction, and height of antenna, the unreliable link issues are commonly encountered in practical applications. There are many methods to ensure the reliable transmission of links in the traditional data gathering methods of wireless sensor networks, but to the best of our knowledge, there is little work for reliable sensor network data gathering method based on compression perception. In addition, the sensitivity of the CS data gathering method to link packet loss is much higher than that of traditional data gathering, so the research on compressed sensing data gathering algorithm under unreliable link is of great significance to the application of compressed sensing technology in real sensor network.

3. Network Model and Problem Description

The CS is a new technique which samples the sparse signal with a frequency below the Nyquist sampling frequency and achieves the projective transformation of the target signal from a high-dimension space to a low-dimension one. The accurate reconstruction of the compressed signal is achieved via the optimal reconstruction algorithm which is widely studied and applied in many areas due to its excellent compression performance.

Assume that N nodes are randomly deployed in the WSN and the gathered data is denoted as $d = (d_1, d_2, \ldots, d_N)^{\mathrm{T}}$. Assuming that d is sparse with respect to base $\Psi_{N \times N}$ and the measurement matrix is $\Phi = (\phi)_{M \times N}$, The received vector $Y_{M \times 1}$ can be expressed as $Y = (y_i)_{M \times 1} = \Phi \cdot \Psi^{\mathrm{T}} \cdot d$. The Sink node can reconstruct the original data with certain accuracy by solving the optimization problem in the following:

$$Y = \Phi \cdot S = \Phi \cdot \Psi^{\mathrm{T}} \cdot d = \Theta \cdot d \tag{1}$$

$$\vec{d} = \arg\min \|d\|_p \tag{2}$$

where $\Theta = \Phi \cdot \Psi^T$ is the sensing matrix and $\|d\|_p$ is the l_p norm of the sensing data vector $\vec{d}$ which is defined as

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} & 0 < p < +\infty \\ \max_{i=1,2\ldots,N} |x_i| & \end{cases} \tag{3}$$

In the data gathering process of WSNs, each round of CS based data gathering is performed with M times of independent measurements, which is expressed as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1} & \phi_{M2} & \cdots & \phi_{MN} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \tag{4}$$

Assume that N ordinary sensor nodes and one immobile Sink node are deployed in the WSN. All the sensor nodes

are uniformly and randomly deployed with fixed locations in a monitoring area of size $a \times a$. The Sink node is at the center of the monitoring area while the sensor nodes periodically gather and transmit the sensing data to the Sink node. Furthermore, the transmission power of the sensor nodes can be adjusted dynamically and adaptively. The Sink node is assumed with strong computation capability so that it can periodically gather and reconstruct the sensing data and acquire the location information for all the nodes in the network. The Minimum Spanning Tree (MST) routing is established by all the nodes in the network to perform data gathering, i.e., a connected undirected graph $G = (V, E(q))$ is constructed where $V = \{v_1, v_2, \ldots, v_N\}$ is the set of sensor nodes, $E(q) = \{e_1(q), e_2(q), \ldots, e_N(q)\}$ is the set of links in the MST, and $e_i(q)$ indicates that the link is connected with probability q. If we set $p=1-q$, then p indicates the packet loss ratio of the link. In addition, the CS technique is employed for data gathering in the WSN, which exhibits the following features: ① Discrete Fourier Transform (DFT) is employed for the sparse transformation base $\Psi_{N \times N}$ of the sensing data vector. The sparse transformation and the orthogonal sparse base are presented in (5) and (6), respectively. When M measurements are received at the Sink end, we employ the Orthogonal Matching Pursuit (OMP) algorithm to reconstruct the original sensing data. ② The relative error η is adopted in (7) as the metric to indicate the CS based reconstruction accuracy and a lower η means more accurate reconstruction. If the relative error is higher than 5%, the reconstruction is considered as a failure.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} x(n) W_N^{kn} \tag{5}$$

$$\Psi_j(t) = \frac{1}{\sqrt{N}} e^{i2\pi jt/N} \tag{6}$$

$$\eta = \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\sqrt{\sum_{n=0}^{N-1} (\hat{x}_n - x_n)^2}}{\sqrt{\sum_{n=0}^{N-1} x_n^2}} \tag{7}$$

The CS based data gathering process on unreliable links under the tree-like topology is illustrated in Figure 1.

If packet loss occurs on the link between S_5 and the Sink end, all the packets corresponding to S_5 as well as the child nodes of S_5 will be lost, which is shown in the frame in Figure 1. Therefore, the loss is more serious if the packet loss occurs on a link closer to the Sink end. Furthermore, since the weighted superposition data packets are transmitted between nodes in the CS based data gathering process, the Sink end also receives a weighted superposition data after each measurement. As a result, the Sink node cannot acquire the information on whether packet loss occurs or the number of lost packets. Assume that the Sink end receives the weighted superposition data of all the nodes in the network as the current measurement, based on which the reconstruction is performed to recover the original sensing data X.

Therefore, the CS based data gathering on unreliable links exhibits the following features: ① one packet loss on the link will result in the data loss of multiple nodes; i.e., the packet
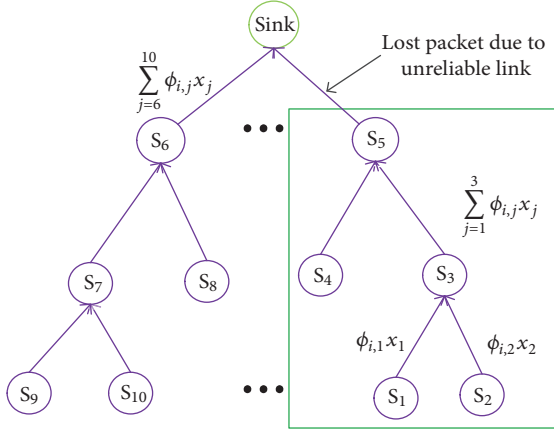
FIGURE 1: CS based data gathering on unreliable links.

loss exhibits the correlation effect. ② The Sink node has no information on the packet loss situation for the nodes in the network and regards the measurement data of the nodes in the network as the data projection to perform reconstruction. That is, the sensing data for compression does not match the sampling of the measurement matrix.

4. Algorithm Design and Realization

4.1. Design for SPLM Measurement Matrix. In order to solve the mismatch problem between the sensing data and the sampling of the measurement matrix, we design a sparse measurement matrix based on Packet-loss Matching (SPLM). In each measurement, the information of the packet loss node is omitted by the measurement matrix. As a result, the packet loss problem for CS based data gathering under tree topology is transformed into the measurement matrix projection problem based on sparse matching. Hence, the large-scale measurement and sampling are accomplished for the data in the network; meanwhile erroneous judgement for the data gathering situation can be avoided at the Sink end. The detailed realization of this process is as follows.

Definition 1 (link state matrix, LSM). The LSM is defined as the matrix which records the link state information with size $M \times N$, where M is the number of measurements and N is the number of nodes in the network. The entries in LSM are defined as follows:

$$L = \left(l\left(i, j\right) \right)_{M \times N} = \begin{cases} 0, & x_j \notin X \\ 1, & x_j \in X \end{cases} \tag{8}$$

Definition 2 (dense random projections, DRP). Each row of the DRP matrix contains $O(N)$ nonzero elements and the DRP matrix is usually constructed with

$$\Phi_d = \left(\phi\left(x, y\right) \right)_{M \times N} \tag{9}$$

The sparse measurement matrix based on packet loss matching employs the randomness of the packet losses on realistic links to construct the random sparse measurement matrix. The construction process can be achieved by multiplying the LSM with the DRP element-wise, as shown in

$$\Phi_s = L\Phi_d \tag{10}$$

If the packet loss ratio of the unreliable link is p, then each entry of the SPLM matrix is defined as follows:

$$\phi_s\left(i, j\right) = \begin{cases} +1, & with\ prob.\ \dfrac{\left(1 - p\right)}{2} \\ -1, & with\ prob.\ \dfrac{\left(1 - p\right)}{2} \\ 0, & with\ prob.\ p \end{cases} \tag{11}$$

The design of the measurement matrix should guarantee that most of the orthogonal base satisfies the RIP constraint. However, the proof of the RIP condition is a NP-hard problem. It was pointed in [11] that if the measurement matrix is full-rank, then, after the projection of this matrix, the data can be accurately reconstructed with a probability arbitrarily close to 1. Due to the fact that each element in the SPLM matrix follows the discrete random distribution in (11), each row of Φ_s can be regarded as a random sequence generated by the random variable ξ_n, which can further expressed by a discrete random process $\{\xi(n),\ n = 1, 2, \ldots, N\}$.

Theorem 3. *For a matrix $\Phi_s = (\xi_1, \xi_2, \ldots, \xi_M)^T$ with independently and identically distributed (i, i, d) random and discrete sequences ξ_i, if the random variable ξ_n which constitutes the sequence follows the distribution in (11), the matrix Φ_s will be full-rank with a probability arbitrarily close to 1.*

Proof. Assuming that matrix Φ_s satisfying the conditions above is not full-rank, i.e., a set of coefficients exists for the ith row of the matrix so that

$$\xi_i = a_1\xi_1 + a_2\xi_2 + \cdots + a_{i+1}\xi_{i+1} + \cdots + a_M\xi_M \tag{12}$$

And not all of the coefficients $a_1, a_2 \ldots a_M$ are zeroes. Define the random process $\{X(n),\ n = 0, 1, \ldots, N\}$ as the row vector ξ_i, the mean, and the variance functions are as follows:

$$EX\left(n\right) = \left[\frac{\left(+1\right)\left(1 - p\right)}{2} \right] + \left[\frac{\left(-1\right)\left(1 - p\right)}{2} \right] + 0 \times p \tag{13}$$
$$= 0$$

$$DX\left(n\right) = E\left[X\left(n\right) - EX\left(n\right)\right]^2 = E\left[X\left(n\right)\right]^2 = 1 - p \tag{14}$$

$\square$

Define the random process $\{Y(n),\ n = 0, 1, \ldots, N\}$ as $a_1\xi_1 + a_2\xi_2 + \cdots + a_{i+1}\xi_{i+1} + \cdots + a_M\xi_M$; then the mean and the variance functions are

$$EY\left(n\right) = E\left[\sum_{j \in [1, M], j \neq i} a_j\xi_j\left(n\right) \right] = \sum_{j \in [1, M], j \neq i} a_j E\xi_j\left(n\right) \tag{15}$$
$$= 0$$

$$DY(n) = E[Y(n) - EY(n)]^2 = E[Y(n)]^2$$

$$= \sum_{j \in [1,M], j \neq i} a_j^2 D\xi_j(n) = \sum_{j \in [1,M], j \neq i} a_j^2 (1-p) \quad (16)$$

Therefore, $X(n)$ and $Y(n)$ denote different random processes, respectively. For the discrete random process $X(n)$, the possible values of the random variable $X(i)$ are $x(i) \in \{+1, -1, 0\}$. Then the length of the state space I_X is 3^N. For the discrete random process $Y(n)$, the possible values of the random variable $Y(i)$ are $-M+1 \leq y(i) \leq M-1$ with $y(i) \in Z$. Therefore, the length of the state space I_Y is $(2M-1)^N$.

Define the case that (12) holds as event A, the case that not all of the coefficients $a_1, a_2 \ldots a_M$ are zeroes as event B, and the case that only one of the coefficients $a_1, a_2 \ldots a_M$ is nonzero as event C. Then

$$P(A \mid B) < P(A \mid C) \quad (17)$$

Solving the probability $P(A \mid C)$ can be transformed into solving the probability that two i.i.d random processes $X_1(n)$ and $X_2(n)$ have the same state simultaneously. According to the distribution in (11), different states have different probabilities in the state space for the random process $X(n)$. For the simplicity of analysis and without loss of any generality, we set the parameter in (11) as $p=1/3$. Then

$$P(A \mid B) < P(A \mid C) = \left(\frac{1}{3}\right)^N \ll 10^{-3} \quad (18)$$

Therefore, the probability of event A is extremely small; i.e., the original assumption does not hold and the matrix Φ_s is full-rank with a probability close to 1.

To evaluate the performance of the SPLI matrix, the classical CS data gathering algorithm, i.e., the CDG algorithm [18], is selected for comparison. Three different methods were used for the comparison of packet loss processing, as shown in Figure 2, where the CDG-DRP means applying the CDG to an unreliable tree topology with correlation effect of packet losses. A dense measurement matrix is used to measure the entire network data; CDG-SPLI indicates that the CDG algorithm is applied to the unreliable tree topology where packet losses exhibit correlation effects. The sparse measurement matrix based on packet loss tags is used to measure the data of the entire network; the CDG-SPLI-NC indicates that the CDG algorithm employs sparse measurement matrix based on packet loss tags to measure the network data without considering the correlation effect of packet losses. It is shown in Figure 2 that the reconstruction accuracy of the CDG-SPLI algorithm and the CDG-SPLI-NC algorithm is significantly higher than that of the CDG-DRP algorithm which utilizes a dense measurement matrix. This is due to the fact that the Sink end of the CDG-DRP algorithm suffers from the incorrect judgment of data packet reception condition. However, this problem can be avoided by the CDG-SPLI and CDG-SPLI-NC algorithms which employ the SPLI matrix for measurement. It can be seen that the misjudgment of the data packet reception condition at the Sink end can seriously deteriorate the reconstruction accuracy of the CS based data gathering algorithm. As for the two algorithms
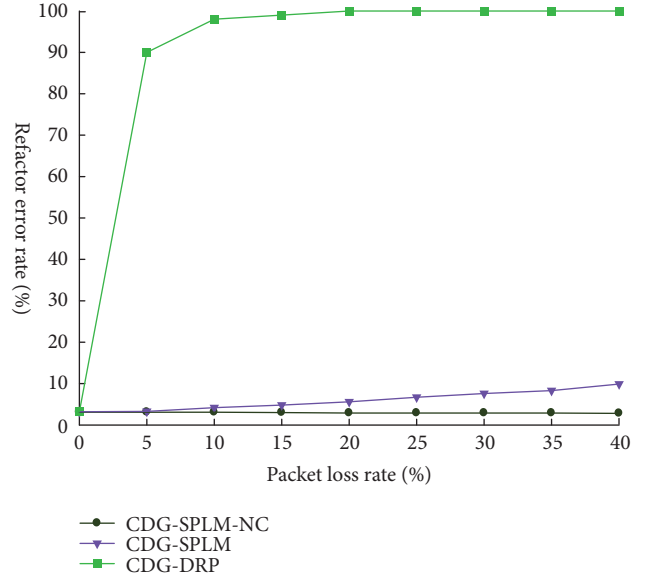

FIGURE 2: Performance comparisons for sparse measurement matrices based on packet loss tags.

employing the SPLI measurement matrix, the CDG-SPLI-NC algorithm does not consider the correlation effect of the packet losses and the CS based data reconstruction exhibit a high accuracy. The relative reconstruction error can be as low as 1.8% when the packet loss ratio is 40%. However, for the CDG-SPLI algorithm where the correlation effect exists, the relative reconstruction error exceeds 5% when the packet loss ratio is beyond 10%; i.e., the reconstruction fails. This is due to the fact that the correlation effect would cause the loss of the entire network data when the packet loss ratio is high. Therefore, the severe lack of CS measurements further deteriorates the reconstruction accuracy. Henceforth, in the data gathering process with correlated packet losses, it is not sufficient if we simply solve the misjudgment of the packet reception situation at the Sink end.

4.2. CS-PLM Algorithm. For the unreliable link under the tree topology, the CS based data gathering not only exhibits the correlation effect for packet losses but also suffers from the problem of misjudgment on the data packet reception situation at the Sink end. However, the SPLI matrix can only solve the misjudgment at the Sink end. Therefore relevant mechanisms still need to be studied to solve the problem of correlation effect in the process of CS based data gathering. Essentially, the correlation effect of the packet loss is caused by the weighted superposition processing of data packets during the CS based data gathering process, which is also the advantage of CS based data gathering. Therefore, the most effective method for solving the correlation effect is to guarantee the reliability of link transmission and avoid the appearance of correlated loss.

The cost of the guarantee for the reliable transmission in the whole network link is huge. To reduce the maintenance cost of the network, according to the performance analysis of the SPLM measurement matrix, this paper designs a hybrid

CS method for the data gathering in the network and divides the nodes of entire network node into traditional forwarding (TF) nodes and CS node, where the TF node only forwards data in a traditional data gathering manner, and the packet loss does not exhibit the correlation effect. The CS node transmits and receives data in a CS based data gathering manner and the packet losses are correlated. Therefore, for the data gathering between TF nodes, simply adopting the SPLM measurement matrix can overcome the impacts of packet losses on CS based data reconstruction. However, for CS nodes, in addition to using the SPLM measurement matrix, a corresponding mechanism must be designed to ensure the reliability of data transmission between CS nodes.

This paper designs a transmission mechanism based on multipath backup routing to ensure the reliability of data transmission between CS nodes. Under normal conditions, the CS node uses MST routing to transmit and receive data packets. If a packet loss occurs on the CS link, the transmitting node S_i of the data packet will choose another the transmission path and use the backup path to send data packets to the Sink. The node S_i can be seen as the source while the Sink is the destination node. There are many ways to construct the route from the source to the destination node. To reduce the energy consumption of the backup path transmission, the minimum energy consumption spanning tree is employed for constructing the routing. The energy consumption model of the network is as follows:

$$E(d) = 2\alpha_1 + \alpha_2 d^n \qquad (19)$$

where α_1 is the energy consumption coefficient of the circuit and α_2 is the power amplifying coefficient, d is the transmission distance, and n is the path loss factor with $2 \leq n \leq 5$. The value of n usually takes 2 in the free space. Therefore, the minimization problem of the energy consumption for the routing can be modelled as the optimization problem as in (20), where d_0 is the distance from the source node to the destination node, K is the number of hops, and d_i is the link distance between nodes.

$$\min \quad \left(\sum_{i=1}^{K} 2\alpha_1 + \alpha_2 d_i^n \right)$$
$$\text{s.t.} \quad \sum_{i=1}^{K} d_i = d_0 \qquad (20)$$

$$d_{char} = \left(\frac{2\alpha_1}{\alpha_2(n-1)} \right)^{1/n} \qquad (21)$$

The above optimization problem is solved using the Lagrangian multiplier method. The energy consumption of the network is the minimum if and only if the distance between the destination node and the source node is the same for each hop. We further present the value of the characteristic distance d_{char} as in (21). Finally, the optimal number of hop K_{opt} takes the maximum between d and d_{char}. We propose a centralized construction method for the backup path, which can be divided into the following four steps: ① according to the location information of each node, the Sink node

calculates the distance d from itself to the nodes; ② the Sink end compares the distance d to each node with the characteristic distance d_{char}. If $d \leq d_{char}$, a single-hop backup path had constructed. Otherwise, the optimal number of hops K_{opt} is calculated first. Then according to K_{opt} and the equal distance principle for each hop on the link from the CS nodes to the Sink end, the ideal locations of the relay nodes are further derived. ③ According to the ideal locations of the relay nodes, the Sink end chooses the CS nodes nearest to the ideal locations as the relay nodes. ④ The constructed routing with the minimal energy consumption is broadcast from the Sink node to the CS nodes and related relay nodes.

The CS-PLM algorithm first divides the nodes in the entire network into TF nodes and CS nodes. The MST routing tree is built for nodes in the entire network where the number of child nodes for node i in the network is ω_i. Specifically, $\omega_{sink} = N$ for the Sink end and $\omega_j = 0$ for the nodes at the end of the links. Define $\varepsilon = M-1$ as the threshold for discriminating the node types and the decision is made by each individual nodes. If $\omega_i > \varepsilon$, the node i participates in the data gathering in a CS method and is defined as the CS node. If $\omega_i \leq \varepsilon$, the node participates in the data gathering in a traditional relay method and is defined as the TF node. Therefore, in the data gathering process, the number of transmitted packets $PN(i)$ for node i is

$$PN(i) = \begin{cases} \omega_i + 1, & \omega_i \leq \varepsilon \\ M, & \omega_i > \varepsilon \end{cases} \qquad (22)$$

During the data gathering process of the CS-PLM algorithm, the TF nodes gather data along the MST routing in the traditional relay method. As shown by the white nodes in Figure 3, the packet loss does not exhibit the correlation effect. However, the CS nodes gather data along the MST routing in the CS method. As shown by the black nodes in Figure 3, the packet losses are correlated. In the MST routing, for the CS nodes directly connected to the TF nodes, all the gathered node data needs to be weighted and superimposed after completing the data gathering for all child nodes. The superimposed data is further combined into one data packet which is suitable for transmission between CS nodes. Related flag bits are included in the node ID part of the data packets from the child nodes to identify the data reception status of each node.

In the case of an unreliable link, no processing will be performed if packet losses occur in the link between TF nodes. However, if a packet loss occurs in the link between CS nodes, the data packet is directly sent to the Sink end using the minimum energy consumption backup path. After each round of data gathering, the Sink node builds an SPLM measurement matrix according to the packet loss during each measurement process and further employs the SPLM measurement matrix and the M measurements received at the sink end to reconstruct the original sensing data.

4.3. Data Gathering and Analyses. The data gathering process of the CS-PLM algorithm can be divided into three stages. The first stage is initialization of the sensor network, in
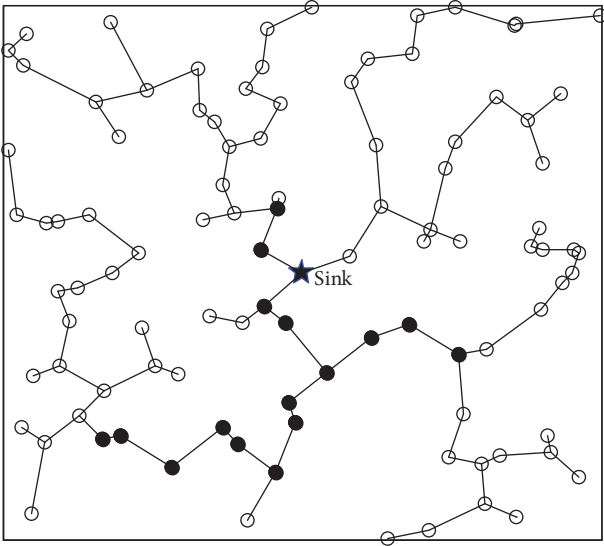
FIGURE 3: Distributions of different node types in the CS-PLM algorithm.

which the node networking, the accumulation of the a priori information for the reception state of each link, and the configuration for the node measurement vector are accomplished. At the second stage, the CS based data gathering is performed on the lossy link and the effective CS based sampling and gathering is achieved for all the nodes in network. At the third stage, the CS based reconstruction is achieved for the sampled data and the original data is further acquired for the nodes in the network. The operation of the algorithm is detailed as follows:

(1) Initialization of the sensor network: initiated by the Sink end, the minimum spanning tree (MST) routing is established to achieve the node networking. Then the Sink end broadcasts the heartbeat data packet to all the nodes in the network. At the reception of the heartbeat data packet, the nodes in the network turn on its own timer. Within the time period $T1$, the nodes transmit and relay the data packet along the routing to the Sink end and the data packet contains the heartbeat information. In this process, each node tracks the real-time packet reception state on its own reception link and stores the result in its memory. This result in the memory will be further employed as the a priori information for the packet loss type prediction which is based on the sliding window scheme. That is, the reception sequence $\{X_i\}$ for each node is initiated till the end of $T1$. Then the Sink end first broadcasts the random seed ξ. At the reception of the random seed, node i combines the seed with its own ID and generates the message (ξ, IDi). A column of the measurement $(\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{iM})^T$ is further generated for each node and stored in its own memory.

(2) CS based data gathering: in the CS based data gathering process, each node multiplies the gathered data di with the corresponding measurement coefficient φ_{ij} according to the routing. Then the messages are added up sequentially and relayed to the Sink end. When a node recognizes its packet loss in this process, the probability density functions $f1(z)$ and $f2(z)$ of the random variable z under two different

packet loss types are determined according to the current bit error rate Pb of the link and the data packet length B. According to (15), the decision threshold ξ for the two types of packet loss is calculated. In the sliding window, the value of the random variable z is calculated and we further compare the value of z with the threshold η. When $z<\eta$, the retransmission scheme is employed for packet recovery. Assuming that the maximal number of retransmission is max_num, the sliding window is updated in a real-time method during the retransmission. If max_num is reached and the received data of the node still remains unrecovered, then the value of z will be compared with the threshold ξ according to the latest update in the sliding window. The current packet loss type is further predicted again according to the comparison result and the corresponding recovery schemes will be adopted. Specifically, the prediction based on the correlation of the temporal sequence is adopted when $z>\eta$ and a k-order temporal correlated sequence Hk is constructed for the node with lost packets. According to (17), the prediction value $h^{\frown}T_i + 1$ is calculated for the lost packet and further transmitted to the next one-hop node, assuming that the predicted packet is identical to the lost packet. Therefore, the CS based measurement and sampling is finished and the prediction order k is updated.

(3) Data reconstruction: when the Sink end receives M measurements in one round of data gathering, the measurement vector $Y = (y_1, y_2, \ldots, y_M)^T$ is constructed and the Sink end reconstructs the measurement matrix $(\varphi_{ij})_{M \times N}$ according to the random seed ξ and the IDs in the whole network. According to the sparse base $\Psi_{N \times N}$, the CS reconstruction algorithm is employed to reconstruct the sparse signal S and we can recover the original signal vector d by the calculation of $d=\Psi \cdot S$. According to the process stated above, the prediction of the packet loss type, lost packet retransmission, and the temporal correlation based prediction can be simply executed serially and the execution complexity is O(1). As a result, for performance-limited single nodes, this algorithm could guarantee the real-time online operation. For the whole network, at least $N \times M$ data packets have to be received or transmitted for one round of data gathering operation of the CS-PLM algorithm where N is the number of nodes in the network while M is the number of measurements. During the operation of the CS based data gathering algorithm, receiving or transmitting one data packet involves one multiplication, some additions, and no complex regressive operations. Furthermore, according to the CS theory, the number of measurements $M \geq O (K \log N)$ where K is the sparsity degree of the original data, which could be regarded as a constant. By considering the whole network, the complexity of the CS-PLM algorithm is O($N \log N$), which indicates that the CS-PLM algorithm causes no additional computations in comparison with conventional CS densely projective data gathering algorithms.

5. Performance Evaluation

In order to evaluate the performance of the CS-PLM algorithm, we employ the tool MatlabR2014a to perform the
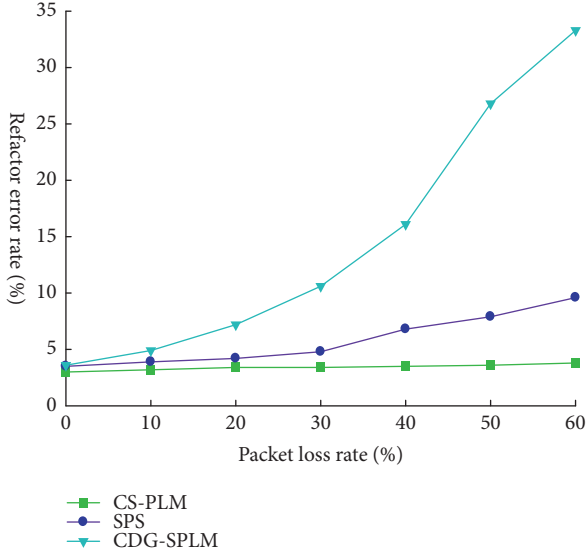
Figure 4: Reconstruction error with different packet loss ratio.

Figure 5: Network packet loss ratio with different sparsity Degree.

simulations. Based on the network model in this paper, we assume that the network is deployed in a 200×200 monitoring area and 400 sensor nodes are placed in the network. The original data sources are assumed to abide by the two-dimensional Gaussian distribution.

The reconstruction accuracy of the CS-PLM algorithm is illustrated with different network packet loss ratio in Figure 4. For comparison, we also investigate the performances of the Sparsest Random Scheduling (SRS) based CS algorithm [2] and the CDG-SPLM algorithm. It is shown in Figure 4 that the reconstruction error performances of these three algorithms increase with the packet loss ratio while the CS-PLM algorithm outperforms the others. The performance gap in terms of reconstruction error between the CS-PLM algorithm and the others also widens with the packet loss ratio. The reason for the performance gap lies in the fact that the SRS algorithm employs the sparsest measurement matrix to reconstruct the original data. On the one hand, the amount of sampled data is far from enough for the purpose of measurement. On the other hand, packet loss will reduce the number of measurements and the reconstruction accuracy is further undermined due to the lack of sampling. Although the CDG-SPLM algorithm could overcome the misjudgment problem for the packet reception state at the Sink end, the correlation of the packet loss in the CS based data gathering process would seriously affect the reconstruction accuracy. However, the proposed CS-PLM algorithm could accomplish as many measurements as possible and overcome the misjudgment problem at the Sink end as well as the packet loss correlation. These improvements could guarantee the reliability of the CS based data gathering process on unreliability links and the highly accurate CS based reconstruction. Furthermore, according to the simulation results, the CS-PLM algorithm could still guarantee the effective reconstruction for the gathered data with a 60% packet loss ratio, which proves the robustness of the proposed algorithm.
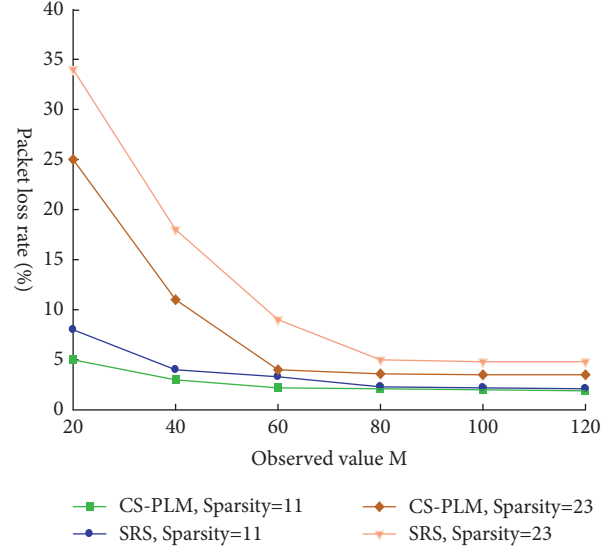
In order to evaluate the performance of the algorithm under different degree of correlation, we choose two data sets with different degree of sparsity; i.e., the results are illustrated in Figure 5 for the reconstruction accuracy of the CS-PLM algorithm and the SRS algorithm, where the link packet loss ratio is set to 20%. It is shown in Figure 5 that, for the same algorithm, a lower sparsity degree would improve the reconstruction accuracy because the data correlation degree increases with the sparsity degree and, according to the CS theories, the reconstruction accuracy can be further improved. When the sparsity degree is set to 11, the proposed algorithm only slightly outperforms the SRS algorithm but when the sparsity degree is 23, the performance of the CS-PLM algorithm is far better than that of the SRS algorithm. This is because when the data is weakly correlated, the SRS algorithm could not recover the data precisely since it is only based on a small amount of compressed samples. However, the CS-PLM algorithm accomplishes as many measurements as possible and the effective reconstruction can thus be guaranteed for data sets with common degree of correlation. Generally, the CS-PLM algorithm could effectively reduce the strong dependency of the CS based data gathering algorithms on the correlation of the data set. The reliable CS based data gathering can be further guaranteed on unreliable links for data sets with ordinary correlation degree.

The network lifetime performances for different algorithms are compared in Figure 6 where the CDG-Retransmission algorithm further adopts the retransmission scheme based on the CDG algorithm to tackle the packet loss on unreliable links. The network lifetime is defined as the number of rounds from the beginning to the first node failure. It is shown in Figure 6 that the proposed CS-PLM algorithm increases the network lifetime by 400% and 140% in comparison with the SRS algorithm and the CDG-Retransmission scheme. This is due to the energy imbalance of the SRS algorithm which causes the capability vacancy and the early failure of the network. The retransmission scheme is adopted
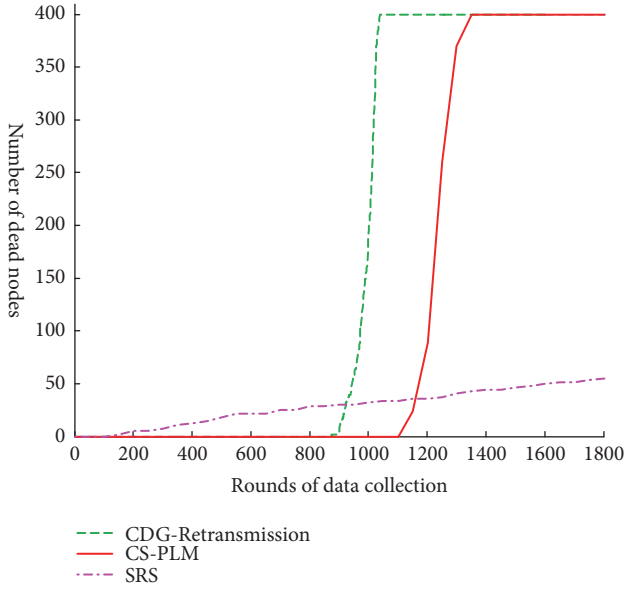
FIGURE 6: Network lifetime for the CS based algorithms.

FIGURE 7: Relative reconstruction errors under different sparse bases.

FIGURE 8: Relative reconstruction errors with different number of measurements.

for the CDG-Retransmission algorithm for packet loss recovery while more packets are transmitted and the network lifetime is therefore shortened.

According to the theory of CS, the correlation between the measurement matrix and the sparse base of the data will affect the reconstruction performance of the algorithm. To verify the performance of the CS-PLM algorithm under different sparse basis, we choose two sparse bases, i.e., the Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT) as a comparison. Related results are shown in Figure 7. It is shown in Figure 7 that, with different packet loss ratio, the proposed algorithm performs better in the DCT base; i.e., the DCT base exhibits weaker correlation with the SPLM matrix in comparison with the DFT base and we can obtain a higher reconstruction accuracy by choosing the DCT base to perceive the changing sparsity degree.

The impacts of packet loss ratio on relative reconstruction error are illustrated in Figure 8 with different number of measurements. The number of measurements for the CS-PLM algorithm is set to 60, 80, 120, and 140, respectively. It is shown in Figure 8 that when the number of measurements is 60, the relative reconstruction error changes significantly with the packet loss ratio and the increasing packet loss ratio would undermine the recovery accuracy. However, when the number of measurements increases, the impacts of packet loss ratio on the reconstruction accuracy are further alleviated. For example, when the number of measurement is 140, the relative reconstruction errors are almost the same for the packet loss ratio of 10%, 30%, and 50%. The reason behind this phenomenon is that the proposed CS-PLM algorithm adopts the measurement matrix based on packet loss matching to sample the data in the network. The increase of the packet loss ratio would lead to a higher sparsity degree in the measurement matrix. In order to maintain the reconstruction accuracy, the number of measurements
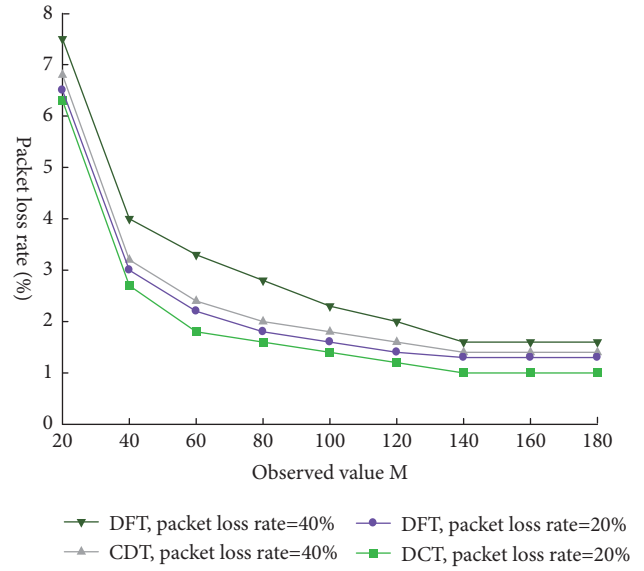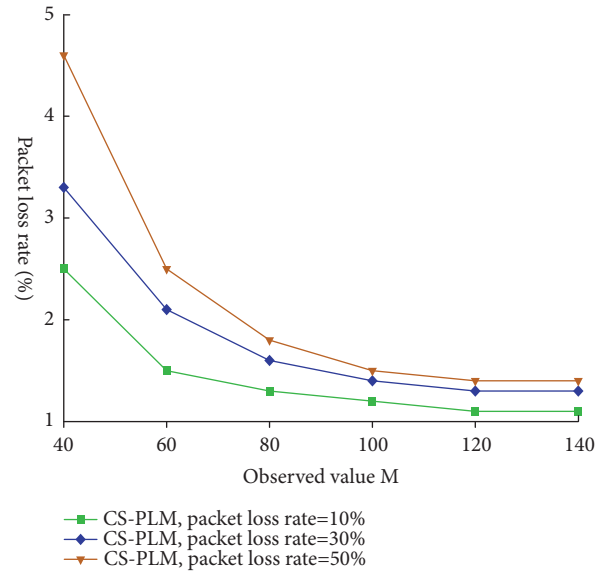
should be increased to deal with the higher sparsity degree. Therefore, more measurements can alleviate the impacts of the packet loss ratio on the reconstruction accuracy to some extent.

When the decay coefficient of the event source is $n=0.01$ and the neighborhood range is $r=2$, the performance of the CS-PLM algorithm is shown in Figure 9. When the bit error rate is relatively low, i.e., $Pb=10^{-5}$, the performances of these three algorithms are almost the same. However, when the bit error rate is as high as $Pb=10^{-3}$, the CS-PLM algorithm outperforms the others. It is shown in Figure 7 that when $Pb=10^{-3}$, the SNR for data reconstruction of the CDG algorithm is 27.33dB and the high bit error rate obviously
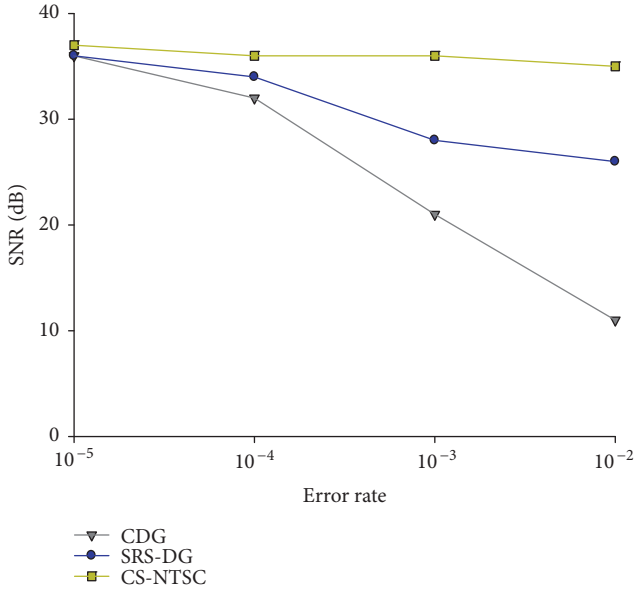
FIGURE 9: Performance analyses of the CS-PLM algorithm.

FIGURE 10: Relative reconstruction error with different network scale.

affects the performance of the CDG algorithm. The SNR for data reconstruction of the SRS-DG algorithm is 29.72dB. Since the SRS-DG algorithm is designed for recovering block loss, the data block is abandoned once transmission error occurs and the error-free nodes are measured with the sparse measurement matrix. Therefore, the measured information is reduced every time. The lost data block is then compensated by increasing the number of measurements through the next round of data gathering and the reconstruction SNR is therefore increased. As a result, the SNR is not high for this round of data reconstruction. The data reconstruction SNR obtained by the CS-PLM algorithm is 35.91dB. The transmission error is predicted by the spatial correlation of the data under certain conditions and the amount of abandoned information is therefore reduced. Henceforth, in the wireless scenarios with high bit error rate, the CS-PLM algorithm does not cause additional communication energy consumption and it can overcome the impacts of erroneous data blocks on data reconstruction. The efficiency of the CS-PLM algorithm is therefore verified.

For the CS-PLM algorithm, under the premise of ensuring the constant data compression rate, different network sizes require different measurement numbers in the data gathering process, which in turn causes the changing threshold for the judgment of node type. Therefore, different network size leads to different proportions of node types, which not only affects the total packet throughput of the network, but also affects the complexity of constructing the backup paths between CS nodes as well as the reliability of their transmission. Therefore, the performance of the algorithm is affected by the network size and related results are illustrated in Figure 10, where the packet loss ratio on the link is set to 20%. According to the curves in Figure 10, it is shown that as the number of nodes increases, the proportion of CS nodes decreases in the network. As for the reason, in large-scale

WSNs, the branches in tree topologies will be rather large and there are many smaller branches. However, according to the node classification rules in hybrid CS schemes, usually only the nodes in the main branch will be classified as CS nodes. Therefore, the proportion of CS nodes decreases, so does the complexity of constructing backup paths between CS nodes. According to the curves of the relative reconstruction error in Figure 10, the reconstruction accuracy of the CS-PLM algorithm improves with the increasing number of nodes in the network. This is due to the fact that as the proportion of CS nodes in the network decreases, the packet losses occurring in the network are uncorrelated with a larger probability. As a result, the reconstruction is less affected by the reliability of the backup paths and higher reconstruction accuracy can be guaranteed. According to the simulations above, the CS-PLM algorithm exhibits better performances in large-scale WSNs. Since the SPLI measurement matrix is designed in the CS-PLM algorithm, according to the CS theory, the correlation between the measurement matrix and the sparse base will affect the reconstruction performance of the algorithm. In order to investigate the performances of the CS-PLM algorithm under different sparse bases, we adopt the DCT base and the DFT base for comparison, as shown in Figure 11. It is shown that, with different number of measurements, the algorithm performs better under the DCT base. Therefore, it can be concluded that the SPLI matrix exhibits a lower correlation with the DCT base. That is, a better reconstruction performance of the algorithm can be ensured by choosing the DCT base to sense the changing of the data sparsity.

6. Conclusions

In order to address the CS based data gathering problem on unreliable links, we have proposed a CS-PLM algorithm. We designed the SPLM measurement matrix by analyzing the influence of the packet loss on CS based data gathering
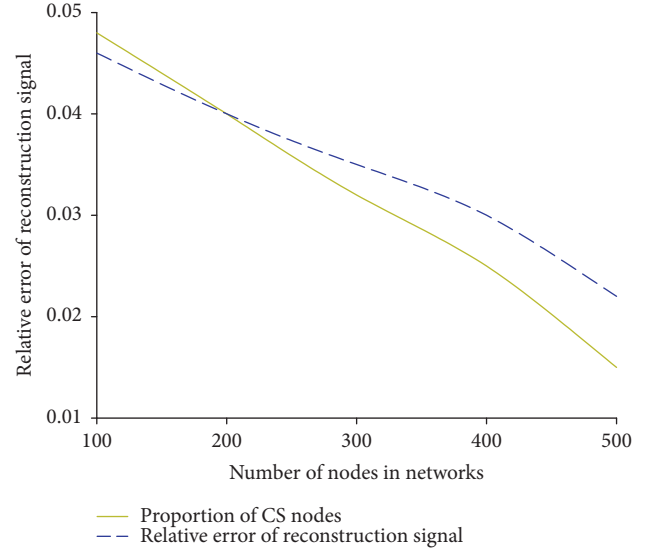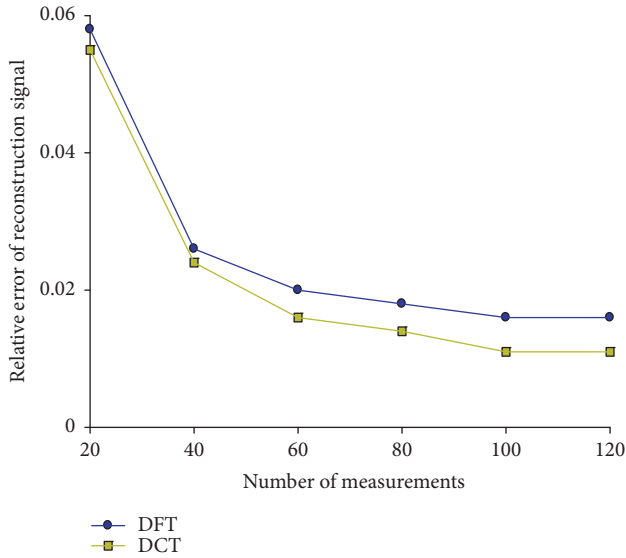
Figure 11: Relative reconstruction error under different sparse bases.

and further verified through simulations that the correlation of packet losses would undermine the reconstruction performance of the SPLM measurement matrix. Therefore, the nodes in the network are divided into TF nodes and CS nodes. Packet losses between TF nodes do not exhibit correlation and we simply adopt the SPLM measurement matrix to perform measurement projection. However, besides adopting the SPLM measurement matrix for measurement projection, the CS nodes also guarantee the transmission reliability by the minimum energy consumption backup paths and avoid the occurrence of correlated lost packets. It was shown in the simulation results that when the packet loss ratio on the link is 60%, the CS-PLM algorithm could still guarantee the effective reconstruction of the compressed data. Compared with other algorithms, the proposed algorithm showed great improvements in terms of reconstruction accuracy and the sparsity degree of the data set. It can accurately reflect the influences of the network size, sparse base, and the number of measurements for data gathering on the performance of the CS-PLM algorithm. Future work may focus on the impacts of packet loss ratio for mobile nodes on the reconstruction accuracy when the network flow is sufficiently large or small.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] W. Fang, X. Yao, X. Zhao, J. Yin, and N. Xiong, "A Stochastic Control Approach to Maximize Profit on Service Provisioning for Mobile Cloudlet Platforms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 4, pp. 522–534, 2018.

[2] T. Wang, J. Zhou, M. Huang et al., "Fog-based storage technology to fight with cyber threat," *Future Generation Computer Systems*, vol. 83, pp. 208–218, 2018.

[3] V. K. Singh and M. Kumar, "A Compressed Sensing Approach to Resolve The Energy Hole Problem in Large Scale WSNs," *Wireless Personal Communications*, pp. 1–17, 2017.

[4] P.-F. Wu, F. Xiao, C. Sha, H.-P. Huang, R.-C. Wang, and N.-X. Xiong, "Node scheduling strategies for achieving full-view area coverage in camera sensor networks," *Sensors*, vol. 17, no. 6, article no. 1303, 2017.

[5] R. Zhou, C. Yang, J. Wan, W. Zhang, B. Guan, and N. Xiong, "Measuring complexity and predictability of time series with flexible multiscale entropy for sensor networks," *Sensors*, vol. 17, no. 4, 2017.

[6] V. Angayarkanni, V. Akshaya, and S. Radha, "Design of a Compressive Sensing Based Fall detection System for Elderly Using WSN," *Wireless Personal Communications*, vol. 98, no. 1, pp. 421–437, 2018.

[7] F. Pareschi, M. Mangia, D. Bortolotti et al., "Energy Analysis of Decoders for Rakeness-Based Compressed Sensing of ECG Signals," *IEEE Transactions on Biomedical Circuits and Systems*, 2017.

[8] M. K. Khan, M. Shiraz, K. Z. Ghafoor, S. Khan, A. S. Sadiq, and G. Ahmed, "EE-MRP: Energy-Efficient Multistage Routing Protocol for Wireless Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, Article ID 6839671, pp. 1–13, 2018.

[9] M. Wu, L. Tan, and N. Xiong, "Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications," *Information Sciences*, vol. 329, pp. 800–818, 2016.

[10] Z. Wu, N. Xiong, Y. Huang, and Q. Gu, "Optimal service distribution in WSN service system subject to data security constraints," *Sensors*, vol. 14, no. 8, pp. 14180–14209, 2014.

[11] C. Yang, D. Puthal, S. P. Mohanty, and E. Kougianos, "Big-Sensing-Data Curation for the Cloud is Coming: A Promise of Scalable Cloud-Data-Center Mitigation for Next-Generation IoT and Wireless Sensor Networks," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 48–56, 2017.

[12] S. Khanna and C. R. Murthy, "Communication-efficient decentralized sparse Bayesian learning of joint sparse signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 617–630, 2017.

[13] B. Sun, Y. Guo, N. Li, and D. Fang, "Multiple Target Counting and Localization Using Variational Bayesian em Algorithm in Wireless Sensor Networks," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2985–2998, 2017.

[14] Z. Sun, X. Xing, C. Li, Y. Nie, and Y. Cao, "A Novel Nonlinear Multitarget k-Degree Coverage Preservation Protocol in Wireless Sensor Networks," *Journal of Sensors*, vol. 2016, 2016.

[15] T. Wang, J. Zeng, Y. Lai et al., "Data collection from WSNs to the cloud based on mobile Fog elements," *Future Generation Computer Systems*, 2017.

[16] H. Djelouat, H. Baali, A. Amira, and F. Bensaali, "Joint sparsity recovery for compressive sensing based EEG system," in *Proceedings of the 2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*, pp. 1–5, Salamanca, September 2017.

[17] Z. Sun, Y. Zhang, Y. Nie, W. Wei, J. Lloret, and H. Song, "CASMOC: a novel complex alliance strategy with multi-objective optimization of coverage in wireless sensor networks," *Wireless Networks*, vol. 23, no. 4, pp. 1201–1222, 2017.

[18] T. Wang, Z. Peng, J. Liang et al., "Following targets for mobile tracking in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 12, no. 4, 2016.

[19] Z. Sun, Y. Shu, X. Xing, W. Wei, H. Song, and W. Li, "LPOCS: A novel linear programming optimization coverage scheme in wireless sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 33, no. 1-4, pp. 173–197, 2016.

[20] G. Xing, M. Li, T. Wang, W. Jia, and J. Huang, "Efficient rendezvous algorithms for mobility-enabled wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 47–60, 2012.

[21] S. Wang, H. Yi, L. Wu, F. Zhou, and N. N. Xiong, "Mining probabilistic representative gathering patterns for mobile sensor data," *Journal of Internet Technology*, vol. 18, no. 2, pp. 321–332, 2017.

[22] Z. Lu, W. W. Li, and M. Pan, "Maximum lifetime scheduling for target coverage and data collection in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 2, pp. 714–727, 2015.

Research Article

High-Throughput Fast-SSC Polar Decoder for Wireless Communications

Xiaojun Zhang [ID],[1] **Xiaofeng Yan,**[1] **Qingtian Zeng,**[1] **Jianming Cui,**[1]
Ning Cao [ID],[2] **and Russell Higgs** [ID][2]

[1]*Shandong University of Science and Technology, Qingdao 266590, China*
[2]*University College Dublin, Belfield, Dublin 4, Ireland*

Correspondence should be addressed to Xiaojun Zhang; buttern@163.com

Polar code has been proven to achieve the symmetric capacity of memoryless channels. However, the successive cancellation decoding algorithm is inherent serial in nature, which will lead to high latency and low throughput. In order to obtain high throughput, we design a deeply pipelined polar decoder and optimize the processing elements and storage structure. We also propose an improved fixed-point nonuniform quantization scheme, and it is close to the floating-point performance. Two-level control strategy is presented to simplify the controller. In addition, we adopt FIFO structure to implement the α_memory and β_memory and propose the 348-stage pipeline decoder.

1. Introduction

Wireless communication is changing our life and has been applied to many scenarios [1–5], and error-correcting codes are utilized to improve its transmission efficiency and reliability. Polar code is a class of error-correction codes proposed by Arikan [6]. Within the ongoing 5th-generation wireless systems (5G) standardization, polar codes have been adopted as channel coding for the enhanced mobile broadband (eMBB) communication service for its excellent error-correction performance. Especially, for the ultra-reliable low-latency communications (URLLC), it should satisfy the high throughput of several tens Gbps [7], which bring in a challenge for polar decoder. In the past, great research efforts have been made on polar codes in decoding algorithm and hardware architecture since Arikan presented the successive cancellation (SC) decoding algorithm. SC has the advantages of low complexity and simple decoding structure. Although polar codes can theoretically achieve channel capacity when code length is infinite, the performance of SC is mediocre for codes of short and moderate lengths. To address this issue, successive cancellation list (SCL) decoding algorithm is proposed in [8]. Different from the SC algorithm, SCL does

not focus on a single candidate codeword; it saves L most reliable candidate codewords at every step. The decoding performance of SCL has been significantly improved. K. Chen and K. Niu proposed CRC-aided SCL (CA-SCL) algorithm [9] based on that the correct codewords can pass the CRC check. And they proposed the successive cancellation stack (SCS) decoding algorithm in [10] and successive cancellation hybrid (SCH) decoding algorithm in [11]. Unlike the SCL decoding which preserves the L most reliable paths in each layer, SCS always extends the most reliable path. Compared with SCL decoding, the performance of SCS is the same as SCL, but the time complexity is lower and the space complexity is higher. The actual time complexity of SCS decoding is far below than that of SCL in the high-SNR regime and is close to SC decoding. SCH algorithm combines the advantages of SCL and SCS, and the performance of SCH is close to that of maximum likelihood (ML) [12]. The researchers of Huawei proposed the adaptive CA-SCL (aCA-SCL) [13] decoding algorithm based on CA-SCL algorithm. aCA-SCL improves the decoding performance by gradually expanding the search width L. aCA-SCL can reduce the software complexity significantly. The above decoding algorithm is proposed for improving the performance, but their throughput is not

ideal. Thus A. Alamdar-Yazdi and F. R. Kschischang propose simplified successive cancellation (SSC) decoding algorithm in [14] based on the location of frozen and information bits. SSC decoding reduces the computational complexity and improves the decoding parallelism by combining some leaf nodes, such as Rate-1 node whose leaf nodes are all unfrozen bits. Simulation results illustrate that the performance is similar to that of SC. G. Sarkis and W. J. Gross divide the leaf nodes into Rate-0, Rate-1, and Rate-R nodes and propose the maximum likelihood SSC (ML-SSC) decoding algorithm in [15] which is mainly to improve the performance of Rate-R nodes decoder in SSC decoding. Compared with the semiparallel SC decoding in [16], ML-SSC decoding improves the decoding throughput by 25 times. In order to further reduce the decoding complexity and improve the throughput, G. Sarkis proposes the fast simplified successive cancellation (Fast-SSC) decoding algorithm, which mainly improves the decoding rules of Rate-R nodes and gives the specific operation for each constituent node [17]. Fast-SSC decoding divides the Rate-R nodes into repetition (REP), single-parity-check (SPC), and REP-SPC nodes and improves the throughput.

In addition, for the decoding hardware architecture of polar codes, a semi-parallel architecture is proposed in [16]. In order to improve resource utilization, this method reuses the processing elements (PE) which effectively reduce the hardware complexity. The overlapped architectures proposed in [18] have advantages in both latency and throughput, which uses precalculation function calculate the possible results firstly and according to the decoded results to choose the corresponding results. It is proved that the decoding latency is reduced by 50% when the code length is larger than 2^7. Then B. Yuan proposed the SCL decoder with multi-bit decision which effectively reduces the decoding latency [19]. And an unrolled hardware polar decoder is proposed in [20] on the basis of Fast-SSC. This decoder loads one frame channel decoding data and outputs a frame of codeword each clock. The PEs are no longer reused and dedicated PEs are assigned to each stage. Graphics processor unit (GPU) provides the flexibility and massive parallel units; the GPU-based polar decoders obtain high throughput [21–23].

In this paper, we investigate the characters of LLRs for different stages of Fast-SSC polar decoder and propose an improved nonuniform fixed-point quantization method. It adopts (6,5,1) quantization scheme; the decoding performance is close to the floating-point decoding performance. The proposed decoder employs deeply pipelined architecture and optimizes the REP decoder and G operation. To simplify the deeply pipelined control, the controller is divided into global controller and local controller. α_memory and β_memory use FIFO architecture to reduce the control logic. Finally, a 348-stage pipeline architecture is devised, which is implemented on Altera Stratix V 5SGXEA7N2F45C2. To test the decoding performance, we design a platform based on FPGA.

The remainder of this paper is organized as follows. A brief review of Fast-SSC decoding algorithm and analysis of the quantization schemes are shown in Section 2. Section 3

depics the deeply pipelined architecture and the PEs. Performance is evaluated in Section 4 and conclusions are drawn in Section 5.

2. Review of Fast-SSC

2.1. Polar Codes. A polar code can be represented by $P(N, K)$, where N denotes the code length and K/N is the code rate. Polar code of length N can be constructed by concatenating two polar codes of length $N/2$. The construction method can be denoted by $x = uG^{\otimes n}$, where $u = \{u_0, u_1, \ldots, u_{N-1}\}$ is the input sequence that to be encoded, and $x = \{x_0, x_1, \ldots, x_{N-1}\}$ denote the codewords. $G^{\otimes n}$ is the n-th Kronecker power of the generator matrix $F = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$. Polar codes select K most reliable channels to transmit information bits, and the other N-K channels transmit frozen bits.

2.2. Fast-SSC Decoding Algorithm. The binary decoding tree of Fast-SSC nodes is divided into four types: Rate-0, Rate-1, REP, and SPC. Compared with SC decoding tree, Fast-SSC has less leaf nodes. Since the polar decoder traverses the entire binary tree during iterations, Fast-SSC decoding algorithm has low latency. Figure 1 shows SC decoding tree and corresponding Fast-SSC decoding tree for a (16, 8) polar code. For instance, the REP node consists of leaf node $\{4, 5, 6, 7\}$ and SPC node includes leaf node $\{8, 9, 10, 11\}$. The leaf nodes of Rate-0 node are all frozen bits. Therefore, its output will be the zero vector. The leaf nodes of Rate-1 node are all information bits. The decoding result of such nodes is obtained by

$$\beta_v[i] = \begin{cases} 0, & \alpha_v[i] \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

For the REP node, only the last bit of it is information bit, and others are frozen bits. The REP node adds all the input α first and then makes a hard decision as

$$\beta_v = \begin{cases} 0, & \text{when } \left(\sum_{i=0}^{N_v-1} \alpha_v[i] \right) \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

where N_v denotes the code length of the node.

The SPC node, of which only the first leaf node is frozen bit, performs threshold detection by (3) on the input LLRs firstly. The parity of all the inputs is calculated by (4). Then the least reliable bit is founded and flipped if the parity constraint is not satisfied. The threshold detection can be written via

$$\beta_v[i] = \begin{cases} 0, & \alpha_v[i] \geq 0 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

The parity of the input is calculated as

$$parity = \bigoplus_{i=0}^{N_v-1} \beta_v[i] \tag{4}$$

(a) SC decoding tree

(b) Fast-SSC decoding tree
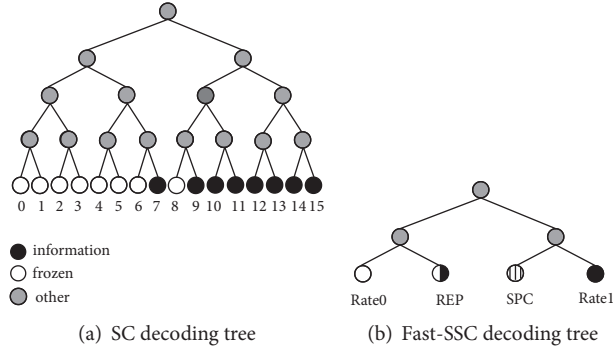
FIGURE 1: Decoding tree for $P(16, 8)$.

Finally, the output of the SPC node is

$$\beta_v[i] = \begin{cases} \beta_v[i] \oplus parity, & \text{when } i = j \\ \beta_v[i], & \text{otherwise} \end{cases}$$

$$i = \underset{j}{\arg\min} \left(|\alpha_v[i]| \right) \quad (5)$$

In addition to the above four type nodes, the rest colored in grey is referred to as other node, as shown in Figure 1. The decoding method of other nodes uses standard SC algorithm as in Figure 2. When node v is activated, it will receive α_v from its parent node p_v and then calculate the soft-valued input to its left child, α_l, which is calculated using the F operation.

$$\alpha_l[i] = F\left(\alpha_v[i], \alpha_v\left[i + \frac{N_v}{2}\right]\right)$$

$$= \text{sign}\left(\alpha_v[i]\right)\text{sign}\left(\alpha_v\left[i + \frac{N_v}{2}\right]\right) \quad (6)$$

$$* \min\left(|\alpha_v[i]|, \left|\alpha_v\left[i + \frac{N_v}{2}\right]\right|\right)$$

Once β_l of the left child node is estimated, it is used to calculate the input to the right child node α_r with G operation.

$$\alpha_r[i] = G\left(\alpha_v[i], \alpha_v\left[i + \frac{N_v}{2}\right], \beta_l[i]\right)$$

$$= \begin{cases} \alpha_v\left[i + \frac{N_v}{2}\right] + \alpha_v[i], & \text{when } \beta_l[i] = 0 \\ \alpha_v\left[i + \frac{N_v}{2}\right] - \alpha_v[i], & \text{otherwise} \end{cases} \quad (7)$$

Finally, β_l and β_r are combined to calculate β_v as

$$\beta_v[i] = \begin{cases} \beta_l[i] \oplus \beta_r[i], & \text{when } i < \frac{N_v}{2} \\ \beta_r\left[i - \frac{N_v}{2}\right], & \text{otherwise} \end{cases} \quad (8)$$

Table 1 lists the number of constituent nodes of the decoding tree for a (1024, 512) polar code. It can be seen that the total number of constituent nodes is 104 of the Fast-SSC decoding, which decreases from 1024 of the SC decoding

TABLE 1: The number of constitute nodes.

Node Types	Rate-0	Rate-1	REP	SPC	total
Numbers	14	40	24	26	104
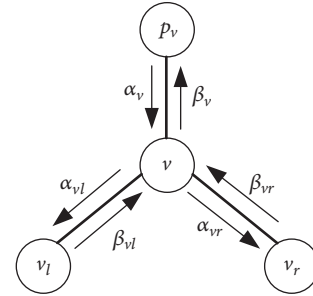
FIGURE 2: Local decoder of node v.

tree. The decoder does not need to traverse the entire decoding tree, it just traverses the pruned tree. Thus Fast-SSC algorithm improves the decoding efficiency and throughput and decreases the latency.

2.3. Quantization Scheme. The quantization scheme is divided into uniform and nonuniform quantization. The uniform quantization is simple, but the consumption of resources is more than that of nonuniform scheme. The nonuniform quantization employs different quantization bits in different decoding stages and uses less storage resources, but the memory structure is not regular [27]. Unlike the conventional SC decoder which memory is shared for the nodes of different stages, the PEs of deeply pipelined decoder in each stage are equipped with a separate memory. In order to reduce the memory consumption, the nonuniform scheme is adopted to quantitate channel LLRs and internal LLRs. In [20], it adopts the all-integer quantization method, where channel LLR is 4 bits and internal LLR is 5 bits. In this paper, an improved quantization scheme is proposed based on LLR distribution of different stages. At the beginning, the internal LLRs is small, and it is quantitated with the same bits as channel LLRs. To avoid catastrophic overflow, the internal LLRs of latter stages are quantitated with larger bits. Let (Q_i, Q_{cf}, Q_f) denote the quantization scheme, where Q_{cf} presents the
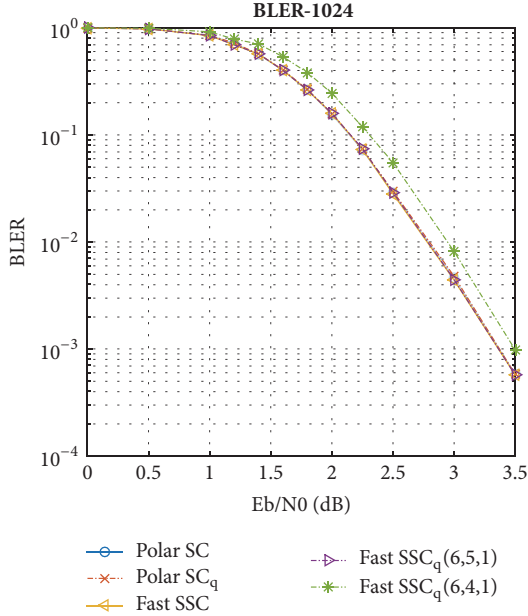
FIGURE 3: The BLER performance of different decoding schemes.

quantization bits of channel LLRs and that of LLRs for the former stages of the decoder, Q_i denotes the internal LLRs of other stages, and Q_f is the fractional bits. Figure 3 shows the block error rate (BLER) performance of SC, Fast-SSC algorithm, and different quantization schemes. The floating decoding performance of Fast-SSC is close to that of SC. In the quantization schemes of Fast-SSC, it can be seen that the performance of (6,5,1) quantization is close to the floating-point performance, but (6,4,1) quantization results in less than 0.2dB performance loss in high E_b/N_0. Therefore, this paper adopts (6,5,1) quantization scheme.

3. Architecture of Fast-SSC Decoder

The decoder is implemented in deeply pipelined architecture to improve the decoding throughput. This paper optimizes the PEs, storage, and control modules to lower the latency.

3.1. Architecture. The structure of Fast-SSC decoder is depicted in Figure 4, which consists of PE, memory, and controller. The PE is composed of various functions, such as F, G function, and Kronecker power module. Memory is divided into α_memory and β_memory, which are utilized to store the channel and internal LLRs and the hard decision of each constituent node, respectively. Because the decoding result of every constituent node needs to multiply G_N, the Kronecker power module contains G_N (N = 4, 8, 16, 32, 64) matrix for the leaf nodes of different length. The entire decoding process is manipulated by the controller module. When the channel LLRs signal (*en_cha_alpha*) is valid, the decoder starts to load one new frame into the decoder, and it outputs the codeword estimates. If the results of current stage are not used immediately by the next stage, it will be stored into α_memory or β_memory.

3.2. Deeply Pipelined. The deeply pipelined architecture for a (1024, 512) polar code is illustrated in Figure 4. The dotted lined rectangles represent the PEs such as REP128, where REP denotes the operation type; the subscript 128 represents the input length of the node. The spotted rectangles represent the RAMs, which are used to store the internal results to give the latter pipelined stages when the current results are not used immediately and the data is larger than 16. The solid lined rectangles represent registers. When the two-stage operation using a certain data is closer and the amount of data is small, the registers are used to store the data temporarily to reduce the memory control signal. The deeply pipelined architecture is designed according to the node activation order of the decoding tree and the operation order of the local decoding of each node. The software simulation which consists of 368 operations and the hardware implementation is a total of 330 operations. In order to achieve high throughput, this paper split the stages with larger latency. For example, REP128 can be split into four stages. The final architecture has 348 stages; thus the decoding delay is 348 clock cycles. Each stage in the pipeline contains one PE.

3.3. Memory. For node v in Figure 2, the inputs data α_v need to be used twice during the decoding process. Firstly, it is used to calculate α_{vl} of left child node; then it is utilized to calculate α_{vr} of right child node. Similarly, the local decoding result β_{vl} of node v_l needs to be input into Kronecker product module to obtain the final decoded words, and it also needs to calculate the local decoding results β_{vr} of its brother node v_r to obtain the result β_v of node v by C operation. The internal α and β need to be used twice in different stages, so they need to be stored. Since the bit widths of α and β are different, they are stored separately. The memory is divided into α_memory and β_memory as shown in Figure 4. If a node produces m internal data at t_1 clock and uses them at t_2, assuming $d = t_2 - t_1$, it will require $m \cdot (d + 2)$ memory unit. When the memory unit is less than 16, it can be stored with registers. Otherwise, we will use RAMs to store the internal results. The access timing of RAM is shown in Figure 5, where Adi denotes the i-th address of RAM. It is clear that the read sequence is the same as write sequence and they only differ d in clock cycles. Therefore, we can use FIFO to replace RAM.

3.4. Processing Elements. The main PEs of Fast-SSC decoder are listed in Table 2. F and $F\_with\_front\_complement$ are used to calculate the left child inputs of the activated node. G, G-0R, $G\_without\_complement$, $G\_without\_front\_complement$, and $G\_without\_latter\_complement$ can calculate the inputs of right child nodes. The C operation is used to combine the local decoding results of the left and right children nodes into the local decoding of the active node v. $R0$-$R1$, REP, SPC, and $R0\_SPC$ are the decoding operation of the corresponding type of leaf nodes, respectively. In addition to the 330 operations, there are two else stages operations. One exists in the first stage and the other one in the last stage. The first stage is used to cache the channel LLRs and occupies one clock. The results of the local nodes need to multiply G_N matrix to recover the local codeword. After the last constituent node
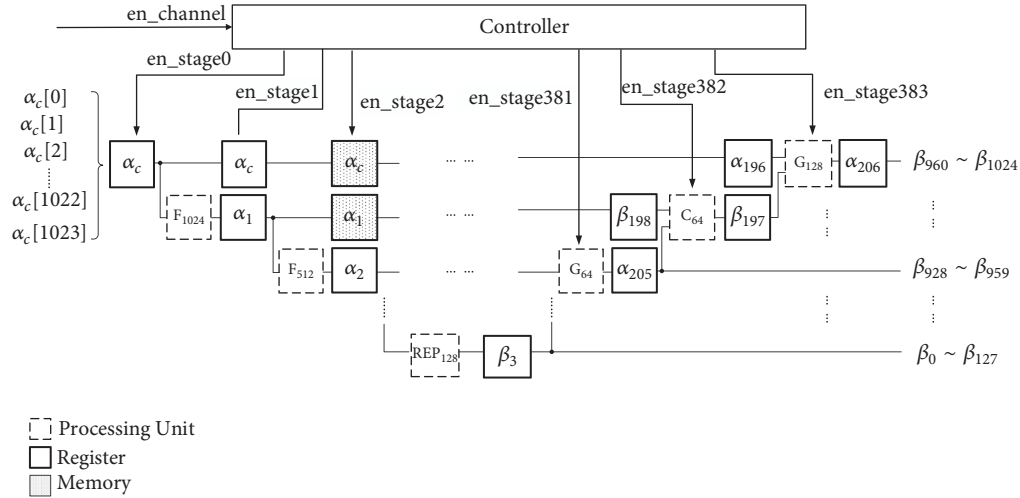
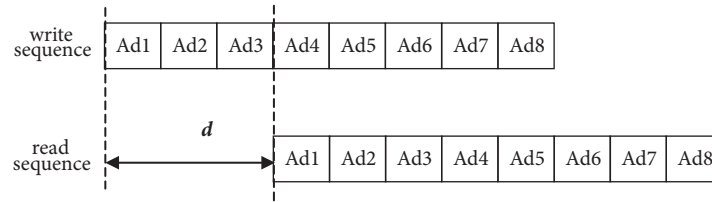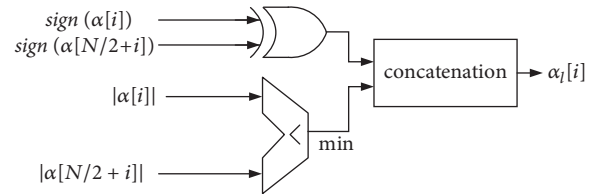FIGURE 4: The architecture of Fast-SSC decoder.

FIGURE 5: Access timing of RAM.

is computed, the converting operation by G_N for it will take one separate clock cycle. According to the above analysis, the deeply pipelined architecture has a total of 332 stages.

To implement the deeply pipelined architecture, we unfold the overall decoder. The *G-0R*, *R0-R1*, and *R0-SPC* operations are introduced to reduce the number of stages. The decoder can directly active the right child when the left child is Rate-0 node; thus it can reduce the decoding latency and the storage capacity. Moreover, in order to balance the pipeline at all stages and lower wire routing congestion, we refine the F and G operations into *F_with_front_complement*, *G_without_complement*, *G_without_front_complement*, and *G_without_latter_complement* operations when the inputs are large.

3.5. F Module. F operation is used to calculate the α_l of left child nods. According to (4), the sign bit of $\alpha_l[i]$ is obtained by XOR operation, and the numerical bits are the minimum of $\alpha_l[i]$ and $\alpha_l[i + 1]$ by compare operation. The structure is shown in Figure 6.

3.6. G Module. The G operation calculates α_r of the right child based on β_l of the left child and α_v of the parent node. According to (5), $\alpha[i]$ adds $\alpha[N/2 + i]$ when $\beta = 0$, and $\alpha[i]$ subtracts $\alpha[N/2 + i]$ when $\beta = 1$. The structure is shown in Figure 7.

When the input size of G is larger than 256, high decoding latency will be brought. It is clear that when the input length

FIGURE 6: Structure of F module.
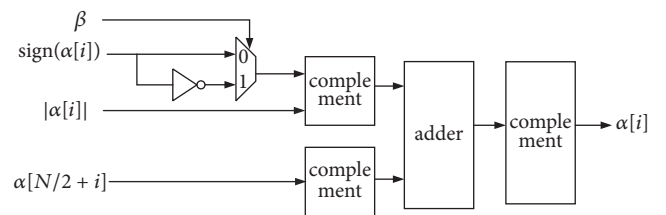
FIGURE 7: Structure of G module.

is long, the next stage after G is usually F operation. Since the complexity of F operation is less than that of G operation, to balance the frequency of the two operations, the complement operation of G is moved into the F operation. The optimized architecture is depicted as Figure 8. The left and right sides of the dotted line are the PEs of two stages, respectively. The complement operation of G colored in grey is performed in the next stage.

TABLE 2: The number of each operation for one frame.

Types	Description	numbers
F	Calculating α_l. (4)	82
F_with_front_complement	Calculating α_l, input complement.	7
G	Calculating α_r. (5)	78
G-0R	Calculating α_r. where $\beta_l = 0$	14
G_without_complement	Calculating α_r, input and output complement	3
G_without_front_complement	Calculating α_r, input complement.	4
G_without_latter_complement	Calculating α_r, output complement.	4
C	Calculating β_v. (6)	85
R0_R1	Decoding Rate-1 nodes, where $\beta_l = 0$.	3
REP	Decoding REP nodes	24
SPC	Decoding SPC nodes	23
R0_SPC	Decoding SPC nodes, where $\beta_l = 0$	3
Total		330
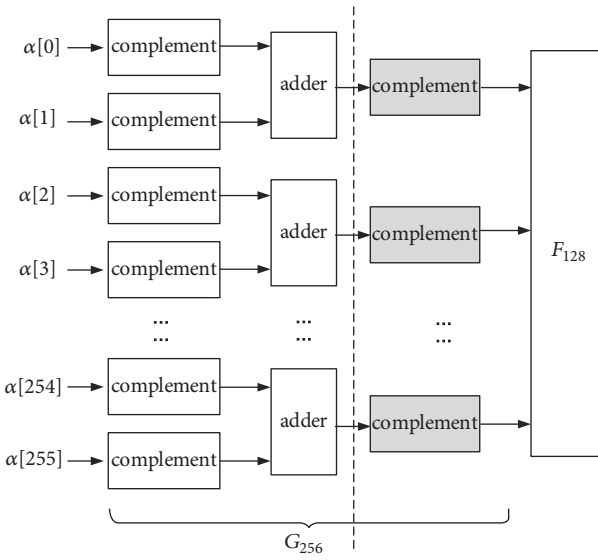
FIGURE 8: Optimized structure of G_{256}.

FIGURE 9: Structure of C module.
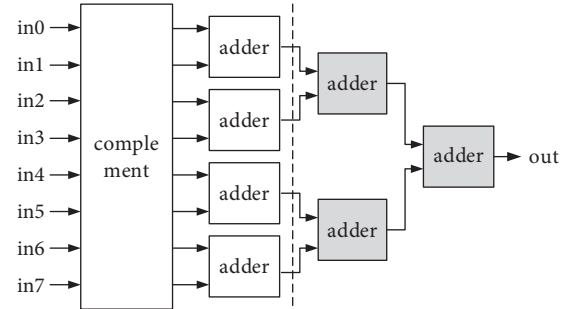
FIGURE 10: Optimized 8-input architecture of REP module.

3.7. C Module.

The C module combines β_l of left child and β_r of right child to calculate β_v of parent node. According to (6), the first half of β_r is obtained by β_l XOR β_r, and the latter half is equal to β_r directly. The structure of C module is shown in Figure 9.

3.8. REP Module.

The number of input ports in REP module is 4, 8, 16, 32, and 64. The 4-input module is the basic REP; other types can be decomposed into 4-input type. The hardware architecture for 8-input REP is presented in Figure 10. When the input length of REP nodes increases, the decoding latency also increases. To improve the working frequency, the 8-input REP is divided into two stages; thus it will use two clock cycles. The first stage translates the input data to complement and generates four internal results. The second stage adds them, and the REP module outputs the result of all the 8-inputs data. The hardware architecture is shown in

Figure 10. The dotted line indicates that the original 8-input REP module is divided into two stages.

Similarly, REP module of other lengths is divided into different stages. For instance, the 16-, 64-, and 128-input REP modules are divided into 2, 3, and 4 stages, respectively.

3.9. SPC Module.

In order to improve the frequency, the length of SPC node is constrained to 4 as shown in Figure 11. $|\alpha[i]|$ denotes the absolute value of $\alpha[i]$, and $\text{sign}(\alpha[i])$ is the sign bit of $\alpha[i]$. The 4MIN1 module selects the smallest LLRs through compare operation. *min01_flag* denotes the index of the minimum of $\alpha[0]$ and $\alpha[1]$, and *min23_flag* denotes the
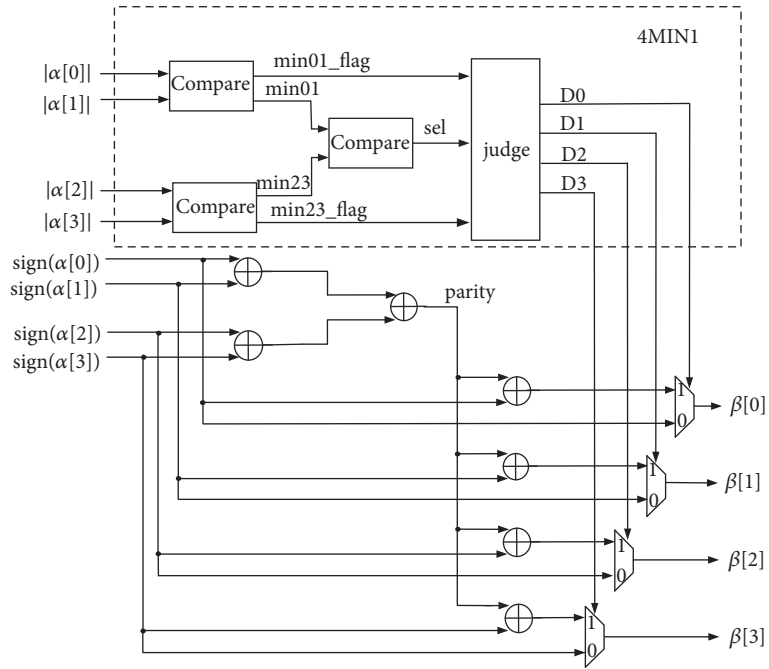
FIGURE 11: Structure of SPC module.
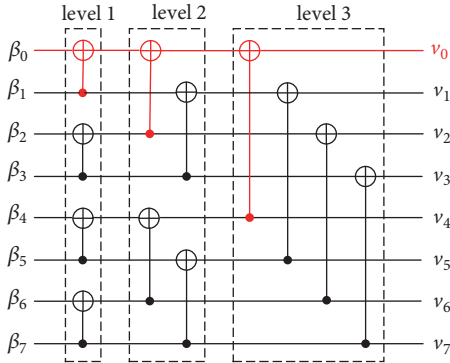
FIGURE 12: Architecture of Kronecker power module.
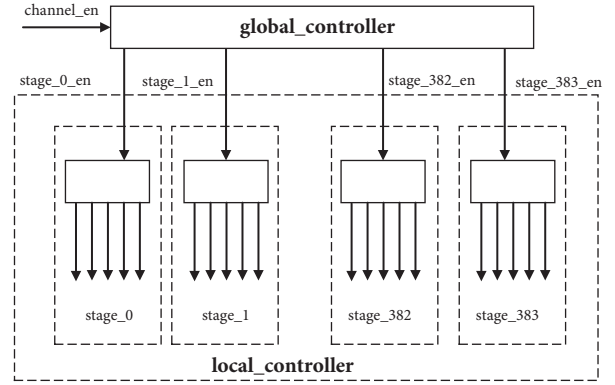
FIGURE 13: Two-level controller.

index of the minimum of $\alpha[2]$ and $\alpha[3]$. If $\alpha[0]$ is less than $\alpha[1]$ $min01\_flag$ is set to zero; otherwise $min01\_flag$ is one. If $min01$ is less than $min23$, sel is set to zero; otherwise sel is one. $D1\sim D3$ is determined by judge module. For instance, if sel and $min01\_flag$ are both zeroes, then $D0$ is set to one and others are set to zero. If sel is zero and $min01\_flag$ is one, then $D1$ is set to one and others are set to zero.

3.10. Kronecker Power Module. The decoding result of the constitute nodes requires the conversion of nth-Kronecker power to get the final result by (9). The code length of the architecture of the Kronecker power module is 8 as shown in Figure 12, where $\oplus$ denotes XOR operation and $\bullet$ denotes that the data is connected directly.

It can be found that if v_i is equal to zero, then it can be obtained as zero directly. As shown in Figure 12, the three XOR operations colored in red can be removed for v_0 is zero.

$$v_1^N = u_1^N F^{\otimes n}, \quad where\ N = 2^n,\ n \geq 0 \qquad (9)$$

3.11. Controller. The controller uses a two-level mode to generate control signals. As shown in Figure 13, the first level generates the global control signals, which assigns an enable signal to each stage to determine whether the corresponding stage works or not. For instance, if $stage1\_en$ is asserted, then stage1 is working; otherwise, stage1 is idle. The second level only generates the local control signals for each stage, such as
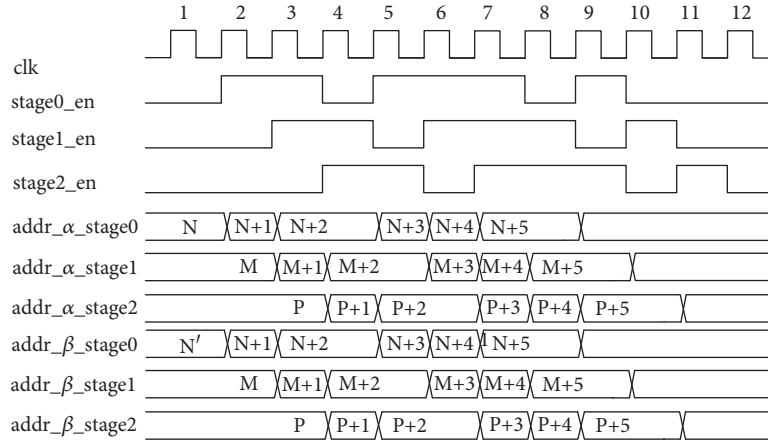
FIGURE 14: Timing of the controller.
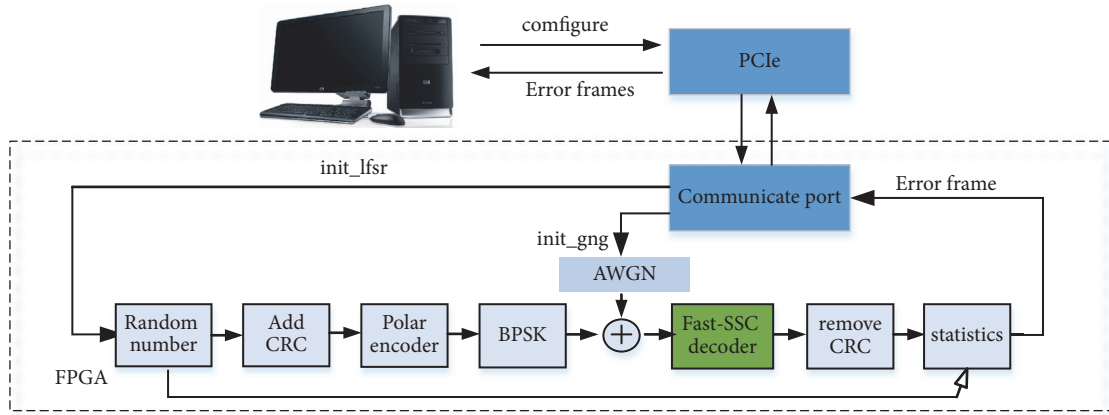
FIGURE 15: Test platform for polar decoder based on FPGA.

the address bus, data, and control signals of the memory, and the enable signals of PEs.

As depicted in Figure 14, *addr_α_stgae2* denotes the address bus of the *α_memory* in stage2. When *satge2_en* is asserted, the corresponding address adds one. Namely, an internal LLR is stored into *α_*memory_stage2. Compared with the one-level controller to generate all the control signals, it can reduce the implementation complexity of the controller.

4. Performance Analysis

4.1. Test Platform. To test the high throughput of Fast-SSC decoder, we implement a test platform based on FPGA. The overall platform including generating test data is completed on the FPGA to reduce the communication cost with the host computer. As depicted in Figure 15, the platform consists of random number generator, CRC check, polar encoder, BPSK AWGN, Fast-SSC decoder, and statistics module. PCIe is responsible for the communication between the host computer and FPGA platform. Meanwhile, this paper designs a software platform based on C++ that compares the results of hardware test and software simulation. At the beginning,

the host computer generates random number seeds, Gaussian noise seeds, the number of test frames, and start signal and transmits them to FPGA. When the decoding is completed, the statistics module uploads the number of error frames; then the host computer calculates the BLER and displays the test parameters. For the (1024, 512) polar code, simulations show that the test platform takes 19.18s at 300MHz to test data with $1.4*10^{10}$ bits.

4.2. Resource Consumption. The (1024, 512) polar decoder is implemented on Altera Stratix V 5SGXEA7N2F45C2 in Quartus II 15.0. The resources used by the decoder based on FPGA are shown in Table 3. It can be observed that the proposed decoder costs more memory compared with other decoder based on FPGA for that 6 bits is used to quantize the LLRs partly. However, it costs less registers compared with [20]. For the decoder in [24], it costs less resource because it does not adopt deeply pipelined architecture.

4.3. Performance. The latency and throughput are the main performance parameters of the polar decoder. Let *freq_decode* be the frequency of the decoder, and *frame_decoder_clocks*

TABLE 3: Statistics of resources.

	ALMs/LUTs	Register	Memory (bits)
Proposed	81498	96762	2367488
[20]	156450	152124	285120
[24]	29828	2332	18356

TABLE 4: Comparison with other polar decoders.

	This work	P. Giard [20]	Park [25]	Dizdar [26]	P. Giard [24]
Decoding. Algo.	Fast-SSC	Fast-SSC	BP	SC	Fast-SSC
IC type	FPGA	FPGA	ASIC	ASIC	FPGA
Tech(nm)	28	40	65	90	40
f(MHz)	300	231	300	2.5	80.6
Latency(us)	1.16	2.4	50	0.4	2.1
T/P(Gbps)	307.2	237	4.68	2.56	0.48

denote the number of clocks to decode one frame. The latency and throughput are calculated by

$$latency = \frac{1}{freq\_decoder} \tag{10}$$
$$* frame\_decoder\_clocks$$

$$Throughput = N * freq\_decoder \tag{11}$$

In this paper, for the (1024, 512) polar decoder, its working frequency can achieve 300MHz. The decoder requires 348 clocks to decode one frame. By (10) and (11), the latency is 1.16us, and the throughput is 307.2Gbps. Table 4 compares the proposed decoder with other polar decoders. In [20], a deeply pipelined decoder based on FPGA is capable of achieving the throughput over 237 Gbps for a (1024, 512) polar code. The latency of the decoder is twice more than this work. And the throughput of the proposed decoder is 1.3 times greater than that. It shows that either the latency or the throughput of this work is better than that in [20, 24, 25]. O. Dizdar and E. Arikan proposed a deeply pipelined polar decoder based on SC decoding algorithm. That decoder operates at lower clock frequency and costs less dynamic power. The proposed decoder has three times higher latency but is over 119 times faster than that in [26].

5. Conclusions

In this paper, a decoder in deeply pipelined architecture has been presented based on Fast-SSC decoding algorithm. The proposed decoder can output 1024 bits at each clock. To optimize the critical path, the PEs are decomposed and re-combined to balance the latency of two adjacent stages. The fixed-point nonuniform quantization scheme lowers storage capacity and obtains a good decoding performance. The two-level mode is proposed to reduce the complexity of the controller. Moreover, we build a platform based on FPGA to test its performance. Numerical results show that the decoder can achieve high throughput.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] W. Liu and X. Luo, "Localization Algorithm of Indoor Wi-Fi Access Points Based on Signal Strength Relative Relationship and Region Division , Computers," *Materials Continua*, vol. 55, no. 1, pp. 71–93, January 2018.

[2] Z. Xia, N. N. Xiong, A. V. Vasilakos, and X. Sun, "EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing," *Information Sciences*, vol. 387, pp. 195–204, 2017.

[3] H. Cheng, Z. Su, N. Xiong, and Y. Xiao, "Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model," *Information Sciences*, vol. 329, pp. 461–477, 2016.

[4] R. Meng, S. G. Rice, J. Wang et al., "A fusion steganographic algorithm based on faster R-CNN," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1–16, January 2018.

[5] H. Cheng, N. Xiong, A. V. Vasilakos, L. Tianruo Yang, G. Chen, and X. Zhuang, "Nodes organization for channel assignment with topology preservation in multi-radio wireless mesh networks," *Ad Hoc Networks*, vol. 10, no. 5, pp. 760–773, 2012.

[6] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[7] "IMT-2020(5G) PG White Paper on 5G Concept," http://www.imt-2020.org.cn/zh/documents/1?currentPage=2&content=.

[8] I. Tal and A. Vardy, "List decoding of polar codes," in *Proceedings of the 2011 Information Theory*, pp. 1–5, St. Petersburg, Russia, August 2011.

[9] K. Niu and K. Chen, "CRC-aided decoding of polar codes," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1668–1671, 2012.

[10] K. Niu and K. Chen, "Stack decoding of polar codes," *IEEE Electronics Letters*, vol. 48, no. 12, pp. 695–697, 2012.

[11] K. Chen, K. Niu, and J.-R. Lin, "Improved successive cancellation decoding of polar codes," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3100–3107, 2013.

[12] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Communications Letters*, vol. 16, no. 12, pp. 2044–2047, 2012.

[13] J. Snyders and Y. Be'ery, "Maximum likelihood soft decoding of binary block codes and decoders for the Golay codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 35, no. 5, pp. 963–975, 1989.

[14] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Communications Letters*, vol. 15, no. 12, pp. 1378–1380, 2011.

[15] G. Sarkis and W. J. Gross, "Increasing the throughput of polar decoders," *IEEE Communications Letters*, vol. 17, no. 4, pp. 725–728, 2013.

[16] C. Leroux, A. J. Raymond, G. Sarkis, and W. J. Gross, "A semi-parallel successive-cancellation decoder for polar codes," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 289–299, 2013.

[17] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast polar decoders: algorithm and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 946–957, 2014.

[18] C. Zhang and K. Parhi, "Low-latency sequential and overlapped architectures for successive cancellation polar decoder," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2429–2441, 2013.

[19] B. Yuan and K. K. Parhi, "Low-Latency Successive-Cancellation List Decoders for Polar Codes with Multibit Decision," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 10, pp. 2268–2280, 2015.

[20] P. Giard, G. Sarkis, C. Thibeault, and W. J. Gross, "237 Gbit/s unrolled hardware polar decoder," *IEEE Electronics Letters*, vol. 51, no. 10, pp. 762-763, 2015.

[21] Y. Li and R. Liu, "High throughput GPU polar decoder," in *Proceedings of the 2nd IEEE International Conference on Computer and Communications, ICCC 2016*, pp. 1123–1127, Chengdu, China, October 2016.

[22] S. Cammerer, B. Leible, M. Stahl, J. Hoydis, and S. Ten Brink, "Combining belief propagation and successive cancellation list decoding of polar codes on a GPU platform," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2017*, pp. 3664–3668, New Orleans, La, USA, March 2017.

[23] X. Han, R. Liu, Z. Liu, and L. Zhao, "Successive-cancellation list decoder of polar codes based on GPU," in *Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 2065–2070, Chengdu, December 2017.

[24] F. Ercan, C. Condo, and W. J. Gross, "Reduced-memory high-throughput fast-SSC polar code decoder architecture," in *Proceedings of the 2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 1–6, Lorient, France, October 2017.

[25] Y. S. Park, Y. Tao, S. Sun, and Z. Zhang, "A 4.68Gb/s belief propagation polar decoder with bit-splitting register file," in *Proceedings of the 2014 IEEE Symposium on VLSI Circuits*, pp. 1-2, Honolulu, Hawaii, USA, June 2014.

[26] O. Dizdar and E. Arikan, "A high-throughput energy-efficient implementation of successive cancellation decoder for polar codes using combinational logic," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 3, pp. 436–447, 2016.

[27] A. Balatsoukas-Stimming, A. J. Raymond, W. Gross, and A. Burg, "Hardware Architecture for List SC Decoding of polar codes," https://arxiv.org/abs/1303.7127.

Research Article

Adaptive Transmission Range Based Topology Control Scheme for Fast and Reliable Data Collection

Haojun Teng [1], **Kuan Zhang,**[2] **Mianxiong Dong** [3] **Kaoru Ota,**[3] **Anfeng Liu** [1,4] **Ming Zhao,**[5] **and Tian Wang**[6]

[1] *School of Information Science and Engineering, Central South University, Changsha 410083, China*
[2] *Department of Electrical and Computer Engineering, University of Nebraska, Lincoln, USA*
[3] *Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan*
[4] *The State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*
[5] *School of Software, Central South University, Changsha 410083, China*
[6] *School of Computer Science, National Huaqiao University, Quanzhou 362000, China*

Correspondence should be addressed to Anfeng Liu; afengliu@mail.csu.edu.cn

An Adaptive Transmission Range Based Topology Control (ATRTC) scheme is proposed to reduce delay and improve reliability for data collection in delay and loss sensitive wireless sensor network. The core idea of the ATRTC scheme is to extend the transmission range to speed up data collection and improve the reliability of data collection. The main innovations of our work are as follows: (1) an adaptive transmission range adjustment method is proposed to improve data collection reliability and reduce data collection delay. The expansion of the transmission range will allow the data packet to be received by more receivers, thus improving the reliability of data transmission. On the other hand, by extending the transmission range, data packets can be transmitted to the sink with fewer hops. Thereby the delay of data collection is reduced and the reliability of data transmission is improved. Extending the transmission range will consume more energy. Fortunately, we found the imbalanced energy consumption of the network. There is a large amount of energy remains when the network died. ATRTC scheme proposed in this paper can make full use of the residual energy to extend the transmission range of nodes. Because of the expansion of transmission range, nodes in the network form multiple paths for data collection to the sink node. Therefore, the volume of data received and sent by the near-sink nodes is reduced, the energy consumption of the near-sink nodes is reduced, and the network lifetime is increased as well. (2) According to the analysis in this paper, compared with the CTPR scheme, the ATRTC scheme reduces the maximum energy consumption by 9%, increases the network lifetime by 10%, increases the data collection reliability by 7.3%, and reduces the network data collection time by 23%.

1. Introduction

Topology control technology has been widely applied to wireless networks such as mobile ad hoc network [1–3], wireless sensor network, and emerging mobile networks (EMN) to reduce the communication interference and energy consumption through altering the underlying network [4–8]. Topology control technology mainly adjusts the transmission power of communication devices to adjust their communication radius so as to achieve the goal of improving network performance. Topology control technology can be fully applied in wireless sensor networks based on the fact

that the transmission power of a wireless sensor node is adjustable. The paper [9] pointed out that the sensor nodes have multiple power levels which can be adjusted. Topology control technology brings benefits to the network. (1) Reducing the energy consumption of nodes [10–14]: in wireless sensor networks, the density of communication devices is usually large [15]. When communication between devices to devices (D2D) is required, the transmission power can be adaptively adjusted according to the distance between devices (or nodes). Since the energy consumption of nodes is related to the communication distance to the second power (even to the fourth power), the adaptive adjustment of the

transmission power of nodes according to the communication distance can effectively reduce the energy consumption of the nodes and extend the network lifetime. (2) Reducing network communication interference, thereby increasing the number of simultaneous network communications, improving network throughput, and at the same time reducing energy consumption [16–18]: if the nodes density is relatively high, appropriately reducing the communication radius of nodes will not only reduce the energy consumption, but also reduce the wireless communication interference between the nodes. Therefore, more nodes can communicate at the same time, thereby increasing the network's throughput.

Most of the topology control studies are based on the ideas mentioned above, that is, adaptively adjusting the communication radius of the nodes, and meanwhile satisfying the communication, the nodes communicate with the smallest possible communication radius. This can reduce energy consumption and interference. Increase throughput improves network life. These studies can be found in [10, 19–21]. Although the main idea of topology control has been widely used in most networks, it leads to good results. However, we have found that for a type of delay and loss sensitive data aggregation wireless sensor networks, the idea mentioned above cannot lead to good results. In some applications, wireless sensor networks need to periodically collect information such as crop temperature and humidity. The users of these applications need to know the minimum value or maximum value rather than every detail value. Therefore, using data aggregation can minimize the volume of data to transmit. In such data collection, any n data packets will be aggregated into one data packet. But the wireless channel is usually lossy due to the complexity of the physical environment. Less delay time of data collection is important for these applications. But more delay time will be caused by the packet loss if the wireless channel is lossy. Therefore, how to adopt the appropriate topology control technology in such applications to achieve fast, efficient, and highly reliable data collection is a challenging issue [22, 23].

Using broadcast method can effectively improve the reliability of data collection [24]. When the sender node transmits data by broadcasting, all the nodes which are in the transmission range of the sender node will receive the data possibly. Only when all the nodes fail to receive the data, the data will be lost. Obviously, this approach can significantly reduce the probability of packet loss but increases energy consumption. However, compared to retransmission method in unicast mode, broadcast still has the following advantages: (a) the sender node only needs to broadcast once and it may achieve high reliability. But the data loss in unicast mode requires multiple retransmissions. (b) Broadcast has a great advantage in data collection speed. In unicast mode, if the packet is lost, retransmission is required, and the retransmission time is very long. But broadcast mode usually does not need retransmission, so it reduces the transmission time.

There are a few studies based on the collection of broadcast data for delay and loss sensitive wireless sensor network. But according to our investigation, there is no topology control method based on broadcast data collection mode. The previous topology control method is to reduce

the transmission radius of the nodes as much as possible to save energy and reduce the interference range. However, topology control in this article takes the exact opposite of the previous topology control strategy. In the topology control of this paper, instead of reducing the transmission radius of the nodes, the remaining energy in the network is fully utilized to increase the transmission radius of the nodes to improve the reliability of data collection and the data collection performance, while ensuring a high network lifetime. In summary, the main contributions of this article's research work are as follows:

(1) An Adaptive Transmission Range based Topology Control (ATRTC) scheme is proposed to reduce delay and improve reliability for data collection in delay and loss sensitive wireless sensor network. In contrast to the previous reduction of communication radius to reduce interference and save energy, the core idea of the ATRTC scheme is to increase the radius of broadcast to speed up the collection of data and improve the reliability of data collection. The reason for increasing the radius of broadcast to optimize the network performance is as follows: the expansion of the transmission range will allow the data packet to be received by more receivers, thus improving the reliability of data transmission. On the other hand, by extending the transmission range, data packets can be transmitted to the sink with fewer hops. Thereby the delay of data collection is reduced and the reliability of data transmission is improved. Extending the transmission range will consume more energy. Fortunately, we found the imbalanced energy consumption of the network. There is a large amount of energy remains when the network died. ATRTC scheme proposed in this paper can make full use of the residual energy to extend the transmission range of nodes. Because the expansion of transmission range, nodes in the network form multiple paths for data collection to the sink node. Therefore, the volume of data received and sent by the near-sink nodes is reduced, the energy consumption of the near-sink nodes is reduced, and the network lifetime is increased as well.

(2) According to the analysis in this paper, compared with the CTPR scheme, the ATRTC scheme reduces the maximum energy consumption by 9%, increases the network lifetime by 10%, increases the data collection reliability by 7.3%, and reduces the network data collection time by 23%.

The rest of the paper is organized as follows. Section 2 reviews related works comparing with our scheme. Section 3 describes the network model and defines problem statements of this paper. In Section 4, we give the design ATRTC scheme for delay and loss sensitive WSNs. In Section 5, we do performance analysis to ATRTC scheme. We conclude this paper in Section 6.

2. Related Work

Advances in technology have made sensors smaller, more accurate, and less expensive. Thanks to this, wireless sensor networks have been widely used in various fields [25–30]. Because building a sensor network requires a large number of sensor nodes, sensor nodes are usually cheap. Due to the cost, the battery, CPU, memory, and communication capabilities of

sensor nodes are limited. The main goal of the optimization of the sensor network is to increase the lifetime and reliability of the sensor network and reduce the sensor delay [25, 31–33].

Topology control technology is an important and commonly used optimization technique in networks fields [34–37]. Topology control technology has considerable studies in WSNs [38, 39]. At present, studies on topology control in WSNs can be mainly divided into network coverage issues and network connectivity issues. And network connectivity issues include Power Management Mechanisms and Power Control Mechanisms [9]. This paper focuses on the Power Control Mechanisms. In general, sensor nodes such as the MICA2 mote have multiple power levels adjustable. By such a fact, one natural question is the following: Is it wise to make radio work with the maximum working power all the time? Research literature later found that the answer is negative. The optimal transmission power is normally between minimum value and the maximum value. The paper [19] proposed a COMPOW Protocol. It can determine the minimum common transmitting range needed to ensure network connectivity. The authors believe that the two nodes that are closer together do not need the maximum transmission power. Reducing the transmission power to reduce the transmission range has two advantages. On the one hand, it can increase the battery lifetime. On the other hand, reducing the transmission radius can reduce the interference of neighbor nodes. It can increase the number of nodes that can transmit at the same time, thereby increasing the network capacity. But in reality, link is not always reliable, and it is difficult to guarantee the link quality under the transmission power of COMPOW. A large number of experimental results in [40] proved that the transmission power is related to the link quality. The relationship of transmission power and received signal strength indicator (RSSI)/link quality indicator (LQI) are analyzed, and a prediction model is established. The authors also proposed the Adaptive Transmission Power Control (ATPC) algorithm, a feedback-based transmission power control algorithm to dynamically maintain individual link quality over time. This algorithm achieves energy efficiency and guarantees the link quality between neighbors. However, the implementation of this algorithm also requires additional data exchanges between neighbor nodes. Therefore, although ATPC has improved in energy saving, the network delay will be increased.

For the issue of link quality, [41] conducted deep research. Marco and Bhaskar use mathematical techniques from communication theory to model and analyze low power wireless links. And they analyzed how multiple factors affect link quality. The paper [41] measured the link quality using packet reception rate (PRR) and introduced the relationship between PRR and distance under certain transmission power conditions. Based on this, this paper analyzes the relationship between transmission distance and the power under the condition of guaranteeing link quality. In applications requiring high reliability, it is not wise to simply increase the link quality by increasing the transmission power. Because the relationship between transmit power and PRR is nonlinear, in the case of relatively high PRR, a huge transmission power improvement can only be obtained in exchange for a slight PRR improvement. In the application of high reliability requirements, the following methods are generally used to improve the reliability of the transmission:

(1) Broadcast technology: broadcast technology is a highly reliable data transmission technology. It allows nodes to transmit data to multiple destination nodes at the same time. As long as one destination node receives data, it can continue to transmit. Therefore, even if the quality of links between nodes is relatively low, the probability of successful transmission in the case of a large number of destination nodes can reach a very high level. Of course, broadcast technologies also have drawbacks. Uncontrolled broadcast technologies such as flooding can significantly increase the amount of data, cause a lot of energy waste, and shorten the network lifetime. This is called broadcast problems [42]. There are many papers on the research of broadcast storm problems and propose improvement strategies. The paper [43] introduces a location aided algorithm to compute the optimal local cover set without delay and without much communication overhead. The approach in [44] is based on selecting a small subset of nodes to form a forward node set to carry out a broadcast process. These methods in [43, 44] require each node to know its k-hop (k >=2) neighbor information, which cause a large amount of extra information exchange, which generates additional energy consumption, so the improvement of energy efficiency is not obvious. An important solution to the broadcast storm problem is to use data aggregation technology. The paper [45] shows that appropriate use of in-network aggregation can significantly reduce the amount of traffic generated over the network. Data aggregation technology is a very important technology to reduce the data volume in wireless sensor networks [46, 47]. There are many researches on the application of data aggregation in wireless sensor networks. The paper [48] adopted the data aggregation technology in industrial wireless sensor networks (IWSNs). And it used cluster heads to reduce the amount of transmissions required to transmit the same information to the BS and reduces the latency at the expense of some reduction in performance. The paper [49] proposed distributing delay efficient data aggregation scheduling to reduce the delay effectively, in which data aggregation is an essential operation.

(2) Retransmission mechanism: in order to achieve high reliability, using the retransmission mechanism is the most direct choice. In the process of data transmission between the node and the destination node, if there is a packet loss, the destination node will send a message to the source node to inform the packets loss. Then, the source node will retransmit the lost packets. If the destination node retransmits successfully, it will return an ACK message. Otherwise, the source node will continue to retransmit and stop retransmitting until it succeeds or reaches the upper limit of the number of retransmissions. Many studies have achieved robust and reliable wireless transmission by using a retransmission mechanism. The papers [50, 51] are examples of using retransmission mechanisms to solve reliability problems. The paper [50] proposes PSFQ (Pump Slowly, Fetch Quickly), a reliable transport protocol suitable for a new class of reliable data applications in wireless sensor networks.

The paper [51] proposes a scalable framework for reliable downstream data delivery that is specifically designed to both address and leverage the characteristics of a wireless sensor network, while achieving the reliability in an efficient manner.

The use of the retransmission mechanism is energy efficient compared to the broadcast mechanism and the following redundancy mechanism, but there are also problems with large delays, because the retransmission mechanism requires multiple transmissions and an acknowledgment message which lead to additional wait time.

(3) Redundancy scheme is also a way to improve reliability [52]. This scheme is usually using Erasure Coding, that is, when transmitting data, encode M source packets into $N+R$ ($N+R>M$) packets for transmission. At the destination node, it can reconstruct M original messages if the destination node receives at least N out of $N+R$ encoded data packets [53]. Theoretical analysis in [53] indicates that when the loss probability remains low or moderate, redundancy scheme is more reliable and energy efficient than retransmission. However, the performance of the redundancy scheme will greatly decline under the condition that the probability of loss is high.

3. The System Model and Problem Statement

3.1. The Network Model. A wireless sensor network for periodic data collection is adopted in this paper. This network is composed of a sink node and n sensor nodes. The sensor node in the network is responsible for collecting the physical characteristics of the surrounding environment and generating data. On the other hand, the sensor node is also responsible for data transmission and data aggregation. The sink node in the network is responsible for collecting all incoming data and processing it for sending to users using the wireless sensor network. This kind of network has widely application scenarios, such as predicting the weather by collecting wind information periodically and monitoring forest fires by collection temperature information periodically. For the deployment of the network, the coverage of the entire network is a circle with radius R. The sink node is located in the center of the network, and n sensor nodes are randomly and uniformly deployed in the entire network range with a distribution density of ρ.

3.2. The Wireless Channel Reliability Model. Due to the complexity of the physical environment, the communication links between nodes are lossy. This paper uses the packet receiving rate proposed by [41] to measure the quality of communication links between nodes. For a data transmission process, the sender is the node that sends the data, and receiver is the node that receives the data. The packet acceptance rate of a node is defined as the ratio of the number of data received by the receiver to the number of data sent by the sender node. According to [41], the packet reception rate (PRR) is calculated as follows:

$$p = \left(1 - \frac{1}{2}\exp^{-(SNR/2)(B_N/V_s)}\right)^{8f} \tag{1}$$

where f is the frame size of the packet, B_N is the noise bandwidth, and V_s is the sending rate of nodes. In general,

TABLE 1: Parameters of sensor node.

Symbol	Description	Value
V_s	Sending rate	19.2$kbps$
V_r	Receiving rate	19.2$kbps$
E_{init}	The initial energy of node	2J
d_0	The reference distance	5m
P_n	The noise floor	-115 dBm
n	The path loss exponent	4
f	The frame size of a data packet	50B
B_N	The noise bandwidth	30HZ

the values of f, B_N, and V_s are fixed in the same network, and their specific values are summarized in Table 1. Substituting specific values of f, B_N, and V_s into (1), the equation of p with respect to $\gamma(d)$ can be expressed as

$$p = \left(1 - \frac{1}{2}\exp^{-(\gamma(d)/2)(1/0.64)}\right)^{8f} \tag{2}$$

Assuming a given transmit power P_t, the signal-to-noise ratio γ at the distance d can be expressed as

$$\gamma(d) = P_t - PL(d) - P_n \tag{3}$$

where Bottom noise P_n is -115 dBm and $PL(d)$ is the loss value during transmission.

The value of $PL(d)$ can be derived from the following equation:

$$PL(d) = PL(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right) + X_\delta \tag{4}$$

where d_0 is a reference distance, n is the path loss exponent, and X_δ is a zero-mean Gaussian distribution of the random variable.

3.3. The Data Collection Model. Wireless communication technology is adopted during the data transmission between nodes in the network. The routing algorithm used for the network is the shortest path algorithm. If the sink node is within the transmission range of the node, the data can be directly transmitted to the sink node. For a node that is far away from the sink node, the data need to be transmitted to a node that is closer to the sink node and forwarded by this node to a node closer to the sink node until it reaches the sink. This type of transmission is called multihop transmission. In the process of data transmission to the sink node, the number of transmissions is called the number of hops. For nodes with the same number of hops, their distribution positions appear as a ring, which we call a tier.

If all the nodes in the network have the same transmission range, the network can be divided into several tiers with the same width. If the ring width is $d(0 < d \le r)$, the network can be divided into $\lceil R/d \rceil$ tiers. From the center to number each tier in ascending order, 1, 2, 3,$\cdots$, $\lceil R/d \rceil$, and T_i denoted the nodes in tier i.

The data collection in the network refers to the process of collecting and processing the data packets generated by all the
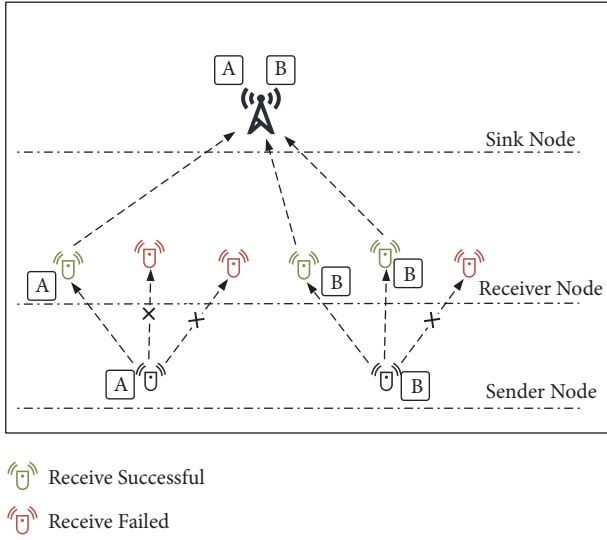
Figure 1: Broadcast mechanism.

Figure 2: Data aggregation mechanism.

sensor nodes in the network and collected in the sink node. The entire data collection process begins with the node in the outermost tier, from the outside to the inside, and the data are collected tier by tier according to a certain collection method until all the data reach the sink node.

The broadcast mechanism is adopted for data transmission in order to improve the reliability. By using the broadcast mechanism, sender node can send data to multiple receiver nodes simultaneously. As long as any node in the receiver nodes receives the data sent by the sender, it can ensure that the data continue to be transmitted without loss. As can be seen from Figure 1, there are two sender nodes sending data. But the data sent by them are not received by all receiver nodes. Although the data sent by the sender on the left is only received by one receiver node, the data can still be transmitted to the sink node. Therefore, using broadcasting can greatly improve the reliability of transmission. In general, the sink node has multiple antennas; therefore, its children can be seen as having multiple parent nodes.

Using broadcast mechanism improves the reliability of data transmission, but it dramatically increases the amount of data transmission as well. As can be seen from Figure 1, after a data packet B sent by a sender node is received by two receiver nodes, both receiver nodes will send B to their parent node (the sink node); i.e., their parent node receives duplicate data causing energy waste. It is obviously not feasible to simply use the broadcast mechanism. Therefore, data aggregation technology is adopted in the process of network transmission. In the process of data collection, the node can use the data aggregation technology to process the received data, and the data size after processing will be greatly reduced.

Data aggregation technology is a technology used in data transmission to reduce the amount of data transmitted. As shown in Figure 2, the data aggregation technology uses a specific aggregate function to process the m input data packets and outputs an aggregated data packet. The size of the output data packet is significantly reduced compared to
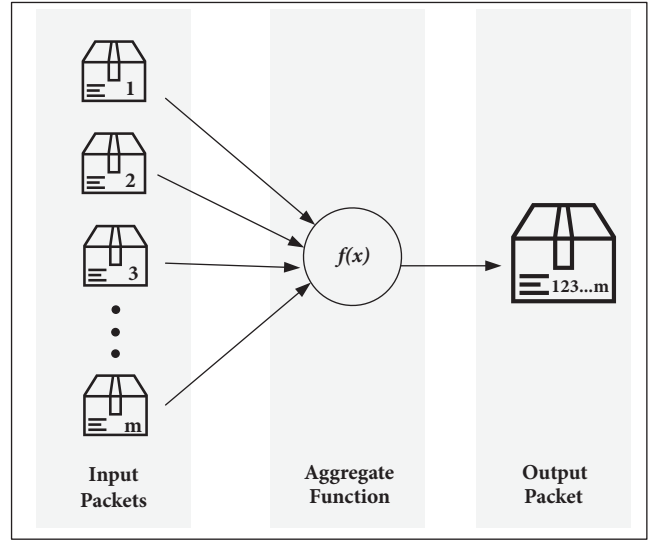
the total size of the input data packet. In this paper, a widely used aggregation model called increasing aggregation model is adopted, same as [54]. When m raw packets are aggregated, the size of the aggregated packet can be calculated according to the following equation:

$$\left|\vartheta'\right| = \left|\vartheta_1\right| + \varepsilon \cdot \left(\left|\vartheta_2\right| + \left|\vartheta_3\right| + \cdots + \left|\vartheta_m\right|\right) \qquad (5)$$

where $|\vartheta'|$ denotes the size of output packet after aggregation and $|\vartheta_1|, |\vartheta_2|, \cdots, |\vartheta_m|$ denote the size of m raw input packets, respectively.

3.4. Problem Statement. For wireless sensor networks, how to achieve better network performance under limited resources is of great importance. For periodic data collection WSNs, improving the reliability of data collection, how to reduce the length of data collection, and how to improve the network life are the main aspects of our efforts to improve.

(1) Data collection time ($\mathscr{T}$): data collection time is the time length for a data collection cycle. It is the time length from which the outermost node starts to transmit data to all nodes in the network to complete the data transmission. The longer the data collection time is, the longer it takes the user to get the latest data. Obviously, the shorter the data collection time, the better. Let t_i denote the transmission time of all the nodes in tier i, and the data collection time ($\mathscr{T}$) can be expressed as the sum transmission time of nodes in all tiers:

$$\min\left(\mathscr{T}\right) = \min\left(\sum_{i=0}^{k} t_i\right) \qquad (6)$$

(2) Data collection reliability ($\mathscr{Q}$): the data collection reliability refers to the probability that all the data transmitted from nodes in all tiers reach the sink node without loss after a data collection cycle is completed. In the application of periodic data collection WSNs, no data loss will affect the result. However, using data aggregation increases the amount

of useful data contained in a packet. Therefore, the probability of packet loss affecting the final result increases. Obviously, improving the reliability of data collection is very important. The greater reliability ($\mathcal{Q}$) of data collection, the better:

$$\max(\mathcal{Q}) = \max\{P(all\ data\ transmit\ to\ sink\ node)\} \quad (7)$$

(3) Network lifetime ($\mathcal{L}$): the lifetime of the network refers to the time when all nodes in the network normally perform the function. Since the WSN is an ad hoc network, the death of any node in the network will affect the performance of the network. Therefore, the network lifetime is generally defined as the lifetime of the first dead node in the network. Since a periodic data collection network is adopted in this paper, the network lifetime can be represented by the data collection cycles of the network until the first node dies. Obviously, the longer the life of the network, the better. Let E_c denote the energy consumption of a node that consumes the highest energy in the network during a data collection cycle, and E_{init} denote the initial energy of the node. The network lifetime can be expressed as follows:

$$\max(\mathcal{L}) = \max\left(\frac{E_{init}}{E_c}\right) \quad (8)$$

To sum up, the goal we want to achieve can be expressed by the following expressions:

$$
\begin{aligned}
\min\quad &(\mathcal{T}) = \min\left(\sum_{i=0}^{k} t_i\right) \\
\max\quad &(\mathcal{Q}) \\
&= \max\{P(all\ data\ transmit\ to\ sink\ node)\} \quad (9) \\
\max\quad &(\mathcal{L}) = \max\left(\frac{E_{init}}{E_c}\right) \\
\text{s.t.}\quad &\mathcal{Q} \geq \mathcal{Q}_{std},\ \mathcal{T} \leq \mathcal{T}_{std}
\end{aligned}
$$

In (9), $\mathcal{Q}_{std}$, $\mathcal{T}_{std}$ represent the minimum standard required by the user. The goal of (9) is to improve the performance as much as possible, while ensuring that $\mathcal{T}$ and $\mathcal{Q}$ reach the minimum standard, while maximizing the network lifetime as much as possible.

4. The Design of ATRTC Scheme

4.1. Research Motivation. There are two main aspects of motivation for this paper: On the one hand, if the transmission power of node can be increased, the transmission range of the node will extend and the PRR does not decrease. On the other hand, in the wireless sensor network under the traditional scheme, the energy consumption of a small part of the nodes near the sink node (near-sink nodes) is much greater than that of the most of the nodes far from the sink node. This leads to the energy remaining of most nodes after the death of the network.

If transmission power of a node can be increased, then the transmission range of node can be extended and PRR
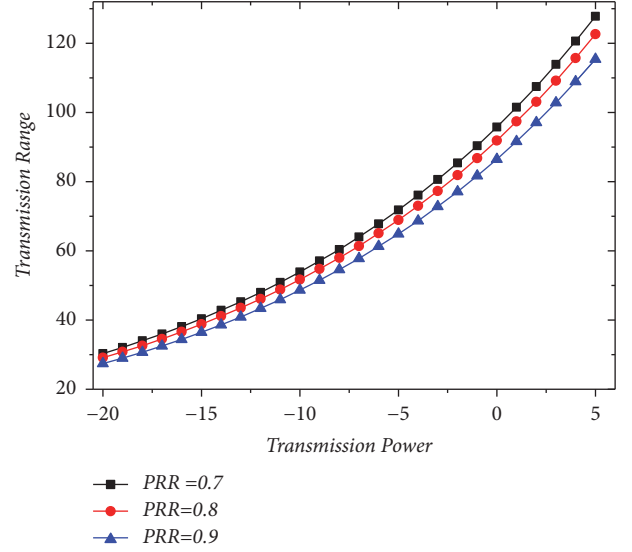

Figure 3: Transmission range with transmission power.

will not decrease at the same time. If the transmission range of the node can be extended to a certain level, the number of hops that node transmit data to sink can be reduced. That means the data reaching sink node need less forwarding times than before which can reduce the transmission time. Since the wireless channel between nodes is lossy, there is a possibility of packet loss during data transmission. During the transmission of a data packet from a node to the sink node, the probability of packet loss increases with each hop. Since extending the transmission range of node will reduce the number of hops, the possibility of packet loss will be reduced during the transmission to the sink node. From the perspective of the entire network, extending the transmission range improves the data collection reliability. Unfortunately, increasing the transmission power will increase energy consumption and shorten the lifetime of the node. It is unwise to increase the transmission power blindly.

In general, there are different power levels that can be adjusted in a wireless sensor node; for example, MICA2 has 26 transmit power levels (-20*dBm* to 5*dBm*). According to the reliability model in Section 3.2, the relationship between transmission range and transmission power under different PRR can be obtained. In Figure 3, the higher the transmission power, the large the transmission range, under the situation of the same PRR. And at the same transmission power, the higher the PRR, the smaller the transmission range. Assume that the network used in this paper requires the PRR to be 0.868. The maximum transmission range that can be achieved under different power levels is summarized in Table 2.

As shown in Table 2, when the transmission power is -20*dBm*, the transmission range is 28 *m*. Because -20*dBm* is the minimum transmission power of the node, 28 *m* is the minimum transmission range that guarantees PRR = 0.868. When the transmission power is -8 *dBm*, the transmission range of the node can reach 55.9*m*, which is about twice the minimum transmission range. And when the transmission

TABLE 2: The transmission range under PRR=0.868.

Transmission power	Range
-20 dBm (0.0100 mW)	28.0 m
-19 dBm (0.0126 mW)	29.7 m
-18 dBm (0.0158 mW)	31.4 m
-17 dBm (0.0200 mW)	33.3 m
-16 dBm (0.0251 mW)	35.3 m
-15 dBm (0.0316 mW)	37.4 m
-14 dBm (0.0398 mW)	39.6 m
-13 dBm (0.0501 mW)	41.9 m
-12 dBm (0.0631 mW)	44.4 m
-11 dBm (0.0794 mW)	47 m
-10 dBm (0.1000 mW)	49.8 m
-9 dBm (0.1259 mW)	52.8 m
-8 dBm (0.1585 mW)	55.9 m
-7 dBm (0.1995 mW)	59.2 m
-6 dBm (0.2512 mW)	62.7 m
-5 dBm (0.3162 mW)	66.4 m
-4 dBm (0.3981 mW)	70.3 m
-3 dBm (0.5012 mW)	74.5 m
-2 dBm (0.6310 mW)	78.9 m
-1 dBm(0.7943 mW)	83.6 m
0 dBm (1.0000 mW)	88.6 m
1 dBm (1.2589 mW)	93.8 m
2 dBm (1.5849 mW)	99.4 m
3 dBm (1.9953 mW)	105.2 m
4 dBm (2.5119 mW)	111.5 m
5 dBm (3.1623 mW)	118.1 m

FIGURE 4: Energy consumption in one cycle.

TABLE 3: Parameters of network.

Symbol	Description	Value
N	total number of node in network	500
R	radius of network	126 m
d	width of tier	21 m
ε	aggregation ratio	0.3
r	initial transmission range of node	28 m
P^{send}	initial transmission power of node	-20 dBm
$P^{receive}$	receiving power of node	-20 dBm

power is $5dBm$, the range can reach $118.1m$, which can reach more than 4 times the minimum range.

(2) Under the traditional Constant Transmission Power and Range (CTPR) scheme, there is a big difference in energy consumption of nodes in different locations in the network. Because the nodes close to sink node need to bear forwarding a large amount of data sent from far-sink nodes, the energy consumption of near-sink node is much great than that of far-sink node. Because of the death of one node in the wireless sensor network, it will have a great impact on the network, so that the entire network cannot complete the predetermined function. So, once a node dies, it is regarded that the entire network is dead. At this time, far-sink nodes still have a lot of energy left. In other words, there is still a large amount of energy in the network that can be used after the network died.

We analyzed a network consisting of 500 wireless sensor nodes and a sink node using traditional Constant Transmission Power and Range (CTPR) scheme. Some specific parameters of this network are summarized in Table 3.

We have calculated the energy consumption of each node of the network in one data collection cycle and the remaining energy of each node after the network died. As shown in Figure 4, the energy consumption decreases from the near-sink node to the far-sink node under the CTPR scheme. The trend of e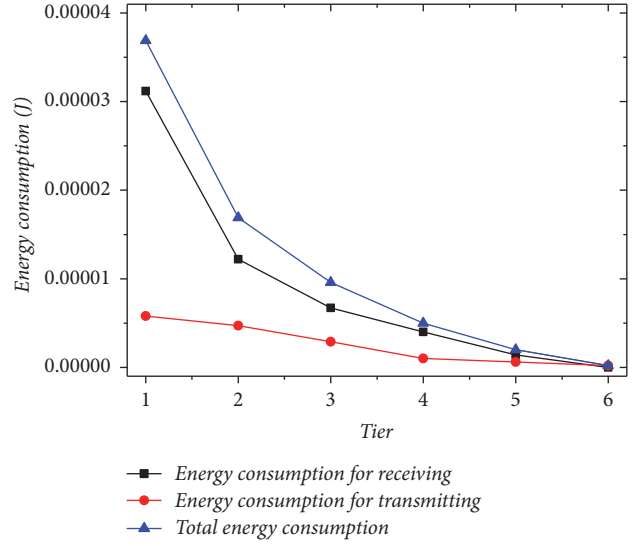nergy consumption for receiving and transmitting is similar to the trend of total energy consumption. The energy consumption of the node in tier 1 is the largest. The total energy consumption and energy consumption for receiving of nodes in tier 1 are greater than twice that of nodes in tier 2. Since nodes in tier 6 are located at the outermost tier of the network, there is no child node and no energy consumption for receiving data, and their total energy consumption and energy consumption of the transmitting are also minimum. The total energy consumption of the first-level nodes is as high as 219 times the energy consumption of nodes in tier 6. We can see the huge difference between them.

Figure 5 shows the lifetime and the residual energy of nodes at each tier after the network died under the CTPR scheme. The black line in the figure shows the lifetime of the node, and the cyan line shows the residual energy after the network died. First, from the black line, it can be seen that the lifetime of the nodes from the first layer to the sixth layer is longer and longer. Since the lifetime of the network is equal to the lifetime of the first dead node, under the CTPR scheme, the lifetime of the network is equal to the lifetime of the node in tier 1. Second, from the cyan line, it can be seen that the residual energy of the node in tier 1 is close to 0, while the residual energy of nodes in other tiers reaches 60% of the total energy.

From the analysis in Figures 4 and 5, we can see that the energy utilization rate of the traditional network using
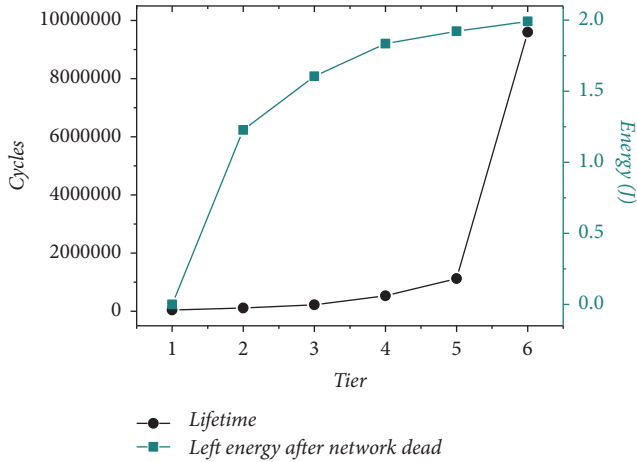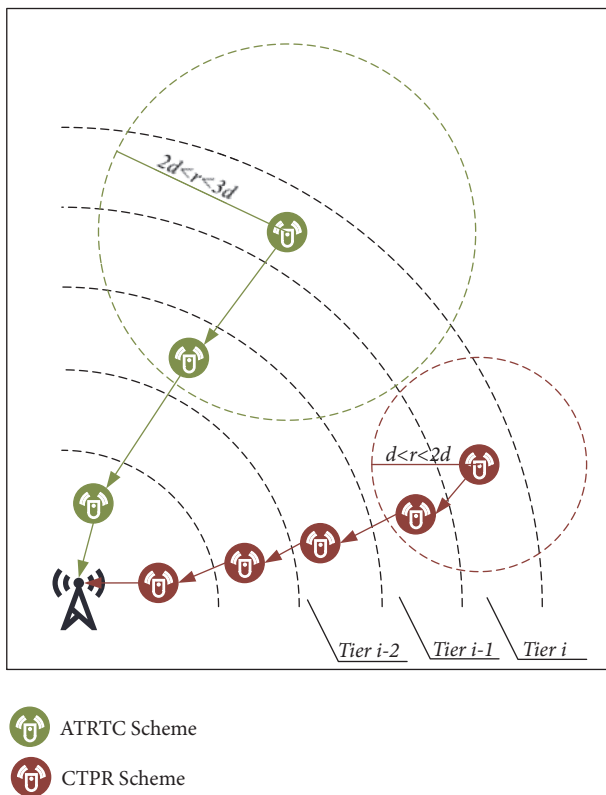
Figure 5: Lifetime and left energy of network.

Figure 6: ATRTC scheme versus CTPR scheme.

can improve the reliability of data collection and reduce the delay of data collection without affecting the network lifetime.

4.2. The ATRTC Scheme. The main idea of the ATRTC scheme is to extend the transmission range of the nodes which have remaining energy by increasing the transmission power of the nodes, so that it can reduce the network collection time and improve the reliability of the network collection. As shown in Figure 6, when the network uses the CTPR scheme, the transmission range r of the node is in the range of $d < r < 2d$. This transmission range can make sure that the node can transmit data to the parent node in the next tier. While the ATRTC scheme increases the transmission distance of the node by increasing the transmission power of the node to a range of $2d < r < 3d$, so that the node in tier i can transmit the data to node both the tier i-1 and tier i-2 nodes. By modifying the transmission protocol, nodes in tier i can transmit data only to the nodes in tier i-2. We call this type of transmission a cross-tier transmission. The ATRTC scheme adaptively extends the transmission range and makes sure PRR reach a level that user required according to the condition of the energy remaining of nodes in different tiers. Through the cross-tier transmission, the number of hops transmitted from the node to the sink node is reduced. In this way, the reliability of the network data collection can be improved and the network data collection delay can be reduced, so that the network performance can be improved.

According to the main idea of the ATRTC scheme, we have designed the ATRTC scheme algorithm (see Algorithm 1).

4.3. The Calculation of Energy Consumption. Calculating energy consumption in the network is very important for the ATRTC scheme. ATRTC scheme needs to adjust the transmission power of node in each tier according to the energy consumption in the network, since the biggest part of the energy consumption of the nodes is the data transmission and the energy consumed by other aspects is negligible. Therefore, the energy consumption of a node is approximately equal to the energy consumption of a node for data transmission. The energy consumption of a node needs to be calculated based on the amount of data transmitted by the node data.

The data transmission of a node includes two parts: receiving data from the child nodes and sending data to the parent nodes. Since broadcast mechanism is adopted in the network, a node can have more than one parent node at the same time. That means the data sent by a node can be received by multiple parent nodes simultaneously. And that also means different parent nodes may have the same child node. It is more complicated than the unicast mechanism where just one node sends data to another node. In the same tier of the network, if the position of nodes is different, the number of their parent nodes is also different. In order to simplify the calculation, we have studied the positions of the most and the least number of parent nodes in a tier and given the calculation methods respectively shown below.

CTPR scheme is very low, and the energy consumption of the network is uneven. After the network died, most of the nodes in the network have a large amount of residual energy. In one word, the CTPR scheme needs to be improved.

Based on the above two points, there is a scheme that can utilize the remaining energy of the far-sink nodes to increase the transmission power of the nodes and improve the network performance. By analyzing the network energy consumption and performance, we propose an Adaptive Transmission Range Based Topology Control (ATRTC) scheme. It

```
1: for i = 1 to tiers:
2:     calculate the parent nodes number of tier i
3:     calculate the child nodes number of tier i
4:     calculate the PRR of tier i
5:     calculate the retransmission time of tier i
6:     calculate the one hop reliability of tier i
7: End for
8: for i = tiers-1 to 1:
9:     calculate the data volume to receive of tier i
10:    calculate the data volume to send of tier i
11:    calculate the energy consumption for receiving data of tier i
12:    calculate the energy consumption for sending data of tier i
13:    calculate the total energy consumption of tier i
14: End for
15: for i = 1 to tiers:
16:    calculate the new transmission power of tier i
17:    calculate the new transmission range of tier i
18: End for
```

ALGORITHM 1: ATRTC Scheme Algorithm.

Theorem 1. *The node v in the outermost of the tier i covers the minimum area of tier where the parent nodes are located. The area is*

$$S_{min} = r_1^2 \cdot (\arccos\angle ACB - \sin\angle ACB \cdot \cos\angle ACB)$$
$$+ r_2^2 \cdot (\arccos\angle CAB - \sin\angle CAB \cdot \cos\angle CAB) \quad (10)$$

where

$$\cos\angle CAB = \frac{l^2 + r_2^2 - r_1^2}{2 \cdot r_1 \cdot r_2},$$

$$\cos\angle ACB = \frac{r_1^2 + l^2 - r_2^2}{2 \cdot r_1 \cdot l} \quad (11)$$

Proof. As shown in Figure 7, the position indicated by point C is the position of the sink node. Point A is the position of the node v which has the least number of parent nodes. All the parent nodes of node v are located in the shadow area. The shadow area can be seen as the area intersected by the circle that center is C and radius is r_1 and the circle that center is A and radius is r_2. Obviously, the longer the distance between two centers, the smaller the area of intersection. And the position of point A leads to the longest distance of two centers, so node v has the least number of parent nodes.

The three sides r_1, r_2, and l of triangle ABC are known, and the cosine value of $\angle CAB$ and $\angle ACB$ can be calculated by cosine theorem. And the area of sector DAB and sector CAB can be calculated by using sector area formula. And the area of isosceles triangle DAB and isosceles triangle CAB can also calculated by using r_1, r_2 and cosine value of $\angle CAB$ and $\angle ACB$. Finally, $S_{\text{shadow area}} = (S_{\text{sector DAB}} - S_{\text{triangle DAB}}) + (S_{\text{sector CAB}} - S_{\text{triangle CAB}})$. □

Theorem 2. *The node v in the innermost of the tier i covers the maximum area of tier where the parent nodes are located. The area is*

$$S_{max} = S_1 - S_2 \quad (12)$$

FIGURE 7: Min area that node covered the parent tier.
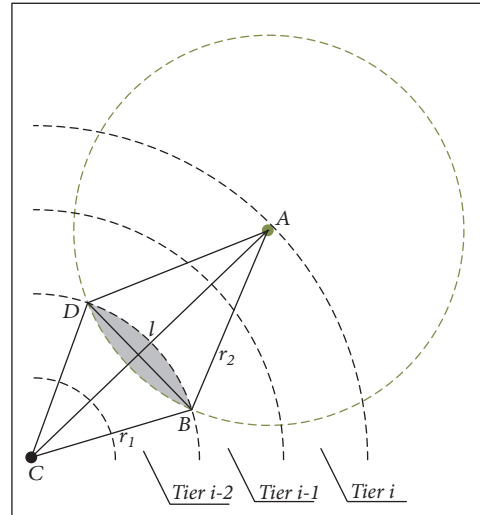
- Node v in outermost position of tier i
- Sink Node

Proof. As shown in Figure 8, S is the shadow area and S_1 is the sum of the areas of S and S_2. Because the shape of S is irregular, it cannot be obtained directly. Fortunately, the area of S_1 can be seen as node v in the outermost of tier i-1 and its parent nodes in tier i-2. Then the area of S_1 can be calculated by Theorem 1. And the area of S_2 can be seen as node v in the outermost of tier i-1 and its parent nodes in tier i-3, so the area of S_2 can also be calculated by Theorem 1. □

Figures 9 and 10 illustrate the minimum area and maximum area of each tier under different transmission range, respectively, in the network (the parameters of network are listed in Table 3). For the same transmission range, the larger the tier number of nodes, the larger the area covered. And it

can be seen that the increase of the coverage area decreases as the tier number increases. For different transmission ranges, the greater the transmission range at the same tier number, the greater the area covered.

$$n = S \cdot \rho = \frac{S \cdot N}{\pi \cdot R^2} \quad \text{where } N \text{ is the total number of nodes in the network, } R \text{ is the radius of the network} \quad (13)$$

Proof. Because N nodes are uniformly deployed in the network, the density of nodes in different area of the network is the same. N and R are given, and the density of node can be calculated: $\rho = N/(\pi \cdot R^2)$. And the area S is given, so the number of nodes in the area is $n = (S \cdot N)/(\pi \cdot R^2)$. $\qquad \square$

The minimum number of parent nodes of nodes in tier i can be calculated by using Theorems 1 and 3, and the maximum number of parent nodes can be calculated by using Theorems 2 and 3.

In order to calculate the receiving data volume of a node, the number of child nodes and the sending data volume of its child nodes need to be calculated at first. The methods to calculate the number of child nodes are given below.

Theorem 4. *Assume that node v in tier i has child nodes in tier j ($i < j$). If the number of parent nodes of a node in tier j (n_j^{parent}) and the total number of nodes in tier j and tier i (n_j and n_i) are known, the average number of child nodes of node v can be calculated by the following equation:*

$$n_i^{child} = \frac{n_j^{parent} \cdot n_j}{n_i} \quad (14)$$

Proof. If a node in tier j has only one parent node in tier i, each node in tier i receives data from average n_j/n_i child nodes in tier j. If broadcasting is used, i.e., a node in tier j has multiple parent nodes in tier i, and each node in tier i receives data from average $n_j^{parent} \cdot n_j/n_i$ child nodes in tier j. $\qquad \square$

The number of child nodes can be obtained by using Theorem 4, and we also need to obtain the sending data volume of child nodes for calculating the receiving data volume of parent node. And the sending data volume of child node needs to be obtained by receiving data volume of child node. The sending and receiving data volume of nodes in the outermost tier is known. So, the receiving and sending data volume can be calculated from the outer tier to the inner tier. Theorem 5 gives the calculation method of the receiving data volume of the node. Theorem 6 gives the calculation method of sending data volume of the node.

Theorem 5. *Assume that node v in tier i has child nodes in tier j ($i < j$). In a data collection cycle, the sending data volume of child nodes in tier j (C_j^{send}) and the number of child nodes of node in tier i (n_i^{child}) are known. If the PRR of data transmission*

from child node to node v is p, and the retransmission time is δ, then the receiving data volume of node v is

$$C_i^{receive} = n_i^{child} \cdot C_j^{send} \cdot \left[1 - (1-p)^{1+\delta}\right] \quad (15)$$

Proof. If the PRR is p and retransmission times is δ, the possibility of a successful data transmission is $1 - (1-p)^{1+\delta}$. If the wireless channel is lossless, the receiving data volume of node v is $n_i^{child} \cdot C_j^{send}$. So, under the lossy wireless channel, the receiving data volume of node v can be calculated; that is, $n_i^{child} \cdot C_j^{send} \cdot [1 - (1-p)^{1+\delta}]$. $\qquad \square$

Theorem 6. *In a data collection cycle, if the retransmission times δ, the receiving data volume $C_i^{receive}$, and the aggregate function $f(data)$ are known, the sending data volume of node v in tier i is*

$$C_i^{send} = (\delta + 1) \cdot f\left(C_i^{receive} + 1\right) \quad (16)$$

Proof. The sending data of node v are generated by aggregate function. And the input data of aggregate function are comprised of the data generated by self (size of 1 data packet) and the data received from its child nodes. The data volume of input data is $C_i^{receive} + 1$. If the aggregate function $f(data)$ is known, the output data volume is $f(C_i^{receive} + 1)$. If node v retransmits data δ times, the sending data volume is $(\delta + 1) \cdot f(C_i^{receive} + 1)$ $\qquad \square$

The sending and receiving data volume of nodes in each tier can be calculated by using Theorems 5 and 6. They are the essential values to calculate the energy consumption of nodes in each tier. And the calculation methods of energy consumption according to data volume are shown in theorem below.

Theorem 7. *Assume the following values are known: sending data volume C_i^{send} and receiving data volume $C_i^{receive}$ of a node in tier i, the power for sending data P_i^{send} and receiving data $P_i^{receive}$, the sending rate V_{send}, and receiving rate $V_i^{receive}$. The energy consumption of a node for sending data in a data collection cycle is*

$$E_{send} = \frac{C_i^{send} \cdot P_i^{send}}{V_{send}} \quad (17)$$
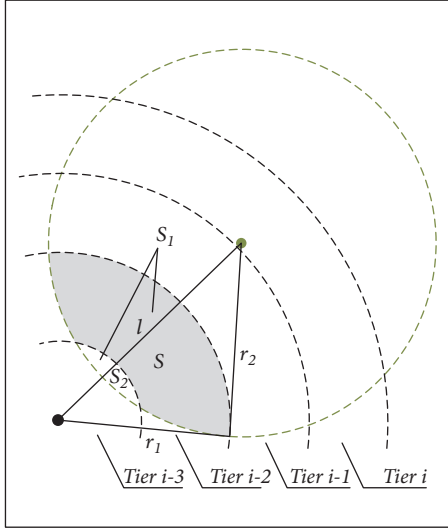
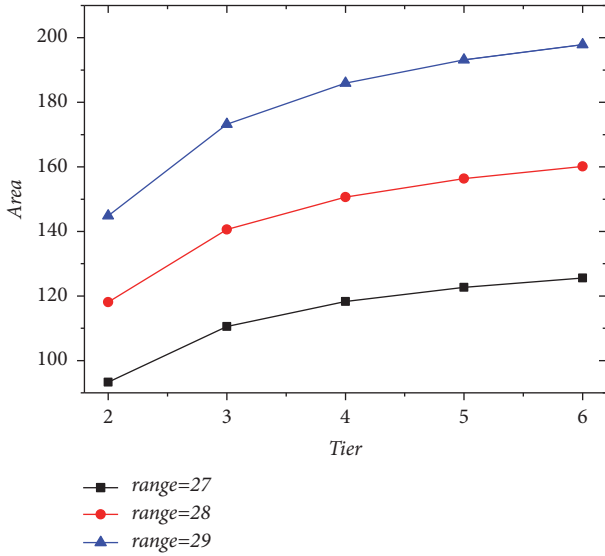FIGURE 8: Max area that node covered the parent tier.

FIGURE 9: Min covered area with ranges.

The energy consumption of a node for receiving data in a data collection cycle is

$$E_{receive} = \frac{C_i^{receive} \cdot P_i^{receive}}{V_{receive}} \qquad (18)$$

The total energy consumption of a node in a data collection cycle is

$$E_{total} = E_{send} + E_{receive} \qquad (19)$$

Proof. If a power P and time t are known the energy consumption E can be calculated as $E = P \cdot t$. If the data volume
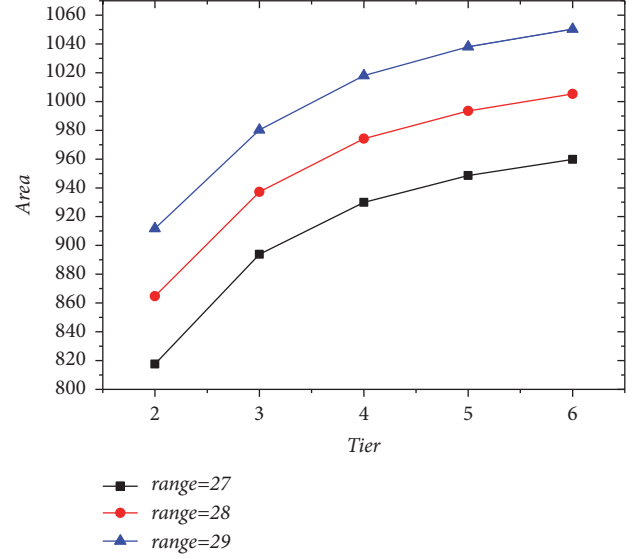

FIGURE 10: Max covered area with ranges.

C and transmission rate V are known, the time t can be calculated as $t = C/V$. So, $E = C \cdot P/V$. To calculate the energy consumption for sending data, just substitute C_i^{send}, P_i^{send}, and the V_{send} into the equation $E = C \cdot P/V$, that is, $E_{send} = C_i^{send} \cdot P_i^{send}/V_{send}$ And the energy consumption for receiving data can be calculated by using the method; that is, $E_{receive} = C_i^{receive} \cdot P_i^{receive}/V_{receive}$. Since the energy of a node is mostly used by sending data and receiving data. Therefore, the total energy consumption of the node is $E_{total} = E_{send} + E_{receive}$. □

4.4. The Calculation of Transmission Range. The key point of the ATRTC scheme is to extend the transmission range properly, and the PRR should not be decreased at same time. To achieve this goal, the transmission power needs to be increased.

Theorem 8. *Assume the following values are known: the energy consumption of node v in tier i for receiving data $E_i^{receive}$, the sending data volume of node v C_i^{send}, the sending rate V_{send}, the maximum energy consumption in the network (usually the node in tier 1) E_{max}, and the max transmission power that the node can adjust determined by hardware P_{limit}. Under the ATRTC scheme, the transmission power of a node can adjust is*

$$P_i^{send\prime} = \min\left\{ P_{limit}, \frac{\left(E_{max} - E_i^{receive}\right) \cdot V_{send}}{C_i^{send}} \right\} \qquad (20)$$

Proof. Under the ATRTC scheme, some nodes can increase the transmission power by using the residual energy. The energy which can be used for sending data is $E_{max} - E_i^{receive}$. According to the equation $E = C \cdot P/V$ in proof of Theorem 7, the power can be calculated as $P = E \cdot V/C$. Since the energy can use $E_{max} - E_i^{receive}$, and the sending rate V_{send} and the sending data volume are known, the new transmission power of node under ATRTC scheme can be

calculated: $(E_{\max} - E_i^{receive}) \cdot V_{send}/C_i^{send}$. And the hardware limit of the transmission power needs to be considered; if the results calculated by the equation are beyond the E_{limit}, the transmission power should be E_{limit}. So, the new transmission power under ATRTC scheme can be expressed as $\min\{E_{limit}, P_i^{send} \cdot (E_{\max} - E_i^{receive})/E_i^{send}\}$. □

Theorem 9. *If the transmission power P_s and the PRR standard p are known, the max transmission range can be calculated by the following equation:*

$$r_{max} = \frac{\left(10^{[PL(r)-PL(r_0)]/10n}\right)}{r_0} \tag{21}$$

where

$$PL(r) = P_s + 1.28 \ln\left[2\left(1 - p^{1/8f}\right)\right] - P_n \tag{22}$$

Proof. According to the reliability model in Section 3.2, the relationship of PRR, transmission power, and transmission range can be known by (2), (3), and (4). And according to (2), the relationship between signal noise ratio γ and PRR p can be known: $\gamma = -1.28\ln[2(1 - p^{1/8f})](2)'$. According to (3), the relationship of $PL(r)$ and γ can be known: $PL(r) = P_t - \gamma - P_n(3)'$. And according to (4), the relationship between r and $PL(r)$ can be known: $r = (10^{[PL(r)-PL(r_0)]/10n})/r_0(4)'$. Substituting (2)' to (3)', and then substituting (3)' to (4)', we can get the relationship between the transmission range r, the transmission power P_s, and the PRR p. □

The max transmission range of node under a certain transmission power and a certain PRR can be calculated by using Theorem 9. But this is not a proper transmission range, because the excessive transmission range will lead to the data redundancy and energy waste. And the data collection time will be increased accordingly. Therefore, after calculating the max transmission range, the new transmission range needs to be adjusted based on the max transmission range. In ATRTC scheme, we suggest that the proper transmission range should satisfy the condition of not more than $3d$ in length and the minimum number of parent nodes is 2.

Under the ATRTC scheme, a node has residual energy that can achieve cross-tier transmission. In order to figure out which tier the node can reach, we show the method in Theorem 10.

Theorem 10. *If the transmission range r of node v in tier i and the width of tier d are known, the tier number of the tier that node v can reach is*

$$tier = i - \left\lfloor \frac{r}{d} \right\rfloor \tag{23}$$

Proof. If the transmission range r and the width of tier d are known, the number of tiers that node v can completely span can be calculated, that is, $\lfloor r/d \rfloor$. Because the order of tier number from the outermost tier to the sink is descending, and the tier number of node v is also known, the tier number of node v can be $i - \lfloor r/d \rfloor$ □
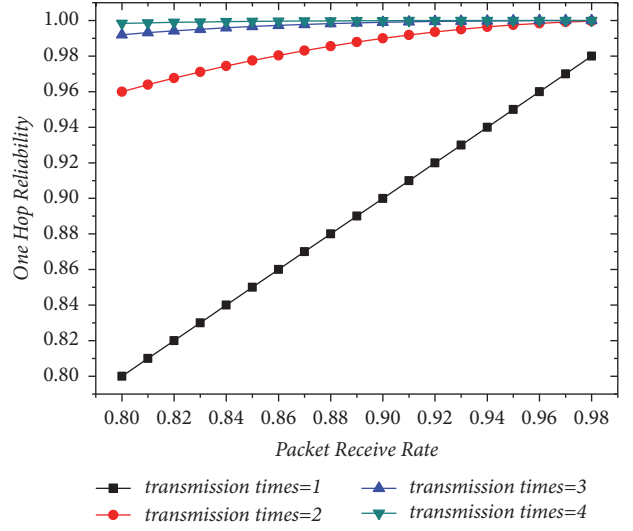

FIGURE 11: One-hop reliability with transmission times.

4.5. The Calculation of Data Collection Reliability. Due to the data aggregation technology used in the network, the amount of information in a single packet increases, the final result generated after data collection is more vulnerable and affected by packet loss. Therefore, improving the reliability of data collection helps to ensure the accuracy of data collection. Theorem 11 gives the calculation method of the reliability of the node to transmit data to the parent node. Theorem 12 gives the calculation method for the reliability of the data sent by node reach to the sink node successfully. And Theorem 13 gives the calculation method of network data collection reliability.

Theorem 11. *The one-hop reliability of node in tier i is defined as the possibility of data sent by node in tier i reach to the parent nodes successfully by using broadcast and retransmission mechanism. If the minimum numbers of parent nodes x, the PRR p, and the retransmission time δ are known, the one-hop reliability of node in tier i is*

$$\varphi_i = 1 - (1 - p)^{x \cdot (1+\delta)} \tag{24}$$

Proof. If the PRR of the data transmission is p, the possibility of packet loss is $(1 - p)$. If the node transmits the same data packet m times, the possibility of packet loss is $(1 - p)^m$. The δ times retransmission can be seen as $\delta+1$ times data transmission, the possibility of packet loss is $(1 - p)^{1+\delta}$. And transmitting data to x parent nodes simultaneously can be seen as x times data transmission, the possibility of packet loss is $(1 - p)^x$. If node transmits data packet to parent nodes by using broadcast and retransmission mechanism, the possibility of packet loss is $(1 - p)^{x \cdot (1+\delta)}$. Therefore, the possibility of successful data transmission, i.e., the one-hop reliability, is $1 - (1 - p)^{x \cdot (1+\delta)}$. □

According to the proof of Theorem 11, the data transmission by broadcast retransmission mechanism can be seen as multiple transmissions of the same data packet. Figure 11

shows the relationship between one-hop reliability and PRR under different transmission times. As shown in Figure 11, first, the greater the one-hop reliability the greater the PRR under the same transmission times. Second, the more transmission times, the greater one-hop reliability under the same PRR. When the transmission times=1, the one-hop reliability of node is equal to the PRR. When transmission time =2, the one-hop reliability of node is greatly improved compared to that when transmission time =1. Even when the PRR=0.8, the one-hop reliability can reach 0.96 under transmission time =2. It can be seen that when the PRR is relatively low, using broadcast and retransmission mechanism can significantly improve the one-hop reliability.

Theorem 12. *Assume that a data packet from a node reaches sink node successfully needs multiple hop transmission. If the one-hop reliability of each hop is $\varphi_1, \varphi_2, \cdots, \varphi_m$, the reliability of a data transmission from the node in tier i to the sink node is*

$$\Phi_i = \prod_{k=1}^{m} \varphi_k \qquad (25)$$

Proof. The data sent by the node in tier i needs to forward by the parent node closer to the sink node. The reliability of successful transmission to the sink node is based on the reliability of each one-hop reliability of node to forward the data. Therefore, the reliability of data transmission from the node in tier i to the sink node can be expressed as $\Phi_i = \prod_{k=1}^{m} \varphi_k$. $\qquad \square$

Theorem 13. *If the reliabilities of data transmission from node in each tire to the sink node $\Phi_1, \Phi_2, \cdots, \Phi_m$ are known, the reliability of network data collection Q is*

$$Q = \min \{\Phi_1, \Phi_2, \cdots, \Phi_m\} \qquad (26)$$

Proof. The distances between the nodes in difference positions and the sinks are different, the reliability of the transmission to the sink is also different. The minimum reliability is used to measure the reliability of network data collection. It can make sure the reliability of the transmission of data sent from any node reach to the sink node is not less than the minimum reliability Q. $\qquad \square$

4.6. The Calculation of Data Collection Time. The data collection time in this paper is the time length from the beginning to the finish of data collection in the network. According to the data collection model in Section 3.3, the parent nodes need to wait until the child nodes finish the data transmission. Under the ATRTC scheme, the transmission ranges of nodes in each tier are different. The node that has residual energy can achieve cross-tier transmission. For example, the nodes in tier 8 can transmit data to the nodes in tier 6 rather than tier 7. And the nodes in tier 7 can transmit data to the nodes in tier 5. So that the nodes in tier 7 need not to wait the nodes in tier 8. The nodes in tier 7 and tier 8 can transmit data to their parent nodes at the same time as long as their parent nodes are not in the same tier. So, under the ATRTC scheme, there are two paths that can transmit data to the sink node

if some conditions are satisfied. The sink node usually has a stable power supply, and the data transmission function is more powerful than the sensor node. It is assumed that the sink node can receive data transmitted from nodes of two tiers at the same time.

Theorem 14. *Assume the node v in tier i has child nodes in tier j. If the number of child nodes of node v n_i^{child}, the average sending data volume of each child node C_j^{send} and the sending rate $V_{receive}$ of nodes are known, and the time of data transmission from child nodes to node v is*

$$t = \frac{n_i^{child} \cdot C_j^{send}}{V_{receive}} \qquad (27)$$

Proof. Assume the sensor node usually has multiple data channels, and nodes in the same tier can transmit data simultaneously if their parent nodes are not the same. The parent nodes can receive data from only one child node at one time. If a parent node has multiple child nodes, the child nodes need to transmit data in order. The sending rate and receiving rate are the same value in this paper. If the wireless channel is lossless, the transmission time from nodes in tier j to their parent nodes in tier i can be calculated by the receiving data volume of parent nodes. But, if the wireless channel is lossy, using the total sending data volume of child nods is more accurate. Therefore, the time of data transmission from child nodes to node v can be expressed as $n_i^{child} \cdot C_j^{send}/V_{receive}$. $\qquad \square$

Theorem 15. *Assume that there are two paths for the data collection, and each path is comprised of nodes in m different tiers. The two paths can transmit data simultaneously. The data transmission times on the first path are $t_{11}, t_{12}, \cdots, t_{1m}$. The data transmission times on the second path are $t_{21}, t_{22}, \cdots, t_{2m}$. The data collection time of the network can be calculated by the following equation:*

$$\mathcal{T} = \max \left\{ \sum_{i=1}^{m} t_{1i}, \sum_{i=1}^{m} t_{1i} \right\} \qquad (28)$$

Proof. Since the parent node needs to wait for all child nodes to complete the transmission to start data transmission, the transmission time on each path is $\sum_{i=1}^{m} t_i$. For the two paths that can be transmitted simultaneously, the data collection time is determined by the transmission time of the last completed path. Therefore, the data collection time of the network can be expressed as $\max\{\sum_{i=1}^{m} t_{1i}, \sum_{i=1}^{m} t_{1i}\}$. $\qquad \square$

The data collection time of the network under ATRTC scheme can be calculated by using Theorems 14 and 15. For that under CTPR scheme, every transmission time from child tier to parent tier need to be calculated by using Theorem 14 and the data collection time of the network under CTPR scheme can be calculated by accumulating them.

4.7. The Calculation of Network Lifetime

Theorem 16. *If the initial energy of node E_{init} and the total energy consumption E_c in a data collection cycle are known, the lifetime of a node is*

$$\ell = \left\lfloor \frac{E_{init}}{E_c} \right\rfloor \tag{29}$$

Proof. The lifetime of a node is defined as the number of data collection cycle from the full state to the empty state of the battery of a node. Therefore, the number of data collection cycle can be easily calculated by the known initial energy of node E_{init} and energy consumption E_c in a data collection cycle: $\lfloor E_{init}/E_c \rfloor$. □

Theorem 17. *Assume the network has m tiers, and the lifetime of nodes in each tier is $\ell_1, \ell_2, \cdots, \ell_m$. The lifetime of the network can be expressed as follows:*

$$\mathcal{L} = \min\{\ell_1, \ell_2, \cdots, \ell_m\} \tag{30}$$

Proof. According to the definition of the network lifetime in Section 3.4, the lifetime of the network is determined by the lifetime of the node first died. That means the lifetime of the network is the minimum lifetime of nodes in all tiers. Therefore, the lifetime of network can be expressed as $\min\{\ell_1, \ell_2, \cdots, \ell_m\}$. □

In a data collection cycle, the data volume received and sent by nodes in different tiers is different. The data transmission time of nodes in different tiers is also different. Therefore, the lifetime expressed by using the data collection cycle is more convenient than using the transmission time. The lifetime of the network can be easily calculated by using Theorems 16 and 17.

5. Analysis and Comparison

In this section, the wireless sensor network under Adaptive Transmission Range Based Topology Control (ATRTC) scheme will be compared with the wireless sensor network under Constant Transmission Power and Range (CTPR) scheme. The network structure, the data collection reliability of the network, the data collection time, and so on will be compared and analyzed in the following contents.

5.1. Network Structure. Under the ATRTC scheme, the transmission power and the transmission range of nodes in each tier are different compared with these under CTPR scheme.

The transmission power of nodes under ATRTC scheme and under CTPR scheme is compared in Figure 12. As can be seen from Figure 12, the transmission power of nodes in most tiers under ATRTC scheme is increased compared with that under CTPR scheme. The transmission power of node in all tiers is the same value under CTPR scheme. And the more the residual energy of the nodes, the higher the transmit power under the ATRTC scheme.

Under the ATRTC scheme, increasing the transmission power of a node can extend the transmission range and
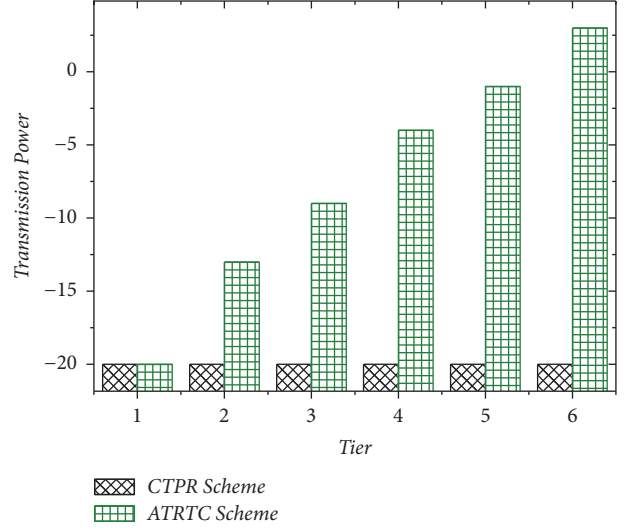

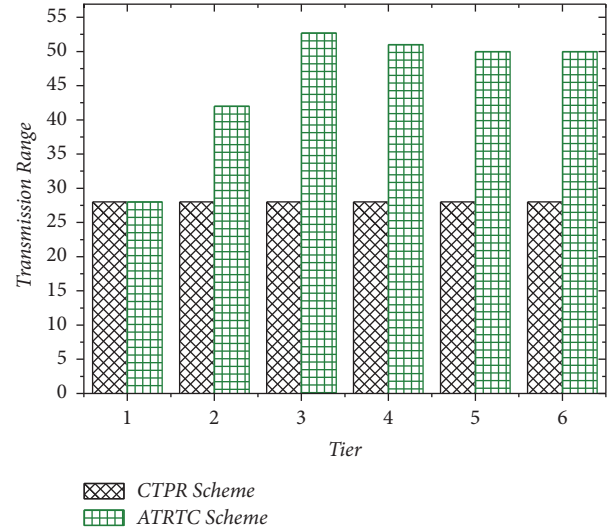Figure 12: Comparison of transmission power.

Figure 13: Comparison of transmission range.

improve the transmission reliability of the node. The transmission range of the nodes under CTPR Scheme and ATRTC Scheme is compared in Figure 13. It can be seen from Figure 13 that the transmission range of nodes in all tiers is the same under the CTPR scheme. The transmission range of node is all increased in tier 2-6 under ATRTC scheme. The transmission range of node in tier 2 is 42m; it can just reach the sink node. The transmission range of node from tier 3 to tier 6 is slightly reduced, because the total area of tier is increased from tier 3 to tier 6.

Figure 14 shows which tier the child node can reach. The number on the vertical axis represents the serial number of tier, and 0 represents the serial number of sink node tier. Under the CTPR scheme, the serial number of the parent tier is always one less than the serial number of the child tier. The transmission range under CTPR scheme is all the same. But under the ATRTC scheme, most serial number of parent tier
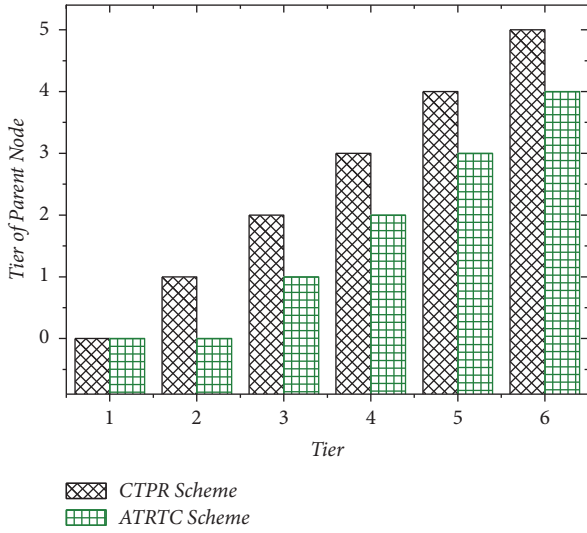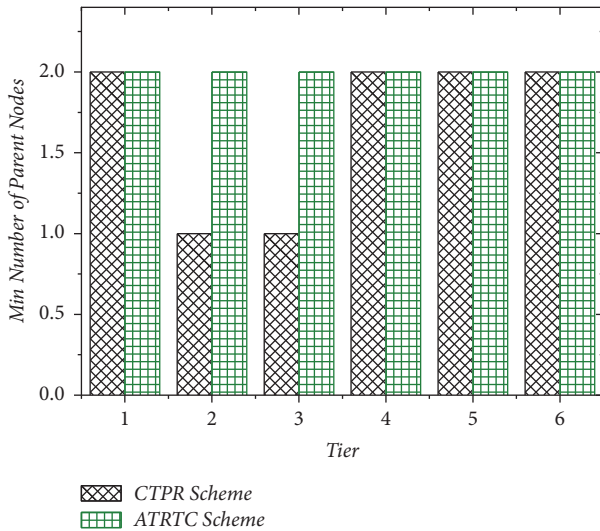
FIGURE 14: Comparison of tier of parent node.

FIGURE 15: Comparison of min number of parent nodes.

is two less than the serial number of the child tier, and nodes in tier 1 and tier 2 have the same serial number of parent tier. Nodes in tier 2 have residual energy and the transmission range of them is extended.

Figure 15 shows the minimum number of parent nodes in each tier under ATRTC scheme and CTPR scheme. Under CTPR scheme, the minimum number of in tier 2 and tier 3 is 1. If a child node only has only one parent node, broadcast mechanism will not improve the reliability of data transmission from a child node to parent node. Therefore, nodes in tier 2 and tier 3 transmit data using retransmission mechanism to improve the reliability under the CTPR scheme. And under ATRTC scheme, the minimum number of parent nodes in each tier is always 2 due to the increase of transmission range.

Figure 16 shows the path of data collection under ATRTC scheme and CTPR scheme. Under CTPR scheme, a path of

data is collected from outer to inner tier by tier. And under the ATRTC scheme, there are two paths of data collection. Nodes in tier 6 transmit data to parent nodes in tier 4 and nodes in tier 5 transmit data to parent nodes in tier 3. Nodes in tier 6 and tier 5 can transmit data simultaneously. As can be seen in the Figure 16, nodes in tier 6 can transmit data to the sink node just through 3 hops under ATRTC and 5 hops under CTPR scheme.

The average number of parent and average number of children under the CTPR Scheme and ATRTC Scheme are compared in Figures 17 and 18, respectively. It can be seen from Figure 17 that the average number of parent nodes of nodes in tier 4-6 under the ATRTC scheme is increased compared to the CTPR scheme due to the increase of the transmission range. As can be seen from Figure 18, under the ATRTC scheme, the average number of child node in tier 1-4 is more than that under the CTPR scheme. The total number of nodes in a tier is increased with the serial number of tier increase according to Figure 19. And the most serial number of child tier under ATRTC scheme is more than that under CTPR scheme with the same parent tier serial number.

5.2. Data Collection Reliability. In Figure 20, the PRR of node in every tier is 0.868 due to the same transmission range and power under CTPR scheme. And under ATRTC scheme, there is a significantly improvement of the PRR in tiers 4, 5, and 6 compared with CTPR scheme. Because the nodes in tiers 4, 5, and 6 have lots of residual energy, they can use a high transmission power under ATRTC. This led to a high PRR. The PRR of nodes in tier 2 and 3 under ATRTC scheme is roughly equal to that under CTPR scheme. The residual energy of nodes in tiers 2 and 3 is not as much as that in tiers 4, 5, and 6. The residual energy is used to extend the transmission range at first. The one-hop reliability of nodes under ATRTC scheme and CTPR scheme is compared in Figure 21. Under the ATRTC scheme, the one-hop reliability of nodes in tiers 4, 5, and 6 has a significant improvement compared with CTPR scheme. And the one-hop reliability of nodes in tiers 1, 2, and 3 under ATRTC scheme is roughly equal to the CTPR scheme. But the one-hop reliability of nodes can reach 0.98 at least both under ATRTC and CTPR scheme.

The reliability of data reaching sink node (RDS) from nodes in each tier under ATRTC scheme and CTPR scheme is compared in Figure 22. Under the CTPR scheme, the RRS is decreased with the increase of tier serial number. The reliability of tier 6 is the lowest, that is, 0.9002. This is because every hop of data transmission has a possibility of data loss, and the further away from the sink node, the more hops required. Under the ATRTC scheme, the RDS from nodes in all tiers except tier 1 is higher than that under CTPR scheme. Transmitting data to the sink node under ATRTC scheme needs less hops than CTPR scheme. And the one-hop reliability of nodes in most tiers under ATRTC scheme is higher than CTPR scheme. The RDS from node in tier 5 is the lowest under ATRTC scheme, and it still reaches 0.9659. According to Theorem 13, the reliability of data collection under ATRTC scheme is 0.9659 and under CTPR scheme is 0.9002. The ATRTC scheme increased by 7.3% over the CTPR scheme on the reliability of data collection.
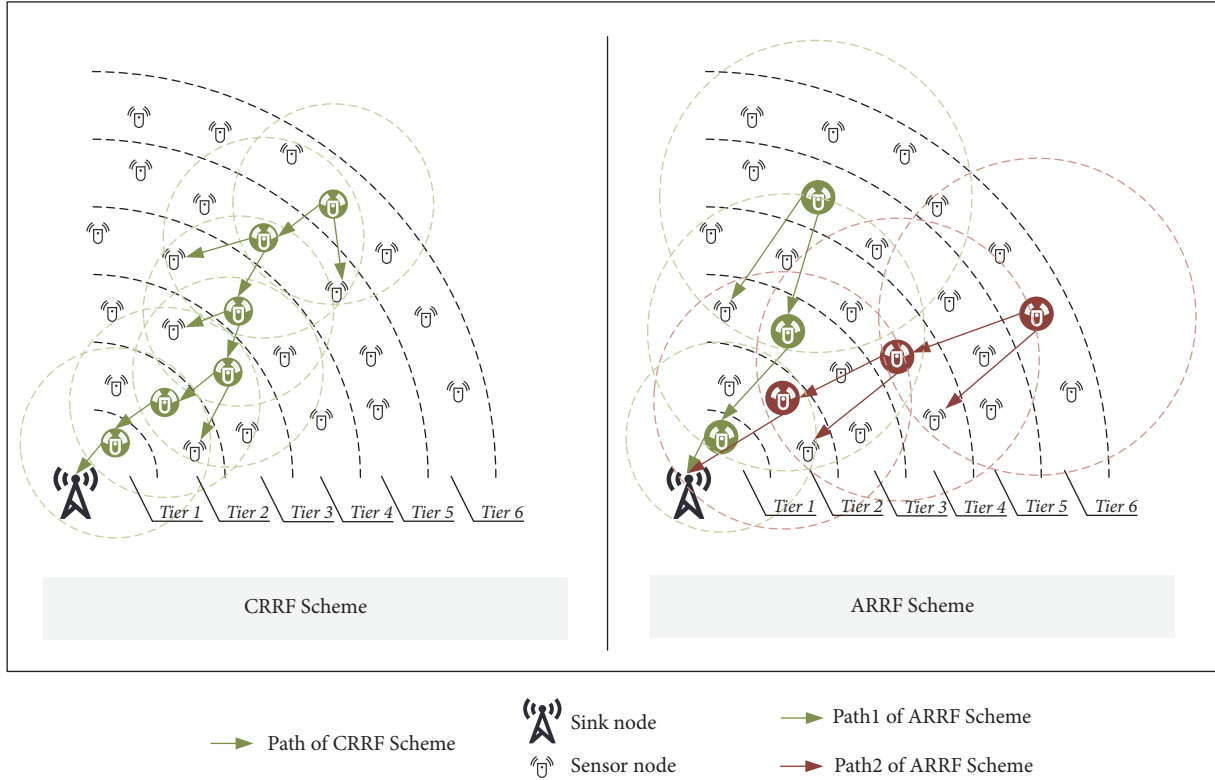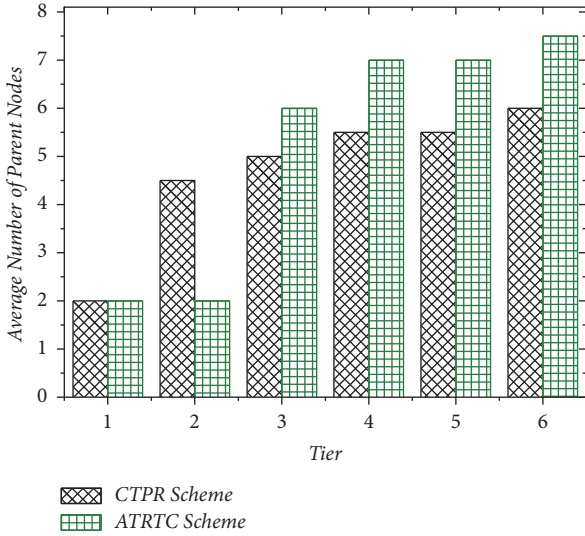
FIGURE 16: Comparison of data collection path.

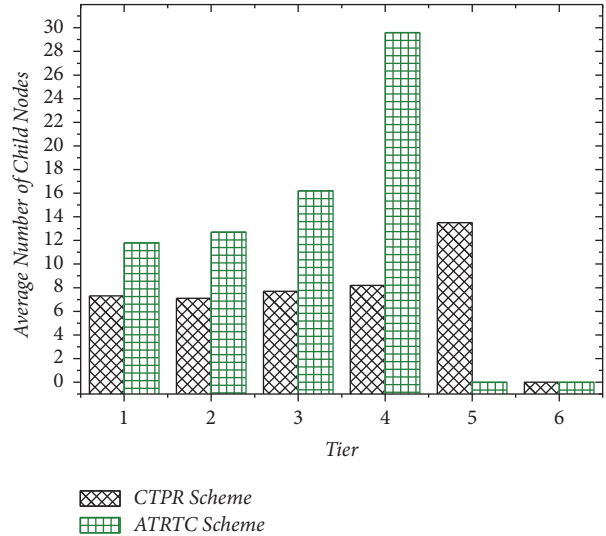FIGURE 17: Comparison of average number of parent nodes.

FIGURE 18: Comparison of average number of child nodes.

5.3. Data Collection Time. The sending data volume and receiving data volume of nodes in each tier under ATRTC scheme and CTPR scheme are shown in Figures 23 and 24, respectively. In Figure 23, the sending data volume of nodes increases as the serial number of tiers decreases. The parent nodes in the inner tier need to forward the data of the child nodes in outer tier and the number of parent nodes is less than the number of child nodes. But under ATRTC scheme the sending data volume is less than CTPR scheme.

As can be seen from Figure 24, under the ATRTC scheme, the receiving data volume of nodes in most tiers is less than the CTPR scheme. But the receiving data volume of nodes in tier 2 is more than the CTPR scheme. And the number of child nodes under ATRTC scheme is more than CTPR scheme and the sending data volume of child nodes is roughly the same according to Figures 18 and 23.

Tables 4 and 5 show the paths of one-hop and the time spent on them under CTPR scheme and ATRTC scheme. As
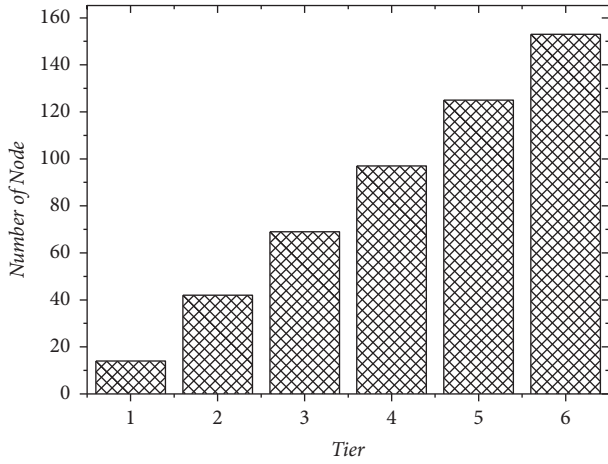
FIGURE 19: Number of node in each tier.

FIGURE 20: Comparison of packet receive rate.

FIGURE 21: Comparison of one-hop reliability.

FIGURE 22: Comparison of reliability of data transmit to sink.

FIGURE 23: Comparison of data volume of node to send.

TABLE 4: Data collection time in CTPR scheme.

Serial number	Path of one hop	Time(s)
1	6->5	0.152083
2	5->4	0.329560
3	4->3	0.695604
4	3->2	1.226709
5	2->1	3.766472
6	1->0	12.112338
Data collection time		18.282766

can be seen from Figures 4 and 5, the closer to the sink node, the more time which is spent on the path of one-hop. The closer to the sink node, the more time which is spent on the one-hop path. Under the same serial number, the time spent on the path of one-hop under ATRTC scheme is mostly more

FIGURE 24: Comparison of data volume of node to receive.

TABLE 5: Data collection time in ATRTC scheme.

Serial number	Path of one hop	Time(s)
1	6->4	0.245833
2	5->3	0.264583
3	4->2	1.532247
4	3->1	2.966007
5	2->0	12.278143
6	1->0	7.459946
Data collection time		14.056223

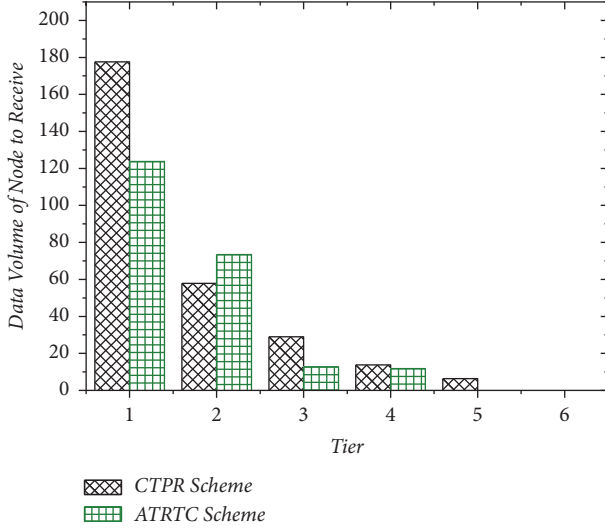than CTPR scheme. For example, when serial number is 1, under CTPR scheme the time spent on one-hop is 0.152083s and 0.245833s under ATRTC scheme. But under ATRTC scheme, data from the nodes in tier 6 can directly reach the nodes in tier 4 in 0.245833s and 0.481643s under CTPR scheme. This is because of the cross-tier transmission and two paths of data collection under ATRTC scheme. Therefore, the data collection time of the network under the ATRTC scheme was reduced by 23% compared to the CTPR scheme.

5.4. Energy Consumption and Network Lifetime. The energy consumption of nodes in each tier and network lifetime under CTPR scheme and ATRTC scheme is compared in this section. The energy consumption for sending data, the energy consumption for receiving data, and the total energy consumption of nodes in each tier in a data collection cycle are shown in Figures 25, 26, and 27, respectively. As shown in Figure 25, the energy consumption of nodes in tier 1 under ATRTC scheme is lower than CTPR scheme, and the energy consumption of nodes in tiers 2-6 under ATRTC scheme is higher than CTPR scheme. Because the nodes in tiers 2-6 have residual energy, the transmission power of them is increased under ATRTC scheme. The trend of Figure 26 is similar to Figure 24, because the receiving power is not changed under ATRTC scheme. As shown in Figure 27, the closer to the sink node, the greater the energy consumption
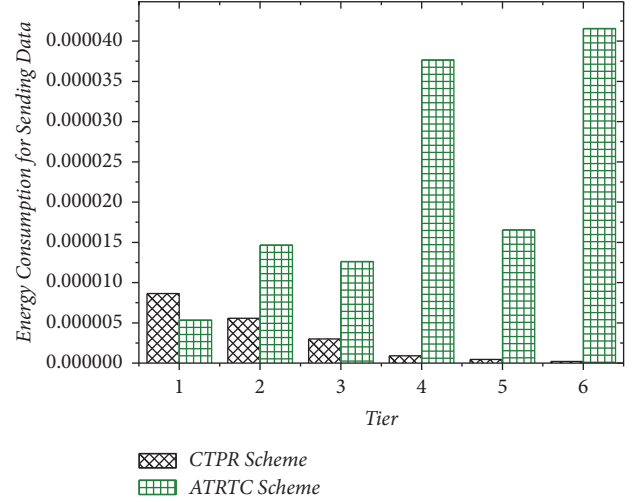
FIGURE 25: Comparison of energy consumption for sending data.
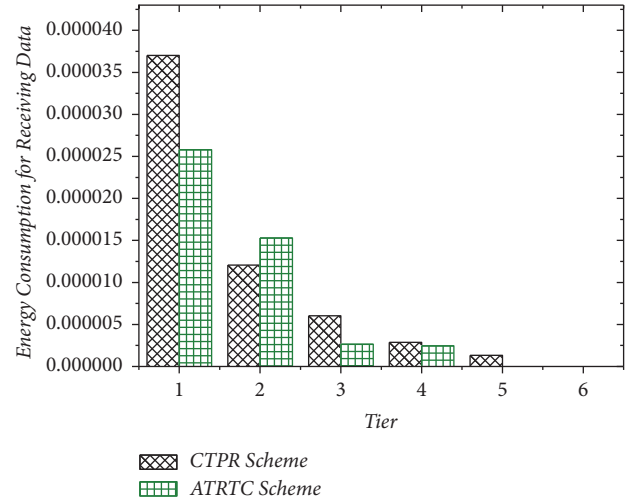
FIGURE 26: Comparison of energy consumption for receiving data.

TABLE 6: Lifetime under ATRTC scheme and CTPR scheme.

Tier	Lifetime under CTPR Scheme	Lifetime under ATRTC Scheme
1	43799	64274
2	113417	66790
3	221450	131059
4	531251	49859
5	1120706	120856
6	9600000	48113
Network Lifetime	43799	48113

of nodes under the CTPR scheme. The maximum and minimum ratio of energy consumption is 219.2 under CTPR scheme and 2.7 under ATRTC scheme. And the maximum energy consumption under ATRTC scheme is 9% less than CTPR scheme.
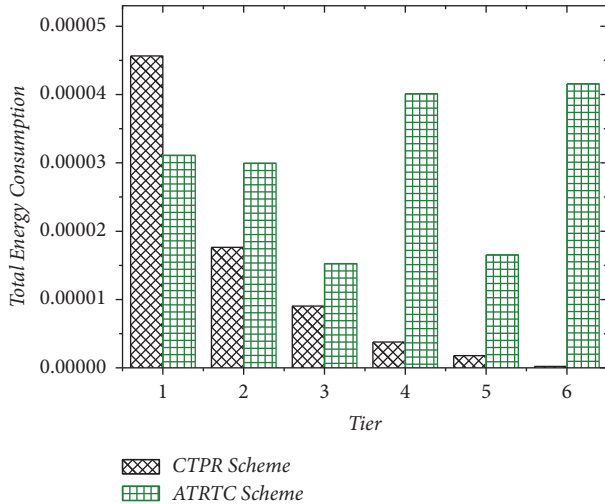
Figure 27: Comparison of total energy consumption.

Table 6 shows the lifetime of nodes in each tier under ATRTC scheme and CTPR scheme. According to the definition of the network lifetime, it is the minimum lifetime of node. The network lifetime under CTPR scheme is 43799 cycles and 48113 cycles under ATRTC scheme. The network lifetime under ATRTC scheme is about 10% higher than CTPR scheme.

6. Conclusion

In the wireless sensor network, the resources of a node are quite limited, and the wireless channel is lossy. Therefore, how to achieve high stability, low power consumption, and low latency network performance is very challenging under limited energy conditions.

This paper focuses on optimizing the lifetime, data collection time, and data collection reliability of periodic data collection wireless sensor networks by controlling the transmission range of the nodes. After careful analysis, we found that the traditional Constant Transmission Power and Range (CTPR) scheme has a shortage that energy consumption near-sink node is much higher than far-sink node. Therefore, the Adaptive Transmission Range Based Topology Control (ATRTC) scheme is proposed in this paper. If a node has residual energy, the transmission power will be increased, the transmission range will be extended, and the transmission quality will be improved under ATRTC scheme. Increasing the transmission range can reduce the number of hops that data reaches the sink node and improve the reliability of data collection. And the increase of paths of data collection leads to reducing of data collection time. According to the analysis in this paper, compared with the CTPR scheme, the ATRTC scheme reduces the maximum energy consumption by 9%, increases the network lifetime by 10%, increases the data collection reliability by 7.3%, and reduces the network data collection time by 23%.

The ATRTC scheme also has some limitations. If the energy consumption difference between the nodes is not very large, the nodes do not have enough residual energy to

substantially extend the transmission range to achieve cross-tier transmission. In this case, the ATRTC scheme will not have a significant performance increase. How to improve the performance in that case can be solved in future work. In addition, ATRTC scheme improves the transmission range of sensor nodes. And larger transmission range will be interfered by more nodes. How to reduce the interference between nodes is also a problem that needs to be solved in the future.

Data Availability

This paper does not include data which can be shared with the public. The data used to support the findings of this study have been described in this paper clearly.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] A. Liu, M. Huang, M. Zhao, and T. Wang, "A Smart High-Speed Backbone Path Construction Approach for Energy and Delay Optimization in WSNs," *IEEE Access*, vol. 6, pp. 13836–13854, 2018.

[2] M. Huang, A. Liu, M. Zhao, and T. Wang, "Multi working sets alternate covering scheme for continuous partial coverage in WSNs," *Peer-to-Peer Networking and Applications*, 2018.

[3] P. L. Nguyen, Y. Ji, Z. Liu, H. Vu, and K.-V. Nguyen, "Distributed hole-bypassing protocol in WSNs with constant stretch and load balancing," *Computer Networks*, vol. 129, pp. 232–250, 2017.

[4] Z. Liu, T. Tsuda, H. Watanabe, S. Ryuo, and N. Iwasawa, "Data Driven Cyber-Physical System for Landslide Detection," *Mobile Networks and Applications*, 2018.

[5] K. Ota, M. S. Dao, V. Mezaris, and F. G. Natale, "Deep Learning for Mobile Multimedia," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 13, no. 3s, pp. 1–22, 2017.

[6] Y. Zhang, X. Chen, J. Li, D. S. Wong, H. Li, and I. You, "Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing," *Information Sciences*, vol. 379, pp. 42–61, 2017.

[7] Q. Xu, Z. Su, and K. Yang, "Optimal control theory-based epidemic information spreading scheme for mobile social users with energy constraint," *IEEE Access*, vol. 5, pp. 14107–14117, 2017.

[8] Q. Xu, Z. Su, Q. Zheng, M. Luo, and B. Dong, "Secure Content Delivery with Edge Nodes to Save Caching Resources for Mobile Users in Green Cities," *IEEE Transactions on Industrial Informatics*, 2017.

[9] M. Li, Z. Li, and A. V. Vasilakos, "A survey on topology control in wireless sensor networks: taxonomy, comparative study, and open issues," *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2538–2557, 2013.

[10] X. Liu, "A novel transmission range adjustment strategy for energy hole avoiding in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 67, pp. 43–52, 2016.

[11] Y. Liu, A. Liu, S. Guo, Z. Li, Y.-J. Choi, and H. Sekiya, "Context-aware collect data with energy efficient in Cyber-physical cloud systems," *Future Generation Computer Systems*, 2017.

[12] K. Xie, J. Cao, X. Wang, and J. Wen, "Optimal resource allocation for reliable and energy efficient cooperative communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4994–5007, 2013.

[13] Y. Liu, K. Ota, K. Zhang et al., "QTSAC: An Energy-Efficient MAC Protocol for Delay Minimization in Wireless Sensor Networks," *IEEE Access*, vol. 6, pp. 8273–8291, 2018.

[14] H. Dai, G. Chen, C. Wang, S. Wang, X. Wu, and F. Wu, "Quality of energy provisioning for wireless power transfer," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 527–537, 2015.

[15] J. Tang, A. Liu, J. Zhang, N. Xiong, Z. Zeng, and T. Wang, "A Trust-Based Secure Routing Scheme Using the Traceback Approach for Energy-Harvesting Wireless Sensor Networks," *Sensors*, vol. 18, no. 3, p. 751, 2018.

[16] T. Han and N. Ansari, "Network Utility Aware Traffic Load Balancing in Backhaul-Constrained Cache-Enabled Small Cell Networks with Hybrid Power Supplies," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2819–2832, 2017.

[17] Y. Guo, F. Liu, Z. Cai, N. Xiao, and Z. Zhao, "Edge-Based Efficient Search over Encrypted Data Mobile Cloud Storage," *Sensors*, vol. 18, no. 4, p. 1189, 2018.

[18] X. Liu, G. Li, S. Zhang, and A. Liu, "Big program code dissemination scheme for emergency software-define wireless sensor networks," *Peer-to-Peer Networking and Applications*, vol. 11, no. 5, pp. 1038–1059, 2018.

[19] S. Narayanaswamy, V. Kawadia, R. S. Sreenivas, and P. R. Kumar, "Power Control in Ad-Hoc Networks: Theory, Architecture," in *Proceedings of the Power Control in Ad-Hoc Networks: Theory, Architecture*, pp. 156–162, 2002.

[20] H. Xin and X. Liu, "Energy-balanced transmission with accurate distances for strip-based wireless sensor networks," *IEEE Access*, vol. 5, pp. 16193–16204, 2017.

[21] J. Xu, A. Liu, N. Xiong, T. Wang, and Z. Zuo, "Integrated collaborative filtering recommendation in social cyber-physical systems," *International Journal of Distributed Sensor Networks*, vol. 13, no. 12, p. 155014771774974, 2017.

[22] J. Cui, Y. Zhang, Z. Cai, A. Liu, and Y. Li, "Securing Display Path for Security-Sensitive Applications on Mobile Devices," *CMC, Computers, Materials & Continua*, vol. 55, no. 1, pp. 17–035, 2018.

[23] A. Liu, W. Chen, and X. Liu, "Delay optimal opportunistic pipeline routing scheme for cognitive radio sensor networks," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, p. 155014771877253, 2018.

[24] Z. Li, B. Chang, S. Wang, A. Liu, F. Zeng, and G. Luo, "Dynamic Compressive Wide-Band Spectrum Sensing Based on Channel Energy Reconstruction in Cognitive Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2598–2607, 2018.

[25] M. Z. A. Bhuiyan, G. Wang, J. Wu, J. Cao, X. Liu, and T. Wang, "Dependable Structural Health Monitoring Using Wireless Sensor Networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 4, pp. 363–376, 2017.

[26] M. Huang, Y. Liu, N. Zhang et al., "A Services Routing Based Caching Scheme for Cloud Assisted CRNs," *IEEE Access*, vol. 6, pp. 15787–15805, 2018.

[27] X. Liu, S. Zhao, A. Liu, N. Xiong, and A. V. Vasilakos, "Knowledge-aware Proactive Nodes Selection approach for energy management in Internet of Things," *Future Generation Computer Systems*, 2017.

[28] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Data, "Data Proessing and Text Mining Technologies on Electronic Medical Records: A Review," *Journal of Healthcare Engineering*, Article ID 4302425, 2018.

[29] Y. Liu, M. Dong, K. Ota, and A. Liu, "ActiveTrust: secure and trustable routing in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 2013–2027, 2016.

[30] X. Liu, M. Dong, Y. Liu, A. Liu, and N. Xiong, "Construction Low Complexity and Low Delay CDS for Big Data Codes Dissemination," *Complexity*, vol. 2018, Article ID 5429546, 10 pages, 2018.

[31] X. Liu, N. Xiong, N. Zhang, A. Liu, H. Shen, and C. Huang, "A trust with abstract information verified routing scheme for cyber-physical network," *IEEE Access*, pp. 3882–3898, 2018.

[32] M. Z. A. Bhuiyan, J. Wu, G. Wang, T. Wang, and M. M. Hassan, "e-Sampling: Event-sensitive autonomous adaptive sensing and low-cost monitoring in networked sensing systems," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 12, no. 1, article no. 1, 2017.

[33] S. Zhao and A. Liu, "High performance target tracking scheme with low prediction precision requirement in WSNs," *International Journal of Ad Hoc and Ubiquitous Computing*, 2017, http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijahuc.

[34] T. Wang, J. Zhou, M. Huang et al., "Fog-based storage technology to fight with cyber threat," *Future Generation Computer Systems*, vol. 83, pp. 208–218, 2018.

[35] J. Li, X. Chen, F. Xhafa, and L. Barolli, "Secure deduplication storage systems supporting keyword search," *Journal of Computer and System Sciences*, vol. 81, no. 8, pp. 1532–1541, 2015.

[36] W. Jiang, G. Wang, M. Z. Bhuiyan, and J. Wu, "Understanding Graph-Based Trust Evaluation in Online Social Networks," *ACM Computing Surveys*, vol. 49, no. 1, pp. 1–35, 2016.

[37] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, "Verifiable computation over large database with incremental updates," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 65, no. 10, pp. 3184–3195, 2016.

[38] J. Li, Z. Liu, X. Chen, F. Xhafa, X. Tan, and D. S. Wong, "L-EncDB: A lightweight framework for privacy-preserving data queries in cloud computing," *Knowledge-Based Systems*, vol. 79, pp. 18–26, 2015.

[39] X. Liu, Y. Liu, N. N. Xiong et al., "Construction of Large-Scale Low-Cost Delivery Infrastructure Using Vehicular Networks," *IEEE Access*, vol. 6, pp. 21482–21497, 2018.

[40] S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He, "ATPC: adaptive transmission power control for wireless sensor networks," in *Proceedings of the Proceeding of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 223–236, New York, NY, USA, November 2006.

[41] M. Zuniga and B. Krishnamachari, "An analysis of unreliability and asymmetry in low-power wireless links," in *Proceedings of*

the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON '04), pp. 517–526, October 2004.

[42] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Wireless Networks*, vol. 8, no. 2-3, pp. 153–167, 2002.

[43] M. T. Sun and T.-H. Lai, "Location aided broadcast in wireless ad hoc network systems," in *Proceeding of the IEEE Wireless Communications and Networking Conference Record (WCNC '02)*, vol. 2, pp. 597–602, 2002.

[44] J. Wu and F. Dai, "A generic distributed broadcast scheme in ad hoc wireless networks," *IEEE Transactions on Computers*, vol. 53, no. 10, pp. 1343–1354, 2004.

[45] C. Joo and N. B. Shroff, "On the delay performance of in-network aggregation in lossy wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 662–673, 2014.

[46] P. Jesus, C. Baquero, and P. S. Almeida, "A Survey of Distributed Data Aggregation Algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 381–404, 2015.

[47] X. Li, A. Liu, M. Xie, N. Xiong, Z. Zeng, and Z. Cai, "Adaptive Aggregation Routing to Reduce Delay for Multi-Layer Wireless Sensor Networks," *Sensors*, vol. 18, no. 4, p. 1216, 2018.

[48] Z. Iqbal and H. Lee, "Low-Latency and High-Reliability Cooperative WSN for Indoor Industrial Monitoring," in *Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–6, Sydney, NSW, June 2017.

[49] B. Kang, P. K. H. Nguyen, V. Zalyubovskiy, and H. Choo, "A Distributed Delay-Efficient Data Aggregation Scheduling for Duty-Cycled WSNs," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3422–3437, 2017.

[50] C.-Y. Wan, A. T. Campbell, and L. Krishnamurthy, "PSFQ: A reliable transport protocol for wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 1–11, USA, September 2002.

[51] S. Park, R. Vedantham, R. Sivakumar, and I. F. Akyildiz, "A scalable approach for reliable downstream data delivery in wireless sensor networks," in *Proceedings of the the 5th ACM international symposium*, p. 78, Roppongi Hills, Tokyo, Japan, May 2004.

[52] F. Xiao, W. Liu, Z. Li, L. Chen, and R. Wang, "Noise-tolerant wireless sensor networks localization via multi-norms regularized matrix completion," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2409–2419, 2018.

[53] H. Wen, C. Lin, F. Ren, Y. Yue, and X. Huang, "Retransmission or redundancy: Transmission reliability in wireless sensor networks," in *Proceedings of the IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems, (MASS' 07)*, Italy, October 2007.

[54] H. Teng, X. Liu, A. Liu, H. Shen, C. Huang, and T. Wang, "Adaptive Transmission Power Control for Reliable Data Forwarding in Sensor Based Networks," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

Research Article

EESS: An Energy-Efficient Spectrum Sensing Method by Optimizing Spectrum Sensing Node in Cognitive Radio Sensor Networks

Zilong Jin [iD],[1,2] **Yu Qiao** [iD],[1] **Alex Liu**,[2] **and Lejun Zhang**[3]

[1]*School of Computer and Software, Nanjing University of Information Science and Technology, 219 Ningliu Rd, Nanjing 210044, China*
[2]*Dept. of Computer Science and Engineering, Michigan State University, 428 S Shaw Ln, MI 48824, USA*
[3]*College of Information Engineering, Yangzhou University, 196 Huayang West Rd, Yangzhou 225127, China*

Correspondence should be addressed to Zilong Jin; zljin85@163.com

In cognitive radio sensor networks (CRSNs), the sensor devices which are enabled to perform dynamic spectrum access have to frequently sense the licensed channel to find idle channels. The behavior of spectrum sensing will consume a lot of battery power of sensor devices and reduce the network lifetime. In this paper, we aim to answer the question of how many spectrum sensing nodes (SSNs) are required. In order to achieve this, SSN ratio effects on the accuracy of spectrum sensing from the perspective of network energy efficiency are analyzed first. Based on these analyses, the optimal SSN ratio is derived for maximizing the network lifetime by optimizing the cooperative detection probability (CDP). Simulation results show that the optimal SSN ratio can guarantee the spectrum sensing performance in terms of detection and false alarm probabilities and effectively extend the network lifetime.

1. Introduction

Traditional wireless sensor networks (WSNs) have been an attractive area of research since last decade and usually operate on the license-free industrial, scientific, and medical (ISM) band. However, as the number of wireless devices and applications has seen explosive growth over the past few decades, there has been increasing pressure on the limited spectrum resources [1, 2]. Moreover, the improvement in throughput of wireless networks largely depends on the utilization of channels [3]. In this situation, the concept of cognitive radio (CR) has emerged to alleviate the scarcity of limited radio spectrum resources by improving the utilization of spectrum resources [4]. The CR is a dynamically programmable and configurable radio that opportunistically uses an idle licensed channel in its vicinity to address the problem of unlicensed spectrum resources shortage and the underutilization of the licensed spectrum resources [5, 6]. Such a radio can be applied in the existing network technologies, like wireless sensor networks, machine to machine

networks, wireless body area networks, etc., enabling the PHY (PHYsic) and MAC (Media Access) layers of the devices to automatically and dynamically detect available channels and then accordingly change their transmission or reception parameters to allow more concurrent wireless communications in a given spectrum band. As a specific application of the CR technology, cognitive radio sensor network (CRSN) is recently regarded as one of the most attractive topics in IoT paradigms [7].

Not the same as the traditional WSNs [8–12], CRSNs operate on licensed bands, periodically sense the spectrums, and determine vacant channels. In order to achieve this, CRSDs (cognitive radio-enabled sensor devices) in CRSNs have to frequently sense the licensed channels to identify an idle one and detect the active state of primary users (PUs) signal with strictly limited interference to PUs [13]. On the other hand, different from CR networks [14, 15], CRSNs inherit the basic limitations of traditional WSNs of which the lifetime is strictly constraint due to energy limitations. In addition, spectrum sensing (SS) [16, 17] is a crucial element

in the implementation of a CRSN, and energy consumption is also a major consideration in spectrum sensing. The more SUs that participate in spectrum sensing will result in higher energy consumption of the network and shorter network lifetime. For this reason, our goal is to improve the network energy efficiency by optimizing the number of spectrum sensing nodes (SSNs) that participate in spectrum sensing while still guaranteeing the spectrum sensing accuracy.

Moreover, it has been shown that cooperative spectrum sensing (CSS) can not only deal with multipath fading and shadow effects but also improve the accuracy of spectrum sensing [18–21]. The idea of CSS scheme is to use multiple SUs and combine their sensing results at a fusion center (FC). There are two possible CSS strategies: the first one is that all nodes perform CSS, and the second one is that some nodes sense spectrums. But if their performance is similar, the second one is obviously more suitable. To this end, there are some existing node selection methods in various works [3, 22–27]. Specifically, an optimal hard fusion strategy was proposed to maximize the energy efficiency in [22]. In [23], an optimal number of multihop-based SUs was derived. To minimize the total energy consumption, a closed-form equation and optimal conditions due to *KKT* were proposed in [24] to determine the SUs which sense the spectrum. An energy-efficient CSS was also proposed in [25] to solve the problem of sensing node selection. Taking into consideration the scenario when only partial information of SUs and PUs is available in [26], an energy-efficient SUs selection algorithm has been proposed to save energy and improve the detection performance. In [3], a correlation-aware node selection scheme was proposed to adaptively select uncorrelated nodes for CSS, because of the openness, dynamics, and uncertainty of wireless environment. Moreover, in [27], general criteria for decision-approach selection were analyzed and derived when there are actual channel propagation effects.

However, when the environment of network changes dynamically, fewer nodes cannot guarantee the accuracy of spectrum sensing, and more nodes involving spectrum sensing will increase the energy consumption of the network. Therefore, all of the above existing work cannot ensure the accuracy of spectrum sensing and less energy consumption at the same time. In addition, there is no efficient mathematical model which quantitatively describes the relationship between the number of SSNs and spectrum sensing performance.

In this paper, we analyze the optimal SSN selection strategy from the following three aspects. The first is to explore the relationship between the SSN ratio and the received signal power of PU, i.e., signal-to-noise ratio (SNR), in randomly deployed networks (which means that both of CRSDs and PUs are randomly deployed) for analyzing the number of SSNs impacts on the performance of individual spectrum sensing. Secondly, a mathematical formula which describes the relationship between the ratio of SSN and cooperative detection probability (CDP) is explored to ensure the accuracy of cooperative spectrum sensing. Finally, an optimization function is proposed to derive the optimal SSN ratio which can prolong the network lifetime and guarantee the accuracy of cooperative spectrum sensing.

The remainder of the paper is organized as follows. Section 2 gives the system model and problem definition. Analysis of optimal SSN ratio is formulated in Section 3. Simulation results are discussed in Section 4. Finally, Section 5 concludes the paper and discusses the future work.

2. System Model and Problem Definition

This section first describes the network model for CRSN and then presents the spectrum sensing. Afterwards, this section gives the definition of the problem for the optimal SSN ratio selection.

2.1. Network Model

2.1.1. The Channel Model. The network environment of a CRSN under consideration, where a set of CRSDs, N, are distributed to monitor the area of interest, includes a PU and a FC, as shown in Figure 1. The CRSDs can be regarded as the secondary users (SUs) in traditional CR networks that can access idle channels opportunistically. According to the practical requirements, CRSDs periodically sense the environment with different sampling rates and then report their sensed data to the FC [28]. In general, the available licensed spectrum consists of multiple primary channels whose active state follows different traffic patterns. For simplicity, in this paper, a single primary channel is assumed, and the PU traffic pattern follows a stationary exponential ON/OFF random process [29].

The ON state denotes that the PU is occupying the channel and the OFF state indicates that the channel is idle. Let H_0 and H_1 denote the hypotheses of the absence and present of the PU, respectively. In addition, let V and L denote the exponential random variables, which describe the idle and occupancy state with means v and l, respectively. Therefore, for a channel, the probability of the channel having occupancy P_{on} and the probability of the channel idle P_{off} are given by

$$P_{off} = \frac{v}{(v+l)}, \quad H_0$$
$$P_{on} = \frac{l}{(v+l)}, \quad H_1. \tag{1}$$

2.1.2. The Signal Propagation Model. To address the optimal number of SSNs selection problem conveniently, the signal propagation model is modeled firstly. According to [30], the relationship between the received power of the CRSDs and the transmission power of the PU can be denoted as follows:

$$P_j(d) = \beta d_j^{-\alpha} P_{PU}, \tag{2}$$

where $P_j(d)$ is the received power of the j-th CRSD, P_{PU} is the signal power of the PU, β is a scalar, d_j is the distance between the j-th CRSD and the PU, and α is the path loss factor. The signal-to-noise ratio (SNR) of the PU received at each CRSD is computed by

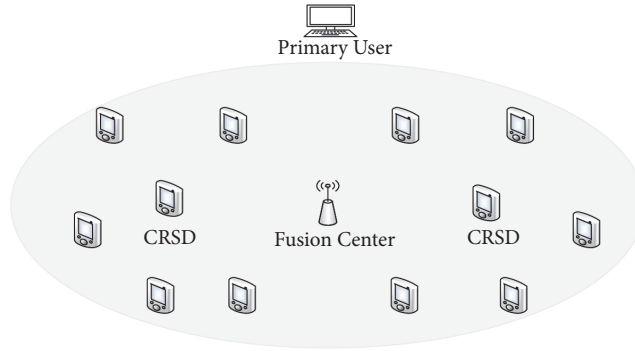$$\gamma_j = 10 \log \frac{P_j(d)}{\sigma^2}, \quad j = 1, 2, \ldots, N, \tag{3}$$

FIGURE 1: The network model of cognitive radio sensor network.

where σ^2 is the noise power and N is the total number of CRSDs in the network.

2.2. Spectrum Sensing

2.2.1. Spectrum Sensing Hypothesis. In a CRSN, the received signal sample of the CRSDs can be formulated as a binary hypothesis testing; the hypothesis H_0 represents that no PU exists in the spectrum, and hypothesis H_1 represents that the PU exists in the spectrum; that is [31],

$$y(m) = \begin{cases} n(m), & H_0 \\ s(m) + n(m), & H_1, \end{cases} \tag{4}$$

where $y(m)$ is the received signal at a CRSD, $s(m)$ is the signal of the PU and is assumed to be an iid random process with zero mean and variance, σ_s^2. The noise, $n(m)$, is assumed to be Gaussian iid random process with zero mean and variance, σ_n^2, and $s(m)$ and $n(m)$ are independent.

In the channel sensing, the probability of detection, P_d, and the probability of false alarm, P_f, are significant indicators for measuring the spectrum sensing accuracy. P_d and P_f are defined as the probabilities of detecting the PU under hypotheses H_1 and H_0, respectively.

2.2.2. PU Detection. The energy detection [32] method can be utilized by the CRSDs as their channel sensing method. The test statistics for energy detector is given by

$$T(y) = \frac{1}{M} \sum_{m=1}^{M} |y(m)|^2, \tag{5}$$

where $T(y)$ is the sample of received signal at CRSDs and $M = 2TW$ where TW represents the product of detection time, T, and signal bandwidth, W. When M is relatively large (e.g., $M > 200$), $T(y)$ can be approximated as a Gaussian random variable under both hypotheses H_0 and H_1, with means μ_0, μ_1 and variances σ_0^2, σ_1^2 respectively, which can be given by [33]

$$\mu_0 = M$$

$$\mu_1 = M(\gamma_j + 1)$$

$$\sigma_0^2 = 2M$$

$$\sigma_1^2 = 2M(2\gamma_j + 1), \tag{6}$$

where γ_j is SNR of the PU signal at the j-th CRSD.

For a given threshold, ϵ, the probability of false alarm and detection probability of the j-th CRSD are given by

$$P_{f,j} = Pr(T(y) > \epsilon \mid H_0) = Q\left(\left(\frac{\epsilon}{\sigma_n^2} - 1\right)\sqrt{M}\right), \tag{7}$$

$$P_{d,j} = Pr(T(y) > \epsilon \mid H_1)$$
$$= Q\left(\left(\frac{\epsilon}{\sigma_n^2} - \gamma_j - 1\right)\sqrt{\frac{M}{2\gamma + 1}}\right), \tag{8}$$

where $Q(\cdot)$ denotes the Gaussian Q-function. Without loss of generality, we set the detection threshold of all the CRSDs to be the same; hence, the false alarm probability of all the CRSDs becomes fixed and is denoted by $\overline{P_f}$. Therefore, (7) and (8) can be represented by (9) and the probability was denoted by [34]

$$P_{d,j} = Pr(T(y) > \epsilon \mid H_1)$$
$$= Q\left(\frac{Q^{-1}(\overline{P_f}) - \sqrt{M}\gamma_j}{\sqrt{2\gamma_j + 1}}\right). \tag{9}$$

2.2.3. Cooperative Spectrum Sensing. In a densely deployed network, we consider that the sensing results of neighboring CRSDs are similar. And the network can be divided into multiple clusters based on the method proposed in Section 3. Therefore, the sensing results in a cluster can be represented by one node which is called SSN in this paper. To reduce the communication overhead, each SSN sends the final 1-bit decision to the FC which is an energy-unconstrained center for data aggregation and decision making and can decode and combine all the decoded information from all SSNs using the logic-OR-rule to make a final decision [35]; that is to say, the PU is considered to be occupying the channel if at least

one sensor node claims the presence of the PU. Then, all the SSNs in different clusters are selected to cooperatively sense the channel; the cooperative detection probability F_d and the cooperative false alarm probability F_f for the channel are as follows:

$$F_f = 1 - \prod_{j=1}^{N_{SSN}} \left(1 - \overline{P_f} \right), \tag{10}$$

$$F_d = 1 - \prod_{j=1}^{N_{SSN}} \left(1 - P_{d,j} \right), \tag{11}$$

where N_{SSN} is the number of selected SSNs cooperating. The cooperative misdetection probability F_m is defined as the probability of not detecting the presence of the PU; i.e., $F_m = 1 - F_d$.

2.3. Problem Definition. This paper focuses on providing the optimal number of SSNs for CSS in the random deployed CRSNs. The selection of optimal SSNs, under the premise that the probability of interfering with the PU should be below a predefined threshold F, must guarantee that all SSNs can accurately detect the available licensed channel information and send their sensed information to the FC for further processing. In other words, there is a constraint on N_{SSN} such that

$$P_{on} * F_m = P_{on} * \prod_{j=1}^{N_{SSN}} \left(1 - P_{d,j} \right) \le F. \tag{12}$$

The ratio of SSN is p, and $0 < p <= 1$. When $p = 1$, it means that all CRSDs are involved in spectrum sensing, and $0 < p < 1$ denotes that some nodes are selected from CRSDs for spectrum sensing. The objective is to select optimal SSN ratio p under the consideration of cooperation detection probability F_d and the constraint of (12) among the whole nodes. Hence, our optimization function is

$$\min_{p \in (0,1]} \quad F_d(p)$$
$$\text{s.t.} \quad P_{on} * F_m \le F, \tag{13}$$

where the objective function $F_d(p)$ is result of a comprehensive consideration of the effect of the SSNs ratio p on the cooperation detection probability F_d, and the relationship between cooperation detection probability F_d and SSNs ratio p will be discussed in Section 3.2.

3. Analysis of the Optimal SSN Ratio

In this section, we first study the impact of the SSN ratio on the PU detection, and the relationship between the SSN ratio and the received signal power of PU is derived. Then a mathematical formula which describes the relationship between the SSN ratio and cooperative detection probability (CDP) is explored to ensure the accuracy of cooperative spectrum sensing. Finally, a cluster method based on spectrum sensing nodes is proposed.

TABLE 1: Table of notations.

Notation	Parameters
N	The total number of nodes
N_{SSN}	The number of SSNs
p	SSNs ratio defined as $p = N_{SSN}/N$
λ_{SSN}	The density of SSNs defined as $\lambda_{SSN} = Np/A = N_{SSN}/A = \lambda p$
λ'	The density of all nodes defined as $\lambda' = \lambda(1 - p)$
R	Transmission radius of the network
r	Transmission radius of SSNs

3.1. The Impact of the SSN Ratio on the PU Detection. In this paper, a densely deployed network is assumed where the connectivity of CRSDs can be guaranteed and each CRSD receives signals from the primary user with different SNRs. Figure 2 shows a specific example where N_{SSN}=2 SSNs randomly deployed in the network, i.e., SSN1 and SSN2. For CRSD1, CRSD2, and CRSD3, because they are close to SSN1, they will join SSN1 by utilizing a clustering method. The other CRSD4, CRSD5, and CRSD6 form another cluster 2 with SSN2 (if $N_{SSN} = N$, it means all of nodes will perform the spectrum sensing; N-N_{SSN} CRSDs will join the nearest SSN as its cluster members). In this way, the SSNs can form a number of clusters and can periodically transmit their sensed data to the FC for further process via data channels. The data channel may not be perfect, but this is beyond the scope of our discussion in this paper.

After the clustering, for example, when primary user appears in cluster 1, SSN is responsible for sensing the spectrum instead of all to perform spectrum sensing and decide the active state of the PU; finally, it shares its spectrum sensing results with the rest of the CRSDs. Therefore, the rest of the CRSDs belonging to cluster 1 do not need to sense the licensed channel to make a decision before transmitting their sensing data. Consequently, the overall spectrum sensing energy in the CRSN can be saved and the lifetime of the network can be prolonged.

Specifically, we assume that all nodes are deployed in a network area A using Poisson point process (PPP) [36] with intensity λ. And then N_{SSN} SSNs are deployed in the network randomly. After clustering process, the network is divided by N_{SSN} Voronoi cells [37, 38]. In order to analyze the signal propagation characteristic in a Voronoi cell, i.e., a cluster, the following parameters are defined in Table 1.

Firstly, the average radius of a cluster is derived. Foss and Zuyew [38] analyzed the geometrical properties of Voronoi cells in a polar coordinate and assumed that a SSN is in 0 point. One of the significant conclusions they got was the expected number of nodes in a Voronoi cell, which can be given as follows (the complete proof of (14) can be seen at [38]).

$$E(N_{tot}) = \lambda' 2\pi \int_0^k l e^{-\lambda_{SSN}\pi l^2} dl = \frac{\lambda'}{\lambda_{SSN}}, \tag{14}$$

where $k \longrightarrow +\infty$, l and $E(N_{tot})$ represent the radius of a cluster and the total number of nodes in a cluster, respectively.
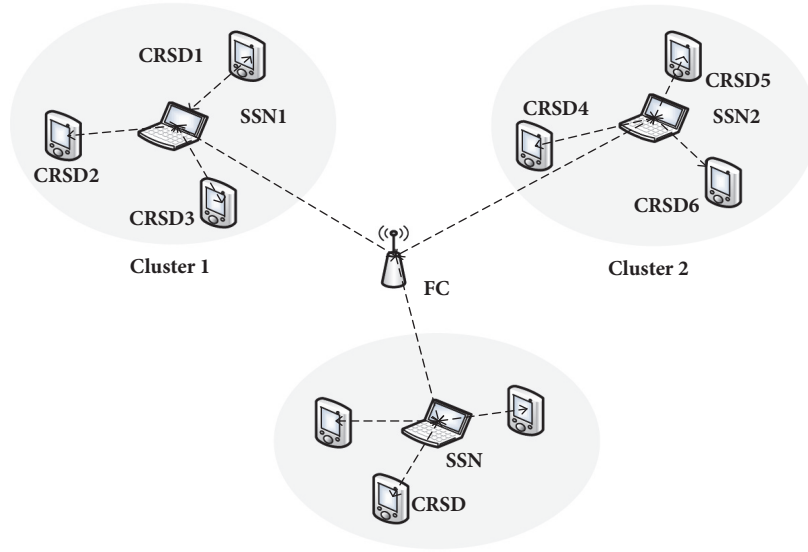
FIGURE 2: The architecture of SSN-based cognitive radio sensor network.

Secondly, the one-hop link lengths between SSN and its cluster member are derived. Since the shape and radius of a cluster are all random, the expected average radius value of a cluster is also random which can be defined as $E[l]$, and the resulting from [38] can be utilized, which showed that the total direct link length of a cluster is $E[L_{tot}] = \lambda'/2\lambda_{SSN}^{3/2}$. Finally, $E[l]$ can be calculated as follows:

$$E[l] = \frac{E[L_{tot}]}{E[N_{tot}]} = \frac{1}{\sqrt{4\lambda_{SSN}}} = \sqrt{\frac{A}{4Np}}, \tag{15}$$

where p represents the ratio of the SSNs to the nodes.

During signal propagation, signal power decreases with distance increase. This phenomenon is statistically modeled using the signal propagation model $P_j(d)$ defined in Section 2.1.2. Let $\overline{P(d)}$ denote the average path loss in a cluster, i.e., $\overline{\gamma} = 10\log(\overline{P(d)}/\sigma^2)$. Based on the result $\overline{\gamma}$ and (14), the average SNR in a cluster can be derived as follows:

$$\overline{\gamma} = 10\log\frac{\overline{P(d)}}{\sigma^2} = 10\log\frac{P\left(\sqrt{A/4Np}\right)}{\sigma^2}, \tag{16}$$

where $P_j(d) = \beta d_j^{-\alpha} P_{PU}$, σ^2 is the noise power, and α and β are determined by the type of approximation and modulation, respectively.

Therefore, $\overline{\gamma}$ can be determined based on the SSNs ratio p and the stationary network parameters, i.e., network area A and the signal power of PU, P_{PU}.

3.2. Optimal SSN Ratio for Maximizing the CDP. The goal of the paper is to improve energy efficiency of the network

by selecting the optimal ratio of SSNs, i.e., p, involved in spectrum sensing while still guaranteeing the spectrum sensing accuracy without interfering with the active state of the PU. Based on the above analysis, analyzing (9) the detection probability is determined by the SNR γ which is also determined by the SSNs ratio p. Hence, the formula of the average detection probability $\overline{P(d)}$ and the SSNs ratio p can derived as follows:

$$\overline{P_d}$$
$$= Q\left(\frac{Q^{-1}\left(\overline{P_f}\right) - 10\sqrt{M}\log\left(P\left(\sqrt{A/4Np}\right)/\sigma^2\right)}{\sqrt{20\log\left(P\left(\sqrt{A/4Np}\right)/\sigma^2\right) + 1}}\right), \tag{17}$$

where $P(\sqrt{A/4Np}) = \beta\sqrt{A/4Np}^{-\alpha}P_{PU}$. Without loss of generality, let $\alpha = 1$ and $\beta = 1$, which are valid to compare the impact of p on P_d; it can yield the following equation:

$$\overline{P_d}$$
$$= Q\left(\frac{Q^{-1}\left(\overline{P_f}\right) - 10\sqrt{M}\log\left(P_{PU}\sqrt{4Np/A}/\sigma^2\right)}{\sqrt{20\log\left(P_{PU}\sqrt{4Np/A}/\sigma^2\right) + 1}}\right). \tag{18}$$

Since the logic-OR-rule is adopted at the FC, according to (11), (18) can be described as the relationship between the cooperative detection probability F_d and the SSNs ratio p, and this is formulated by the following optimization function $F_d^*(p)$:

$$F_d^*(p) = \arg\max_{p\in(0,1]}\left(1 - \prod_{j=1}^{Np}\left(1 - Q\left(\frac{Q^{-1}\left(\overline{P_f}\right) - 10\sqrt{M}\log\left(P_{PU}\sqrt{4Np/A}/\sigma^2\right)}{\sqrt{20\log\left(P_{PU}\sqrt{4Np/A}/\sigma^2\right) + 1}}\right)\right)\right). \tag{19}$$
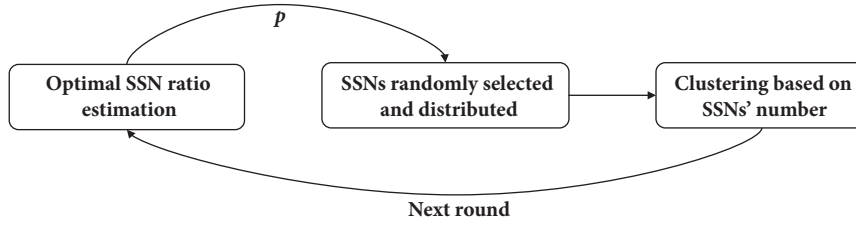
FIGURE 3: Illustration of SSNs selection and cluster formation process.

The objective function in (19) is the final result of the relationship between p and F_d. Therefore, p is obtained which maximizes the cooperation detection probability with the constraint in (12), and the optimization function in (13) can be rewritten as follows:

$$\underset{p \in (0,1]}{\arg\min} \quad F_d^* (p)$$

$$s.t. \quad P_{on} * F_m \le F,$$

(20)

Figure 5 shows the results of (20), which indicates that the optimal SSNs ratio p values are 0.3, 0.15, and 0.1 for nodes densities of 300, 400, and 500, respectively, and the performance of the cooperative detection probability (CDP) is guaranteed.

3.3. The Process of Clustering Based on Spectrum Sensing Node. After the process of the optimal SSN ratio analysis, the optimal p values are obtained at different node densities. In order to guarantee the balance of the network load, we randomly select nodes in the network as SSN nodes to participate in spectrum sensing based on p values. When the SSNs are randomly deployed in the network, a SSN will send a message which contains a counter (initially set to 0) and its ID to neighbor nodes. If the counter is less than a constant C, the ID and counter will be recorded in the neighbor node's memory. Then the neighbor will forward it after increasing the counter by 1 and set a timer. If the timer is out, the node (e.g., CRSD) will join the SSN that has the lowest counter. If there are similar messages, it will select SSN that sent the first message. We illustrate the main idea of addressing the SSNs selection and cluster formation process in Figure 3 and summarize the detailed procedures of our solution in Algorithm 1.

4. Simulation Results

The performance of the proposed schemes is evaluated by extensive simulations on MATLAB. The radius of the simulation area is 100m ∗ 100m, the FC is located in the center, the PU station is assumed to be the DTV station which can cover the area, and the number of randomly deployed CRSDs is 300, 400, and 500. The SSNs are also randomly selected, each node selects the closest SSN to join based on cluster formation algorithm, and then the SSN in a cluster senses the spectrum and transmits its sensing result to the FC for decision making.

TABLE 2

Parameters	Values
Network size	$100m \times 100m$
FC station	(50, 50)
Detection time T	0.001
Signal bandwidth W	3MHz
Sampling frequency	6MHz
Path loss factor α	1
Scalar β	1
Occurrence probability of H_1 P_{on}	0.5
Occurrence probability of H_0 P_{off}	0.5
Initial energy of each node	0.5J
Data packet	4000 bits
Power amplification factor (multipath)	0.0013pJ/bit/m^4
Power amplification factor (free)	10pJ/bit/m^2
Energy consumption per 1-bit data	50nJ/bit

The specific parameters are shown in Table 2.

In Figure 4, we consider that the number of nodes is 300, they are evenly distributed in the network, and the area of interest is covered by the PU signal, and the position of the FC is in the center of the area. The circle represents the nodes and the red five-pointed star represents the FC.

For the CRSN, the most important issue in spectrum sensing is to ensure a high detection probability. In addition, we know that cooperative detection probability is more accurate than single detection probability. Thus the cooperative detection probability and SSN ratio are investigated with our proposed scheme. The relationship between the SSN ratio and the CDP is shown in Figure 5. From the figure, it can be seen that the optimal SSN ratio for the three densities is 0.3, 0.15, and 0.1, respectively. As the number of the nodes increases, the nodes in the network have better connectivity. Therefore, the proportion of the SSNs needed for CSS is reduced. In addition, it also can be found from the figure that the more the number of SSNs that are selected to participate in cooperative spectrum sensing, the closer the cooperative detection probability to 1.

After the calculation and analysis of optimal number of SSN nodes, Figure 6 shows the distribution of spectrum sensing nodes, on all the subfigures, at different nodes densities.

According to the proposed cluster formation Algorithm 1, Figure 7 shows the relationship between the number of

Input: Optimal SSNs ratio p based on the Eq. (20);
　　　　　All number of nodes in the network N;
　　　　　The rest number of nodes in the network M^*.
initialization: counter $c = 0$; constant C;
　　　　$N_{SSN} \longleftarrow Np, M^* \longleftarrow N - Np$
　　　　for $i \in \{1 \ldots N^*\}$ **do**
　　　　　　randomly selected SSN i sends a message containing a counter c and its ID to neighbor node
　　　　end of for
　　　　repeat until timer stops
　　　　　　for $i \in \{1 \ldots N_{SSN}\}$ **do**
　　　　　　　Timer i is on;
　　　　　　　$c \longleftarrow c + 1$;
　　　　　　　if $c <= C$
　　　　　　　then The neighbor node of SSN i forward its ID and count c;
　　　　　　end of for
　　　　end of repeat
　　　　for $j \in \{1 \ldots M^*\}$ **do**
　　　　　　for $i \in \{1 \ldots N_{SSN}\}$ **do**
　　　　　　　if node j has the lowest count c among every SSN in $\{1 \ldots N_{SSN}\}$
　　　　　　　then node j join the SSN i
　　　　　　end of for
　　　　end of for
　　　　for $j \in \{1 \ldots M^*\}$ **do**
　　　　　if node j has similar messages among every node in $\{1 \ldots M^*\}$
　　　　　then node j select the SSN i who sent the first message
　　　　end of for
Output: clusters $i, \ldots, N_{SSN}$

ALGORITHM 1: Cluster formation algorithm based on optimal SSN selection.

○ Spectrum Sensing Nodes
☆ Fusion Center
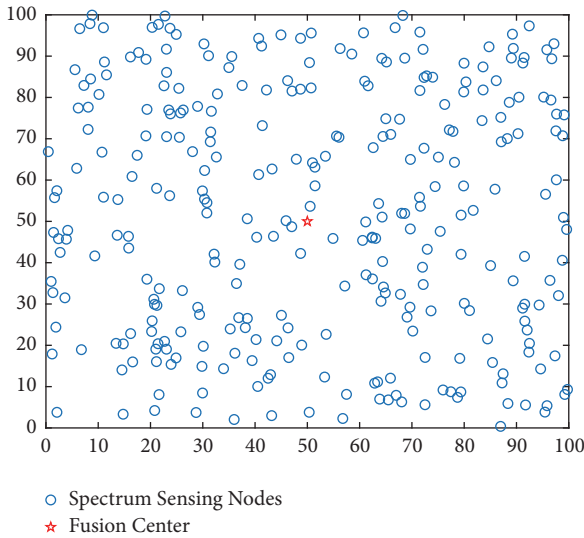
FIGURE 4: Nodes of CRSN deployment diagram ($N = 300$).

—□— N=300
—◇— N=400
—×— N=500
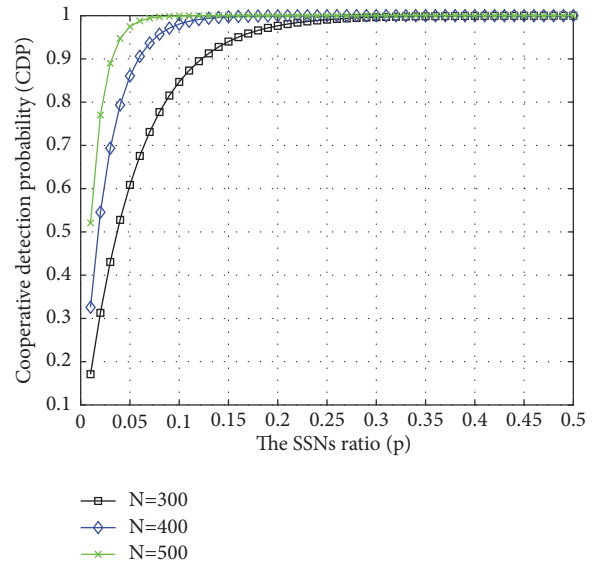
FIGURE 5: The impact of SSNs ratio p on the cooperative detection probability (CDP).

iterations and the number of SSN nodes when the number of network nodes is 300, 400, and 500, respectively.

Figure 7 reveals that, on all the subfigures, the greater the node density, the lower the number of nodes that need to be used as SSNs which also represent the number of clusters and the longer the network lifetime, which confirms that cluster formation algorithm can effectively and reasonably partition the network to prolong the lifetime of the network.

In Figure 8, we show the receiver operating characteristic (ROC) curves of two different methods for CSS. The first indicates the simulated and theoretical values when all nodes

(a) The number of nodes deployed is 300

(b) The number of nodes deployed is 400
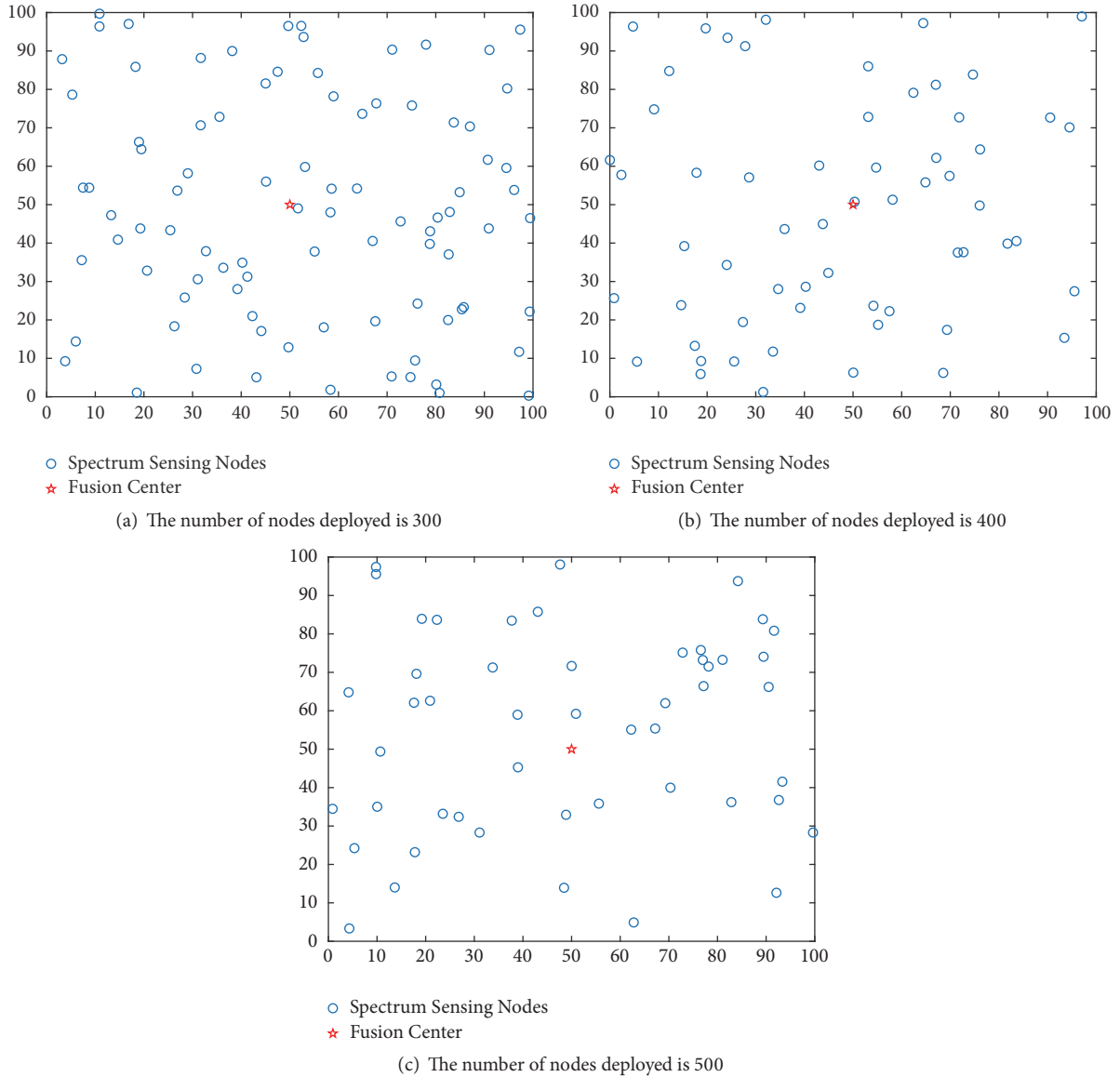

(c) The number of nodes deployed is 500

FIGURE 6: The distribution of spectrum sensing nodes at different node densities.

perform spectrum sensing, while the second represents the simulated and theoretical values only when the selected SSNs participate in spectrum sensing. It can be seen that the simulated values and the theoretical values of these two methods are close to each other, which indicates that the derivation of (10) and (11) is correct. At the same time, it is proved that only a part of nodes is selected to perform spectrum sensing; the detection probability of the network does not decrease, which also indicates that our scheme is effective.

To evaluate the impact of the optimal SSN ratio selection strategy on the energy efficiency of the network, the simulation is performed for 5000 rounds, and Figure 9 shows that the proposed optimization strategy can greatly improve the energy efficiency of the network and thus prolong the lifetime of the network.

Finally, it can be clearly seen that the proposed optimal SSN ratio selection strategy not only can optimize the number of nodes that participate in spectrum sensing while still guaranteeing the spectrum sensing accuracy without interfering in the active state of the primary user, but also can prolong the lifetime of the network due to the reduced number of nodes for SS.

5. Conclusions

In this paper, an optimal SSN ratio selection scheme is proposed to improve the network energy efficiency by optimizing the number of nodes that participate in spectrum sensing while still guaranteeing the spectrum sensing accuracy without interfering in the active state of the primary user. Firstly, the mathematical analysis of the relationship

(a) The number of nodes deployed is 300

(b) The number of nodes deployed is 400
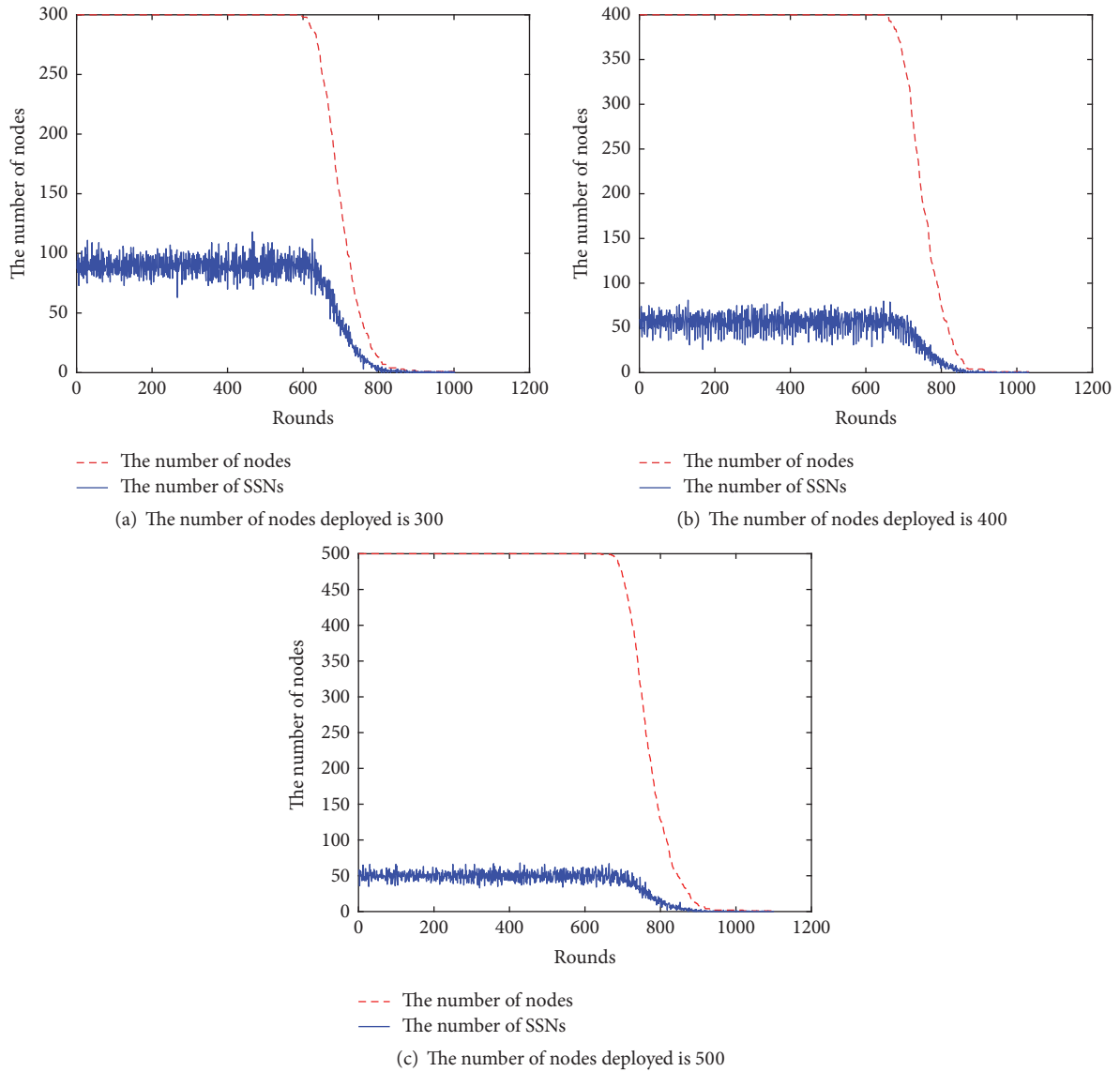
(c) The number of nodes deployed is 500

FIGURE 7: Comparison of cluster numbers at different nodes densities.

between SNR and the SSN ratio is carried out, and then an optimization problem is given to choose the optimal SSN ratio for cooperative spectrum sensing to maximize the cooperative detection probability. Finally, the simulation results demonstrate that this scheme not only can choose optimal SSN ratio to maximize the cooperative detection probability but also can effectively reduce the energy consumption to prolong the network lifetime. To the best of our knowledge, this is the first time to associate the SSN ratio with cooperative detection probability for CSS. Our future work is to explore the relationship between the selection of optimal SSNs and more decision-making schemes.

Data Availability

The data shown in the manuscript is available for the readers.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
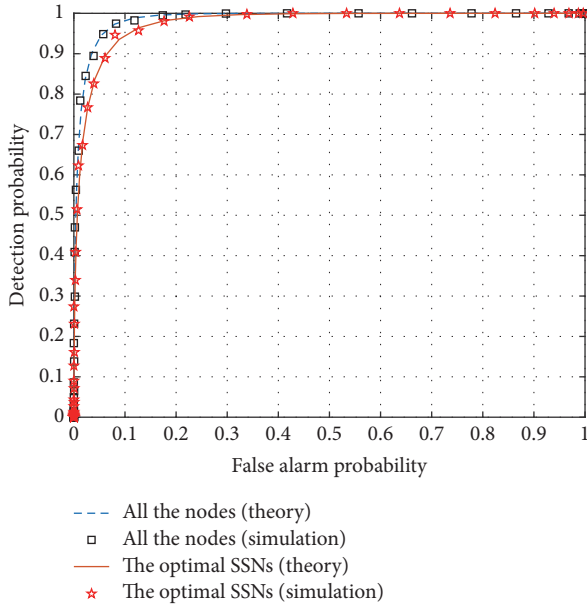
Acknowledgments

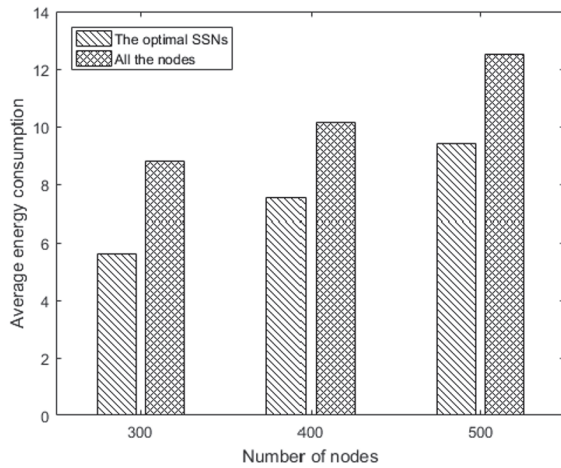FIGURE 8: Detection performance at different number of nodes.

FIGURE 9: Average energy consumption at different nodes densities.

References

[1] J. Cheng, R. Xu, and X. Tang, "An Abnormal Network Flow Feature Sequence Prediction Approach for DDoS Attacks Detection in," *Big Data Environment , Computers, Materials Continua*, vol. 55, no. 1, pp. 95–119, 2018.

[2] W. Liu, X. Luo, Y. Liu et al., "Location algorithm of indoor wi-fi access points based on signal strength relative relationship and region division," *Computers, Materials and Continua*, vol. 55, no. 1, pp. 71–93, 2018.

[3] A. S. Cacciapuoti, I. F. Akyildiz, and L. Paura, "Correlation-aware user selection for cooperative spectrum sensing in cognitive radio ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 297–306, 2012.

[4] E. Hossai and V. K. Bhargava, "Cognitive wireless communication networks," *Springer Science and Business Media*, vol. 16, no. 16, pp. 112-113, 2007.

[5] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[6] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: a survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, 2011.

[7] O. B. Akan, O. B. Karli, and O. Ergul, "Cognitive radio sensor networks," *IEEE Network*, vol. 23, no. 4, pp. 6152–6159, 2009.

[8] V. Srinivasan, C.-F. Chiasserini, P. S. Nuggehalli, and R. R. Rao, "Optimal rate allocation for energy-efficient multipath routing in wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 891–899, 2004.

[9] M. Zheng, H. Yu, J. Zheng, and P. Zeng, "Tradeoff between utility and lifetime in energy-constrained wireless sensor networks," *Control Theory and Technology*, vol. 8, no. 1, pp. 75–80, 2010.

[10] J. Chen, W. Xu, S. He, Y. Sun, P. Thulasiraman, and X. Shen, "Utility-based asynchronous flow control algorithm for wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1116–1126, 2010.

[11] J. Zhu, K.-L. Hung, B. Bensaou, and F. Nait-Abdesselam, "Rate-lifetime tradeoff for reliable communication in wireless sensor networks," *Computer Networks*, vol. 52, no. 1, pp. 25–43, 2008.

[12] H. Cheng, Z. Su, N. Xiong, and Y. Xiao, "Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model," *Information Sciences*, vol. 329, pp. 461–477, 2016.

[13] E. Larsson and M. Skoglund, "Cognitive radio in a frequency-planned environment: some basic limits," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4800–4806, 2008.

[14] Y. Pei, Y. Liang, K. C. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: optimal sensing strategy, power allocation, and sensing order," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, 2011.

[15] Y. Wu and D. H. K. Tsang, "Energy-efficient spectrum sensing and transmission for cognitive radio system," *IEEE Communications Letters*, vol. 15, no. 5, pp. 545–547, 2011.

[16] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: state-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[17] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[18] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, 2009.

[19] G. Noh, H. Wang, J. Jo, B.-H. Kim, and D. Hong, "Reporting order control for fast primary detection in cooperative spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 4058–4063, 2011.

[20] Y. Zou, Y. D. Yao, and B. Zheng, "Cooperative relay techniques for cognitive radio systems: spectrum sensing and secondary user transmissions," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 98–103, 2012.

[21] Q.-T. Vien, G. B. Stewart, H. Tianfield, and H. X. Nguyen, "Efficient cooperative spectrum sensing for three-hop cognitive wireless relay networks," *IET Communications*, vol. 7, no. 2, pp. 119–127, 2013.

[22] S. Maleki, S. P. Chepuri, and G. Leus, "Optimal hard fusion strategies for cognitive radio networks," in *Proceedings of the*

IEEE Wireless Communications and Networking Conference (WCNC '11), pp. 1926–1931, Cancún, Mexico, March 2011.

[23] A. Singh, M. R. Bhatnagar, and R. K. Mallik, "Performance of an improved energy detector in multihop cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 2, pp. 732–743, 2016.

[24] M. Najimi, A. Ebrahimzadeh, S. M. H. Andargoli, and A. Fallahi, "A novel sensing nodes and decision node selection method for energy efficiency of cooperative spectrum sensing in cognitive sensor networks," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1610–1621, 2013.

[25] A. Ebrahimzadeh, M. Najimi, S. M. H. Andargoli, and A. Fallahi, "Sensor selection and optimal energy detection threshold for efficient cooperative spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1565–1577, 2015.

[26] M. Najimi, A. Ebrahimzadeh, S. M. H. Andargoli, and A. Fallahi, "Energy-efficient sensor selection for cooperative spectrum sensing in the lack or partial information," *IEEE Sensors Journal*, vol. 15, no. 7, pp. 3807–3818, 2015.

[27] A. S. Cacciapuoti, M. Caleffi, L. Paura, and R. Savoia, "Decision maker approaches for cooperative spectrum sensing: Participate or not participate in sensing?" *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2445–2457, 2013.

[28] J. Ren, Y.-X. Zhang, and K. Liu, "Multiple k-hop clusters based routing scheme to preserve source-location privacy in WSNs," *Journal of Central South University*, vol. 21, no. 8, pp. 3155–3168, 2014.

[29] N. Zhang, H. Liang, N. Cheng, Y. Tang, J. W. Mark, and X. S. Shen, "Dynamic spectrum access in multi-channel cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2053–2064, 2014.

[30] M. Pan, P. Li, Y. Song et al., "Spectrum clouds: a session based spectrum trading system for multi-hop cognitive radio networks," in *Proceedings of the IEEE INFOCOM '12*, vol. 131, pp. 1557–1565, pp. 1557-1565, Orlando, FL, USA, March 2012.

[31] A. Ghasemi and E. S. Sousa, "Opportunistic spectrum access in fading channels through collaborative sensing," *Journal of Communications*, vol. 2, no. 2, pp. 71–82, 2007.

[32] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, 2007.

[33] D. Cabric, A. Tkachenko, R. W. Brodersen et al., "Experimental study of spectrum sensing based on energy detection and network cooperation," in *Proceedings of the TAPAS 2006*, vol. 73, pp. 527–533, 2006.

[34] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.

[35] A. Singh, M. R. Bhatnagar, and R. K. Mallik, "Cooperative spectrum sensing in multiple antenna based cognitive radio network using an improved energy detector," *IEEE Communications Letters*, vol. 16, no. 1, pp. 64–67, 2012.

[36] M. Kaynia, N. Jindal, and G. E. Øien, "Improving the performance of wireless ad hoc networks through MAC layer design," *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 240–252, 2011.

[37] S. N. Chiu, D. Stoyan, W. S. Kendall et al., *Stochastic Geometry and Its Applications*, vol. 45, John Wiley and Sons, 2013.

[38] S. G. Foss and S. A. Zuyev, "On a Voronoi aggregative process related to a bivariate Poisson process," *Advances in Applied Probability*, vol. 28, no. 4, pp. 965–981, 1996.

Research Article

Controllable Effective Threshold Based Fusion Coverage Algorithm in Mobile Sensor Networks

Yong Lu (iD) **and Na Sun**

School of Information Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Yong Lu; amycun@126.com

Academic Editor: Jaime Lloret

The coverage quality and network lifetime are two key parameters in the research of sensor networks. The coverage quality shows direct influences on the network lifetime. Meanwhile, it is influenced by many other factors such as physical parameters and environmental parameters. To reveal the connection between the coverage quality and the parameters of target node concerned, a fusion coverage algorithm with controllable effective threshold is proposed based on the sensing probability model. We give the model for the membership function of coverage intensity as well as the prediction model for the fusion operator. The range for the effective threshold is presented according to the membership function model. Meanwhile, the maximum of the effective coverage intensity for the target nodes within the monitoring area is derived. The derivation of the maximal fusion coverage intensity is elaborated utilizing a processing function on the distances from the target node to the ones in the sensor node set. Furthermore, we investigate different network properties within the monitoring area such as network coverage quality, the dynamic change of parameters, and the network lifetime, based on the probability theory and the geometric theory. Finally, we present numerical simulations to verify the performances of our algorithm. It is shown under different settings that, compared with the demand coverage quality, the proposed algorithm could improve the network coverage quality by 15.66% on average. The simulation experiment results show that our proposed algorithm has an average improvement by 10.12% and 13.23% in terms of the performances on network coverage quality and network lifetime, respectively. The research results are enlightening to the edge coverage and nonlinear coverage problems within the monitoring area.

1. Introduction

The wireless sensor networks (WSNs) are constructed by thousands of sensor nodes in a self-organized and multihop way. The WSN accomplishes data collection, data calculation, and data communication as well as data storage and achieves an organic unification between information of the physical surroundings [1]. The development of the information technology has enabled the appliance of WSNs to various areas like national defense, environment monitoring, rescuing, smart home, e-health, agricultural production, and transportation [2].

Network energy consumption and the coverage quality are two crucial problems for the research of WSNs [3]. The coverage quality determines the monitoring effectiveness for the mobile target. The behavioral characteristic of the coverage quality mainly manifests in the node deployment.

The key to the network energy consumption problem is to effectively control the rapid energy exhaustion, extend the entire network lifetime, and enhance the quality of service (QoS). The network energy efficiency and coverage quality are mainly determined by the reasonability of the deployment of the sensor nodes [4]. Normally, restrained by environmental factors such as the landscape, the sensor nodes are deployed in a random way. No a priori information on the geological information of nodes is available because of the randomness. Meanwhile, multiple sensor nodes may exist within a certain monitoring area or monitoring spot; that is, k-coverage comes into existence. In another case, the randomness of deployment may also lead to uncovered monitoring areas or monitoring spots, that is, coverage dead zones which can only be eliminated by introducing more deployed sensor nodes. Coverage is guaranteed for both cases described above. However, some disadvantages inevitably

exist. Firstly, the high-density clustering of a massive number of sensor nodes would cause a large amount of redundant information and the congestion of the communication link, which restrains the scalability of the network, undermines the network QoS, and shortens the network lifetime. Secondly, real-time monitoring and coverage for the target node concerned cannot be guaranteed when it is located in the coverage dead zone. As a result, the sensor nodes fail to transmit accurate data information to aggregation nodes, which further leads to bias and uncertainty in the collected data information of the aggregation node. Furthermore, the sensor nodes exhibit heterogeneity after several operation cycles, which cause the changes of the node coverage area. Once the target node concerned is no longer covered, the data information collected for the target node will become meaningless. Finally, the nature of coverage is the effective coverage for the target node concerned, instead of all the target nodes. The coverage effectiveness is elaborated as follows. As the target node concerned moves through the area, head nodes accomplish the coverage while its joint nodes are awakened to achieve joint coverage and transmit the collected date information on the mobile target node to the aggregation node. For a deeper understanding of the problem described above, a fusion coverage algorithm based on effective threshold is proposed which could maximize the coverage area with the least sensor nodes and ensure the real-time coverage for the target node concerned.

Based on the discussion above, we perform investigations on the two key problems of the sensor network, that is, the coverage problem and the network lifetime. However, there are still some deviations between the theoretical results and the realistic engineering results for the coverage quality and the network lifetime. Therefore, accurately locating the target nodes, enhancing the coverage intensity for a certain target node, and improving the coverage quality for the target node concerned utilizing fusion coverage are still three open problems which remain to be solved.

To solve these problems, we employ the membership function of coverage intensity and the prediction model for the fusion operator. In order to have a more accurate estimation for the network coverage quality and extend the system lifetime, the analyses are then performed on the range for the effective threshold of the membership function and the maximization problem for the effective coverage intensity as well as the tunable range for the dynamic parameters. Meanwhile, the derivation of the maximal fusion coverage intensity is elaborated utilizing a processing function on the distances from the target node to the ones in the sensor node set. Generally speaking, the proposed algorithm is enlightening for improving energy efficiency for sensor nodes and extending the network lifetime.

2. Related Works

The coverage for the target node is accomplished by the cooperation of many sensor nodes. Many experiments have proven that as the target node enters the monitoring area,

sensor nodes could sense the target node and then accomplish the full coverage for the target node. However, effective coverage, instead of full coverage, is sufficient for practical applications. The reason is that full coverage is accomplished at the cost of too much network energy consumption which accelerates the network collapse. Recently, the coverage problem of the sensor network has inspired much interest. An ID authentication based optimized coverage protocol was proposed in [5] where the coverage over multiple target nodes is accomplished by different IDs. Meanwhile, the derivation for the coverage quality was presented when multiple target nodes are covered. The different coverage angles of the sensor nodes were exploited by [6] to derive the required area for the coverage sector. Then the coverage quality for the target node from a certain angle can be obtained. A coverage algorithm for multiple targets was presented in [7] based on linear programming and the clustering structure. The optimized coverage over multiple targets is achieved via the calculations of the network coverage probability and the remaining energy for the nodes. The artificial intelligence algorithm was employed in [8] during the coverage process of the sensor nodes, that is, the swarm intelligence algorithm. The complete deployment of the sensor nodes in the entire network monitoring area is accomplished. For the optimization period of the coverage probability, these two intelligent algorithms are employed iteratively for the optimization of the network coverage probability refinement function. Finally, the complete coverage over the monitoring area is achieved. A coverage control algorithm was proposed based on node scheduling strategies [9]. The network coverage quality to demand coverage quality ratio is taken as the indicator of the network performance. Exploiting this relationship, a scheduling strategy was established for the sensor nodes. As a result, the connectivity and coverage over the entire network can be guaranteed with the minimal number of nodes. An enhanced coverage control algorithm (ECCA) was proposed in [10]. The derivation for the expectation of the coverage quality was elaborated when the coverage over the monitoring area was guaranteed. The functional relationships among different parameters were also verified when the random variables were mutually independent. It was also proven that the coverage over the monitoring area can be achieved effectively by adjusting the tunable parameters. The mobility characteristic of the sensor nodes was utilized by [11] to achieve the coverage over the target node, assuming that the network connectivity is guaranteed. Two heuristic algorithms were proposed in [11] where the Voronoi diagram was adopted as the model for the active area of the target nodes. The energy consumption for the sensor nodes is then reduced to accomplish the network energy balance and further improve the network coverage probability. A fence coverage algorithm was proposed in [12] to maximize the network lifetime. This algorithm starts once the moving target node crosses the monitoring area. The correlation between the neighbors and the sensor node is employed to achieve the consistent coverage on the mobile target node. Finally, the complete coverage over the target is achieved. A distributed connectivity and coverage maintenance algorithm was proposed in [13]. In this algorithm,

the redundant nodes are awakened by the working nodes and the network connectivity is therefore guaranteed. The effective coverage over the target node is maintained with only a small number of mobile sensor nodes. Therefore, the energy consumption caused by the long-distance monitoring can be controlled and the network lifetime can be prolonged. An energy-efficient target coverage algorithm was proposed in [14], which constructs the network clusters based on non-linear programming. The energy evaluation is performed for each cluster node. The sensor nodes satisfying the coverage condition are chosen as elements of the optimal coverage set. Therefore, the coverage effect is optimized and the network lifetime is prolonged. A probability-driven mechanism based coverage algorithm was proposed in [15]. This algorithm first establishes the coverage model for the sensor nodes and the target nodes. Then the coverage ratio between the sensor nodes and the target nodes is calculated based on probability theory. Finally, the state transitioning process for the nodes is accomplished according to the node scheduling algorithm. As a result, the network lifetime is prolonged.

All the algorithms stated above could accomplish the effective coverage over the monitoring area under some particular conditions. However, they exhibit some disadvantages. For example, the algorithms in [5–7] are high in terms of computational complexity. The maximization of the network lifetime was not considered in [8–10]. The refinement functions in [11–13] vary significantly, which leads to inaccurate calculation results, while the models in [14, 15] are too ideal for practical engineering applications. Therefore, we employ the sensing probability model and propose a fusion coverage algorithm based on controllable effective threshold. This algorithm achieves the network fusion coverage by adjusting the tunable threshold parameter and efficiently prolongs the system lifetime.

The structure of this paper is described as follows. Section 3 performs a theoretical analysis on the network model and presents related definitions and the energy decay function for the sensor nodes. The CETFC algorithm is elaborated [16]. Furthermore, the membership function is employed to indicate the coverage intensity. For a further improvement on the coverage quality for the target node concerned, the coverage over the monitoring area is achieved by adjusting the effective threshold. The functional relationship for the distances between the nodes in the sensor node set and the target node is employed to derive the maximum of the fusion coverage intensity. We then display numerical simulation results in Section 4. Compared with other algorithms, the proposed CETFC algorithm exhibits prominent effectiveness and feasibility. Finally, we conclude this paper in Section 5.

3. Methodology

3.1. Basic Definitions. For a better study on the coverage problem, we make simplifications by introducing the five following assumptions [10].

Assumption 1. At the initial phase, the sensing ranges and communication ranges for all the nodes are in the shape of a circle while the communication radius is twice the value of the sensing radius.

Assumption 2. The sensor node set lies within the monitoring area. All the nodes are synchronized during the working phase. The boundary effect can be neglected.

Assumption 3. The location information for each node can be obtained via the positioning algorithms while no extra positioning device is needed.

Assumption 4. At the initial phase, all sensor nodes possess the same sensing coverage intensity membership function.

Assumption 5. The sensing area of the sensor nodes is far smaller than the monitoring area.

Definition 6 (coverage intensity). Assuming that the location of an arbitrary sensor node s_i is (x_s, y_s) while that of the target node t_i is (x_t, y_t), then the coverage intensity from s_i to t_i is defined as

$$
I(s_i, t_i) = \begin{cases} 0 & d(s_i, t_i) \geq r + r_m \\ e^{-\alpha[d(s_i,t_i)-r]^\beta} & r < d(s_i, t_i) < r + r_m \\ 1 & d(s_i, t_i) \leq r, \end{cases} \quad (1)
$$

where α and β are physical parameters with $\alpha \in (0, +\infty)$ and $\beta \in (0, 1]$, r is the sensing radius for the sensor node after decay, r_m is the decay distance for the sensor nodes, $r + r_m$ is the sensing radius of the sensor nodes, and $d(s_i, t_i)$ indicates the Euclidean distance from the target node to the sensor node.

Definition 7 (effective coverage intensity). Assuming that the effective threshold is e_{th} with $e_{th} \in [0, 1]$, when the coverage intensity exceeds e_{th}, that is, $I(s_i, t_i) > e_{th}$, the coverage intensity $I(s_i, t_i)$ is defined as effective coverage intensity and denoted as $I_e(s_i, t_i)$. Note that the effective threshold $e_{th} = \exp(-\alpha r)^\beta$ if and only if $d(s_i, t_i) = 0$.

Definition 8 (membership function). The function employed to indicate the effective coverage intensity is defined as the effective coverage membership function and denoted as $u(x, y)$. The range of the membership function is $[0, 1]$. The physical meaning of the membership function is the membership degree of the target node with respect to the coverage from the sensor node. A larger membership function indicates a higher membership degree, and vice versa.

Definition 9 ((M, C_{th}) fusion coverage). The fusion coverage intensity from the nodes in the sensor node set M to an arbitrary target node within the monitoring area is denoted as I_M. Assuming that the effective coverage threshold is C_{th} with $C_{th} \in (e_{th}, 1]$ and $I_M \geq C_{th}$, then the sensor node set M fusion covers the target node, which is denoted as (M, C_{th}). Note that the sensor node set M is composed of m sensor nodes.

The signal is transmitted over the air and received at the sensor nodes [17, 18] within the monitoring area. During the transmission, the signal is attenuated by environmental factors such as physical obstacles, which would lead to uncertainty for the received signal at the sensor nodes [19–21]. Furthermore, this will cause errors in the calculation results. Affected by the factors above, the sensing radius of the sensing node mainly depends on the environmental factors in the monitoring area [22, 23]. During the coverage process from the sensor node to the target node, the loss of energy due to distance is expressed in the exponential form.

$$P_L(d) = P_L(d_0) + 10n \log_{10}\left(\frac{d}{d_0}\right) + N_\sigma, \tag{2}$$

where d indicates the Euclidean distance between the sensor node and the target node, d_0 is the reference range physical parameter, n stands for the energy decay parameter due to the link distance between the sensor node and the target node, N_σ is a random variable with mean 0 and variance σ^2, and $P_L(d_0)$ is the function for the average decay loss with reference point d_0, while $P_L(d_0)$ is the average decay loss function with reference point d.

The received signal at the sensor node can be modelled as follows:

$$P_r = P_t - P_L(d_0) - 10n \log_{10}\left(\frac{d}{d_0}\right) - N_\sigma, \tag{3}$$

where P_r is the power of the received signal and P_t is the transmission power. Due to the impacts of the environmental factors, the power of received signals from different receiving angles is different, even for the sensor nodes at the same location.

3.2. Fusion Coverage with Restricted Threshold.

The sensor node set composed of m sensor nodes is denoted as M; that is, $M = \{s_1, s_2, s_3 \ldots, s_m\}$. For an arbitrary target node t within the monitoring area, the coverage intensity for the target node t is $I(s_i, t)$. According to Definition 9, the fusion coverage intensity of the set M at node t is I_M. Define the fusion operator as $F : [0, 1]^M \to [0, 1]$. Then the effective fusion coverage intensity is

$$\begin{aligned} I_M &= F[I(s_1, t), I(s_2, t), \ldots, I(s_m, t)] \\ &= \min\left(1, \frac{1}{\lambda}\left(\prod_{i=1}^{M}[1 + \lambda I(s_i, t)] - 1\right)\right), \end{aligned} \tag{4}$$

in which λ is a ratio with range $\lambda \in [-1, 0)$. When the fusion coverage intensity from the set M to the target node is no smaller than C_{th}, the target node is (M, C_{th}) covered. When the coverage intensity diminishes, the value of λ can be adjusted to improve the coverage intensity and achieve the effective coverage over the target node [3].

According to (4), k nodes within the set M could be found out to achieve the maximal fusion coverage intensity as follows:

$$I_k(M, t) = \max_{\{s_1, s_2 \cdots s_k\} \in M} F[I(s_1, t), I(s_2, t) \cdots I(s_k, t)]. \tag{5}$$

According to (5), during the coverage from the set M to the target node set T, the minimal fusion coverage intensity is

$$\begin{aligned} &I_{k\_T}(M, T) \\ &= \min_{t \in T}\left\{\max_{\{s_1, s_2 \cdots s_k\} \in M} F[I(s_1, t), I(s_2, t) \cdots I(s_k, t)]\right\}. \end{aligned} \tag{6}$$

Theorem 10. *When m equidistant sensor nodes perform the fusion coverage for the target node, the distance d from the sensor node to the target node should satisfy $d \leq r + (-1/\lambda \ln(C_{th}/\sqrt[q]{m}))^{1/\beta}$, where q is the fusion factor with $q \geq 2$.*

Proof. When m equidistant sensor nodes perform the fusion coverage for the target node, the effective fusion coverage intensity should satisfy

$$I_m = F(I(s_1, t), I(s_2, t), I(s_3, t) \cdots I(s_m, t)) \geq C_{th}. \tag{7}$$

According to (4),

$$I_m = \min\left[1, \sum_{i=1}^{m} I(s_i, t)^q\right]^{1/q} \geq C_{th}. \tag{8}$$

Since the m sensor nodes are equidistant, that is,

$$I(s_1, t), I(s_2, t), I(s_3, t) \cdots I(s_m, t) = I_m, \tag{9}$$

substituting (9) into (8), we have

$$\frac{C_{th}}{\sqrt[q]{m}} \leq I_m. \tag{10}$$

According to Definition 6, when $r < d < r + r_m$, the value of I_m is

$$I_m = \exp\left\{-\lambda[d(s, t) - r]^{-\beta}\right\}. \tag{11}$$

Substituting (11) into (10), we have

$$I_m = \exp\left\{-\lambda[d(s, t) - r]^\beta\right\} \geq \frac{C_{th}}{\sqrt[q]{m}}. \tag{12}$$

We further simplify (12):

$$d \leq r + \left(\frac{-1}{\lambda \ln(C_{th}/\sqrt[q]{m})}\right)^{1/\beta}. \tag{13}$$

Therefore, the proof is completed. □

Corollary 11. *When m equidistant sensor nodes perform the effective coverage for the target node, the critical distance for the effective coverage intensity is*

$$d = r + \left(\frac{-1}{\lambda \ln(C_{th}/\sqrt[q]{m})}\right)^{1/\beta}. \tag{14}$$

Proof. According to Theorem 10, the coverage from the equidistant sensor nodes is equivalent. The same is true for the effective coverage intensity. In other words, $I(s_1, t) =$

Parameter	Value
Area I	$100 * 100$
Area II	$200 * 200$
Area III	$300 * 300$
r	5 m
Time	100 s
R	10 m
$E_{R\text{-elec}}$	50 J/b
$E_{T\text{-elec}}$	50 J/b
ε_{fs}	10 (J/b)/m^2
Energy	2 J

$I(s_2, t) = \cdots = I(s_m, t) = I_m$. For the set composed of m sensor nodes, the equivalent distance is

$$d = r + \left(\frac{-1}{\lambda \ln \left(C_{th} / \sqrt[\beta]{m} \right)} \right)^{1/\beta}. \qquad (15)$$

This is exactly the critical distance for an arbitrary sensor node. Therefore, the proof is completed. □

4. Experimental Result

The energy loss at the transmitting end is modeled as follows:

$$
\begin{aligned}
E_{Tx}(l, d) &= lE_{T\text{-elec}} + E_{amp}(l, d) \\
&= \begin{cases} lE_{T\text{-elec}} + l\varepsilon_{fs} d^2, & d < d_0 \\ lE_{T\text{-elec}} + l\varepsilon_{amp} d^4, & d \geq d_0. \end{cases}
\end{aligned} \qquad (16)
$$

The energy loss model for the receiving end is

$$E_{Rx}(l) = E_{R\text{-elec}}(l) = lE_{elec}, \qquad (17)$$

in which $E_{T\text{-elec}}$ and $E_{R\text{-elec}}$ are the energy consumption for transmitting and receiving 1-bit data, respectively, d_0 is the Euclidean distance threshold from the node to its communication neighbors, ε_{amp} stands for the energy consumption parameter for multipath loss, and ε_{fs} is the energy consumption for the node in the free space. When the node performs transmission, the path loss indices are 2 and 4, respectively. Related simulation parameters are elaborated in Table 1.

We make some comparisons, under different algorithm operating time, between the demand coverage probability and the calculated coverage probability. The impact of the dynamic parameter λ is also investigated while the simulation area has the size $100 * 100 \text{ m}^2$ and the results are obtained by averaging 300 repeated simulations. The simulations are depicted in Figures 1 and 2.

Figures 2 and 3 show the relationship between the demand coverage probability and the network coverage probability for different λ and different time. Observe from Figures 2 and 3 that the network coverage probability gets improved with the increasing λ. The reason is that when
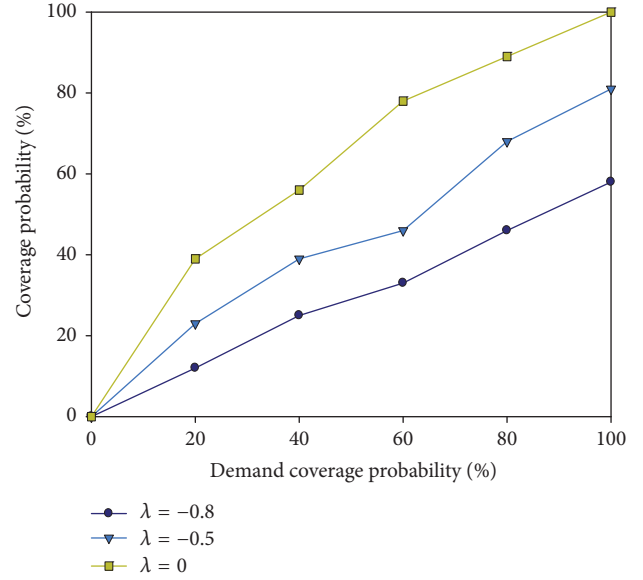

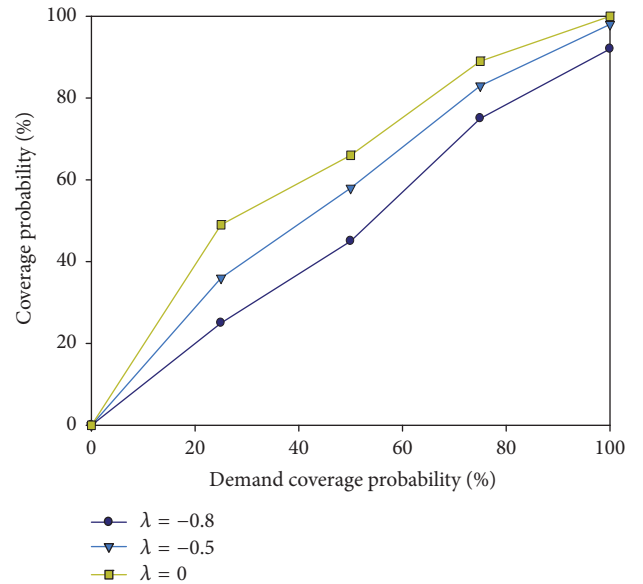FIGURE 1: The demand coverage probability versus the calculated results when $t = 100$ s.

FIGURE 2: The demand coverage probability versus the calculated coverage probability when $t = 200$ s.

λ increases, the effective fusion coverage intensity gets strengthened. When $\lambda = 0$, the effective fusion coverage intensity is 1 which is the maximal value. The network coverage probability increases with time in Figures 2 and 3. The reason is that there are more nodes in the sensor in the set M; that is, the effective fusion coverage intensity is increased for the target nodes in the monitoring area.

Then, we compare the CETFC algorithm with the ECTA algorithm [13] on network lifetime and operating time.

Figures 3 and 4 show the comparison between our proposed algorithm and the ECTA algorithm on network lifetime and operating time. It can be observed from Figure 3
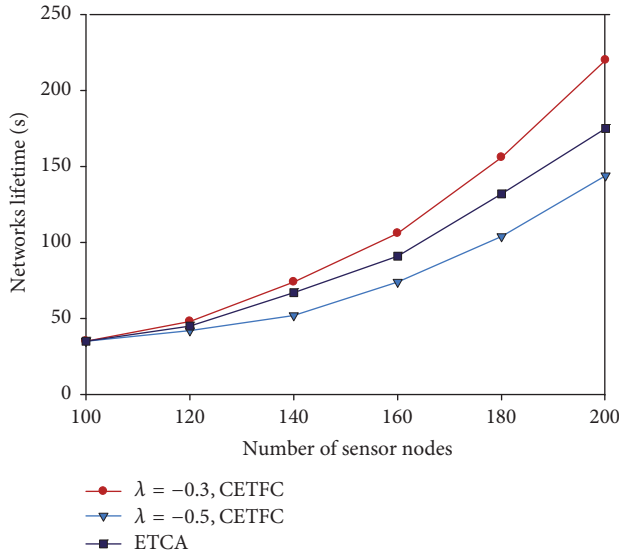
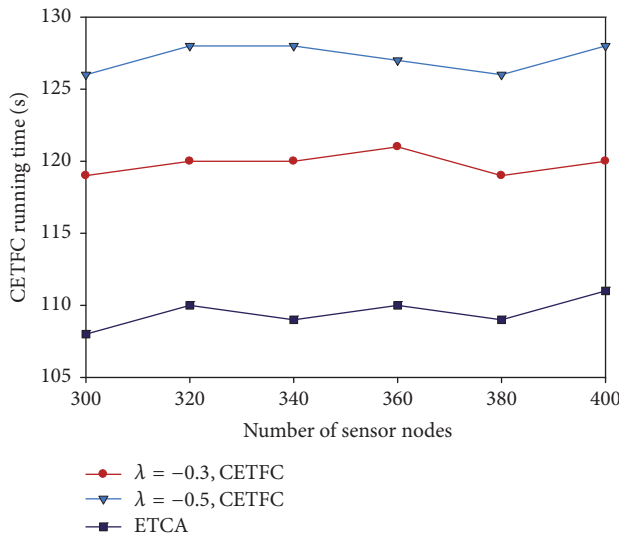Figure 3: Comparison of network lifetime in monitoring area 200 $\ast$ 200 m^2.

Figure 4: Comparison of operation time.

Figure 5: Comparison of network coverage probability for the three algorithms in simulation area 200 $\ast$ 200 m^2.

Figure 6: Comparison of network coverage probability for the three algorithms in simulation area 200 $\ast$ 200 m^2.

that, at the initial phase, these two algorithms exhibit similar network lifetime which can be further extended for both algorithms with a larger number of sensor nodes. However, since the ECTA algorithm achieves monitoring over the target nodes in a nonlinear and consistent way, more energy consumption is required by the ECTA algorithm. When there are more than 180 sensor nodes, the network lifetime of these two algorithms tends to be stabilized. The average system lifetime of the CETFC algorithm is higher than that of the ECTA algorithm by 12.78%. The algorithm operating time is presented in Figure 4 as a function of sensor node number. The ECTA algorithm employs a link structure for the storage of node energy. The node energy is ranked by an ergodic algorithm and the node with more energy is given higher priority to achieve the coverage over the target node. As a
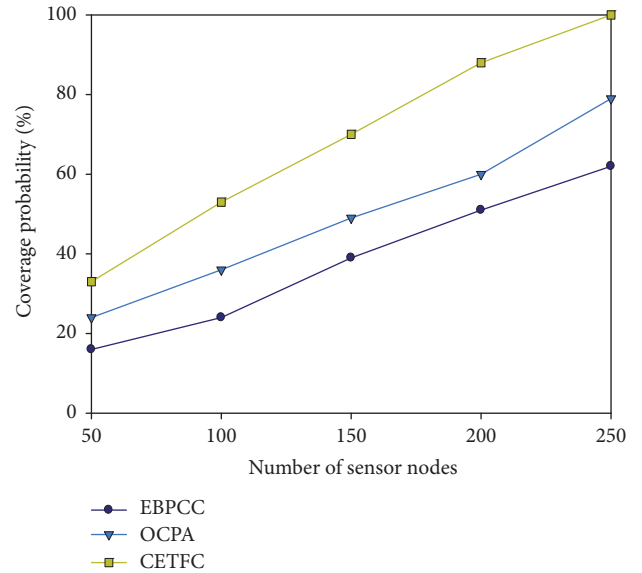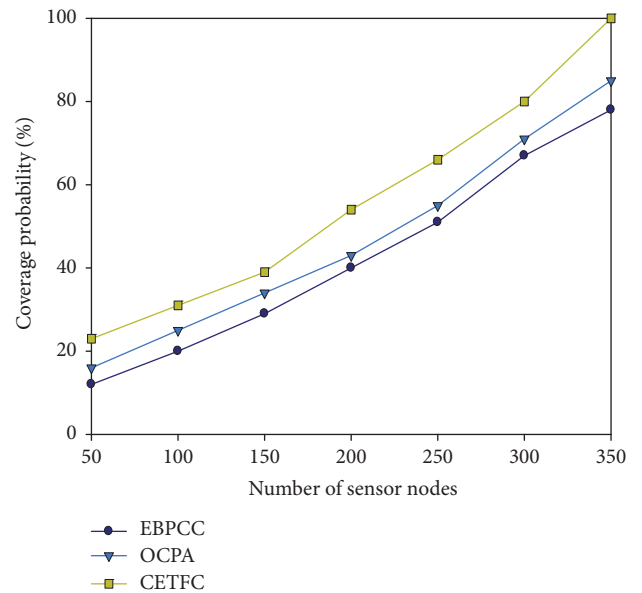
result, the ECTA algorithm possesses lower complexity and requires less operating time in comparison with our proposed algorithm.

We compare our proposed algorithm with the algorithms in [2, 9] in terms of network coverage probability and network lifetime. The simulation areas are 200 $\ast$ 200 m^2 and 300 $\ast$ 300 m^2, respectively, and the operating time is $t = 200$ s. The results are obtained by averaging 300 simulations. Simulation results are illustrated in Figures 5–7.

It is clearly evident from Figures 5 and 6 that the coverage probability gets improved for all of the three algorithms when there are more sensor nodes. However, our proposed
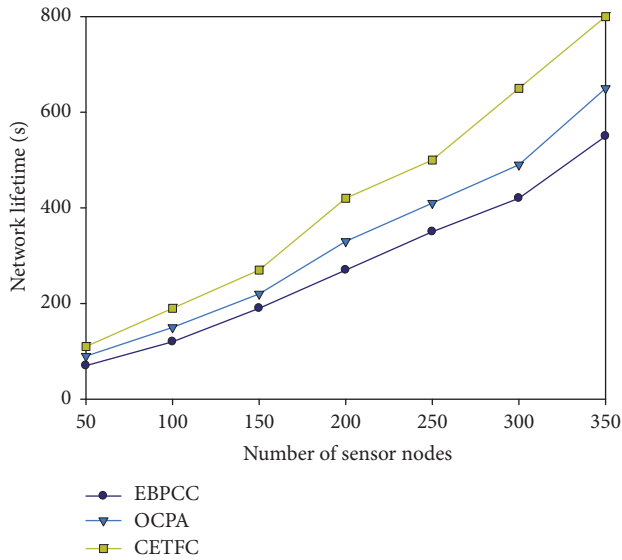
Figure 7: Comparison of network lifetime for the three algorithms in simulation area $200 * 200\,\mathrm{m}^2$.

algorithm shows a higher improvement speed than the others. This is due to the fact that the CETFC algorithm adopts the concept of fusion coverage intensity where the coverage intensity of all the nodes in the set M is taken into consideration. As a result, the effective fusion coverage intensity from the entire set M to the target node gets further improved. When there are equal numbers of sensor nodes, the proposed algorithm shows higher network coverage probability than any of the other algorithms while the average improvement rate is 10.12%. Observe from Figure 7 that, with the same number of sensor nodes, the proposed algorithm shows longer network lifetime than the other two algorithms and the average improvement rate is 13.68%. The reason behind the improvement on the network lifetime is the same as that for the improvement on the network coverage probability.

5. Conclusions

For a better investigation on the coverage problem in WSNs, we employed the probability theory to investigate the network coverage quality and the network lifetime. We revealed the proportional relationship between the fusion coverage intensity and the membership function during the coverage over the target node. We also investigated the range of the dynamic parameter λ and the fusion coverage process over the target node set. Finally, we obtained the following conclusions.

(1) The energy of the sensor node decays exponentially and exhibits a reverse proportional relationship with the Euclidean distance.

(2) The restricted threshold is employed to indicate the fusion coverage intensity and provide an accurate calculation for the range of the fusion coverage intensity. As a result, the effective coverage process from the sensor nodes to the target node can be reflected effectively.

(3) By setting the effective threshold, the coverage intensity can by adjusted to optimize the network resource allocation, reduce the network cost, improve the network energy efficiency, and prolong the network lifetime.

(4) By adjusting the tunable parameters, the fusion coverage intensity can be increased effectively to improve the coverage quality over the monitoring area.

We focused on the probability model and proposed a controllable effective threshold based fusion coverage algorithm. The node energy decay model was further simplified for computational considerations. The CETFC algorithm shows certain reference value to subsequent study on the network coverage quality. Since the study has been performed theoretically so far, it requires future work to put this algorithm into practice and improve the network coverage quality in engineering applications.

Conflicts of Interest

The authors declare that they do not have any commercial or associative interests that represent conflicts of interest in connection with the work submitted.

Acknowledgments

References

[1] A. E. Zonouz, L. Xing, V. M. Vokkarane, and Y. Sun, "Hybrid wireless sensor networks: A reliability, cost and energy-aware approach," *IET Wireless Sensor Systems*, vol. 6, no. 2, pp. 42–48, 2016.

[2] R. V. Kulkarni and G. K. Venayagamoorthy, "Particle swarm optimization in wireless-sensor networks: a brief survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 2, pp. 262–267, 2011.

[3] Z. Sun, Y. Zhang, Y. Nie, W. Wei, J. Lloret, and H. Song, "CASMOC: a novel complex alliance strategy with multiobjective optimization of coverage in wireless sensor networks," *Wireless Networks*, 2016.

[4] J.-T. Meng, J.-R. Yuan, S.-Z. Feng, and Y.-J. Wei, "An energy efficient clustering scheme for data aggregation in wireless sensor networks," *Journal of Computer Science and Technology*, vol. 28, no. 3, pp. 564–573, 2013.

[5] H. Mahboubi, K. Moezzi, A. G. Aghdam, K. Sayrafian-Pour, and V. Marbukh, "Distributed deployment algorithms for improved coverage in a network of wireless mobile sensors," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 163–175, 2014.

[6] Y.-C. Tseng, P.-Y. Chen, and W.-T. Chen, "K-angle object coverage problem in a wireless sensor network," *IEEE Sensors Journal*, vol. 12, no. 12, pp. 3408–3416, 2012.

[7] N. Tamboli and M. Younis, "Coverage-aware connectivity restoration in mobile sensor networks," *Journal of Network and Computer Applications*, vol. 33, no. 4, pp. 363–374, 2010.

[8] C.-C. Hsu, M.-S. Kuo, S.-C. Wang, and C.-F. Chou, "Joint design of asynchronous sleep-wake scheduling and opportunistic routing in wireless sensor networks," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 63, no. 7, pp. 1840–1846, 2014.

[9] F. Lin, X. Z. Zhou, and W. H. Zeng, "Sparse online learning for collaborative filtering," *International Journal of Computers Communications & Control*, vol. 11, no. 2, 2016.

[10] M. R. Senouci, A. Mellouk, L. Oukhellou, and A. Aissani, "An evidence-based sensor coverage model," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1462–1465, 2012.

[11] F. Lin, W. Zeng, L. Yang, Y. Wang, S. Lin, and J. Zeng, "Cloud computing system risk estimation and service selection approach based on cloud focus theory," *Neural Computing and Applications*, vol. 28, no. 7, pp. 1863–1876, 2017.

[12] J.-H. Seok, J.-Y. Lee, W. Kim, and J.-J. Lee, "A bipopulation-based evolutionary algorithm for solving full area coverage problems," *IEEE Sensors Journal*, vol. 13, no. 12, pp. 4796–4807, 2013.

[13] X. F. Xing, G. J. Wang, and J. Li, "Polytype target coverage scheme for heterogeneous wireless sensor networks using linear programming," *Wireless Communications and Mobile Computing*, vol. 14, no. 14, pp. 1397–1408, 2014.

[14] K. Derr and M. Manic, "Wireless sensor network configuration-part II: adaptive coverage for decentralized algorithms," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1728–1738, 2013.

[15] F.-Z. Meng, H.-Z. Wang, and H. He, "Connected coverage protocol using cooperative sensing model for wireless sensor networks," *Acta Electronica Sinica*, vol. 39, no. 4, pp. 772–779, 2011.

[16] Z. Y. Sun, W. G. Wu, H. Z. Wang, H. Chen, and X. F. Xing, "A novel coverage algorithm based on event-probability-driven mechanism in wireless sensor network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, article 58, pp. 1–17, 2014.

[17] M. Wazid, A. K. Das, S. Kumari, and M. K. Khan, "Design of sinkhole node detection mechanism for hierarchical wireless sensor networks," *Security and Communication Networks*, vol. 9, no. 17, pp. 4596–4614, 2016.

[18] H. Barati, A. Movaghar, and A. M. Rahmani, "EACHP: Energy Aware Clustering Hierarchy Protocol for Large Scale Wireless Sensor Networks," *Wireless Personal Communications*, vol. 85, no. 3, pp. 765–789, 2015.

[19] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1960–1971, 2006.

[20] A. Vajdi, G. Zhang, Y. Wang, and T. Wang, "A New Self-Management Model for Large-Scale Event-Driven Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 16, no. 20, pp. 7537–7544, 2016.

[21] X. F. Xing, G. J. Wang, and J. Li, "Square region-based coverage and connectivity probability model in WSNs," in *Proceedings of the Mobile Ad-Hoc and Sensor Networks 2010 Sixth International Conference (MSN)*, pp. 79–84, 2011.

[22] Z. Sun, Y. Shu, X. Xing, W. Wei, H. Song, and W. Li, "LPOCS: A novel linear programming optimization coverage scheme in wireless sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 33, no. 1-4, pp. 173–197, 2016.

[23] C.-L. Yang and K.-W. Chin, "Novel algorithms for complete targets coverage in energy harvesting wireless sensor networks," *IEEE Communications Letters*, vol. 18, no. 1, pp. 118–121, 2014.