

RESEARCH

Open Access



A low redundancy data collection scheme to maximize lifetime using matrix completion technique

Jiawei Tan¹, Wei Liu², Mande Xie^{3*}, Houbing Song⁴, Anfeng Liu¹, Ming Zhao⁵ and Guoping Zhang⁶

Abstract

Sensor nodes equipped with various sensory devices can sense a wide range of information regarding human or things, thereby providing a foundation for Internet of Thing (IoT). Fast and energy-efficient data collection to the control center (CC) is of significance yet very challenging. To deal with this challenge, a low redundancy data collection (LRDC) scheme is proposed to reduce delay as well as energy consumption for monitoring network by using matrix completion technique. Due to the correlation of the location-dependent sensing data, some data without being collected can still be recovered by the matrix completion technology, thereby reducing the data amount for data collection and transmission, reducing the network energy consumption, and accelerating the process of data acquisition. Based on matrix completion technique, LRDC scheme can select only part of the nodes to sense data and transmit less data to CC. By doing so, the data collected by the network can be greatly reduced, which can effectively improve the network lifetime. In addition, LRDC scheme also proposes a method for quickly compensate sample data in cases of packet loss, whereby part of redundant data is sent in advance to the area closer to CC. If the data required for matrix completion is lost, these redundant data can be quickly obtained by CC, so the LRDC scheme has low delay characteristics. Simulation results demonstrate that LRDC scheme can achieve better performance than the traditional strategy, and it can reduce the maximum energy consumption of the network by 27.6–57.9% and reduce the delay by 0.7–17.9%.

Keywords: Wireless sensor networks, Matrix completion technique, Data recovery, Energy efficiency, Delay

1 Introduction

In the past several years, we have witnessed a dramatic advancement in network-enabled sensors/actuators, which have been closely related to our lives, including medical monitoring [1–3], smart-home [1–5], smart-grid, smart vehicles [6–9], intelligent transportation systems, and smart cities [10–13]. This profound change is mainly due to the rapid development of the manufacturing process of intelligent electronic devices, which has led to an exponential increase in the number of these network-enabled sensor devices. It is estimated that almost 50 billion devices will be interconnected by 2020 [8, 14]. At the same time, these sensor devices are powerful and able to sense the information regarding

people or things, such as location, environment, and behaviors, thus providing a foundation for wireless sensor networks (WSN). On the other hand, these huge numbers of sensor devices can have huge data acquisition capabilities than before, and thus can form systems to facilitate deep interaction between human and objects, and many applications based on big data have been derived. The report from Cisco shows that the amount of data generated by the Internet of Things accounted for 69% of the total Internet traffic in 2014, 30 times the amount of data in 2000 [1, 14], and the growth of data is still in the state of acceleration. With the explosive growth of the number of sensor devices and the amount of data generated [15–17], extensive research has been conducted on the sensing and algorithms aspects [18–20], which can improve efficiency [21, 22], reduce cost [23, 24], and or make our life much more convenient

* Correspondence: xiemd@zjgsu.edu.cn

³School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

Full list of author information is available at the end of the article

[25, 26]. However, two key issues for WSNs are still not well resolved.

The first of important issue is how to reduce energy consumption and the amount of data to be acquired [27–32] for WSNs. The U.S. Environmental Protection Agency (EPA) report [6] pointed out that the United States' data centers (DCs) consumed about 61 billion kilowatt-hours of electricity in 2006, and its worth about \$4.5 billion. It was also observed that in 2007, the global energy consumption of 30 million worldwide servers was 100 TWh, costing \$9 billion, and it is expected to increase to 200 TWh in the next few years [6, 7]. The energy consumption in the report only includes the energy consumption of data centers. However, the number of sensor devices is much larger than the number of data centers, so their energy consumption for perception and communication is more than ten times the energy consumption of DCs for data calculation and storage data (relative to the energy consumption of perception and communication, the energy consumption of sensor devices computing and storage is negligible [6]). As one of important and basic operations for WSNs to increase its collection efficiency, it is critical for data collection to reduce the amount of data and its energy consumption.

The second key issue is delay [33–37] for data collection of WSNs. In many WSNs, the decision to make requires gathering sufficient data. It takes a lot of time to collect the data of the sensing interesting area, causing a large latency. Moreover, on the other hand, since sensor-based devices are used in carious environments, the characteristics of wireless communication cause the data packets to be easily lost. However, WSNs are very sensitive to delay, for example, in the health monitoring of the elderly, delayed decision will put off the best treatment time and lead to serious consequence. Therefore, in these applications, the system is required to be able to make decisions with partial data, that is, data collection and rapid decision making in the absence of partial data. On the other hand, tolerating partial data loss can also reduce the amount of data collected [38, 39] and reduce the amount of data that needs to be transferred, thus reducing network congestion and speeding up the transmission of data.

To deal with those challenges, in this paper, a low redundancy data collection (LRDC) scheme is proposed to reduce delay as well as energy consumption for WSNs by using matrix completion technique. The main contributions of this paper work are as follows:

1. In LRDC scheme, it does not collect data that can be recovered by the matrix completion technique, which can effectively reduce the amount of data that needs to be collected, and at the same time, the monitoring quality of the monitoring system will not decline. Due to the correlation of the location-dependent sensing data, some data without being acquired can be recovered through matrix completion techniques. Based on such principle, in LRDC scheme, compared with traditional data collection, only a subset of nodes is selected to sense data and route data to control center (CC), so the data collected by the network can be greatly reduced, which can effectively improve the network lifetime.
2. A method to quickly supplement the sample data is proposed to deal with some samples lost in the way to CC while maintaining a high lifetime. In loss and delay sensitive sensor-based networks, packets can be lost in random along the long way to CC. Thus, if only the minimum number of data packets are collected, due to packet loss, the data packets received by the CC can be less than the minimum amount for data recovery, causing failure for recovering the data. However, collecting more data packets may cause a large amount of redundancy and reducing network lifetime. Therefore, LRDC scheme proposes a novel solution for quickly supplementing sample data. The main idea of this method is when the process of collecting data, the amount of data collected must be greater than the amount of data required by the matrix completion technique to present for the loss of data in the route. The data that is more than the requirement of matrix completion technique is called the backup data set, while the data needed for matrix completion technique is called the basic data set. When the data of the basic data set is lost in the route and the CC does not receive it completely, the data of the backup data set can be used to supplement the data lost in the basic data set. The strategy of this paper is not the same as the previous strategy. Due to the amount of data forwarded by sensor nodes in the far CC area is small and there is residual energy, in LRDC scheme, the basic data set is directly routed to the CC, while the backup data set is routed to the nodes with a certain distance to CC. When the basic data is lost due to the unreliability of the network transmission, the supplementary data can be quickly acquired by CC, so LRDC scheme has low delay characteristics while maintaining a very high network lifetime.
3. The LRDC scheme can be applied to grid network as well as planar network. A large number of theoretical and experimental results show that the LRDC scheme can achieve better performance than the previous strategy, which can reduce the maximum energy consumption of the network by 27.6–57.9%, and reduce the delay by 0.7–17.9%.

The rest of this paper is organized as follows. Section 2 reviews the related work. The system model and problem formulation are presented in Section 3. In Section 4, we present the low redundancy data collection (LRDC) scheme in details. The theoretical analysis simulation results are presented in Section 5. Section 6 concludes this work.

2 Methods

In WSNs, the energy stored in the sensor node battery is limited, so reducing the network energy consumption is important. Therefore, a method of reducing network energy consumption and reducing the delay is proposed. Reducing the amount of data collected can effectively reduce the energy consumption of network. Because of the coherence between some data, sensor nodes collect a lot of redundant data. These data can be recovered by matrix completion technology, so only part of data need to be collected, thus reducing the amount data collected, and it also reduces the energy consumption of the network. However, reducing the amount of data collected has no effect on balancing network energy consumption, and data may be lost in transmission. Therefore, retransmission mechanism is needed to ensure the reliability of transmission, but the retransmission caused a lot of delay. The matrix completion technology needs to collect a part of the data to make the decision, so a lot of delay will reduce the efficiency of the network. Therefore, a method of quick supplement data is adopted. The network collects a large amount of data, some of which are stored in nodes on its transmission path. When data that needs to be transferred to CC is lost, data can be quickly transmitted from the nodes that store the data. Therefore, the matrix completion technique can quickly recover the matrix, which can make the decisions quickly, thus reducing the decision delay. The background and related work on this method can be found in Section 3.

3 Background and related work

In WSNs, there are two important challenges. One is how to reduce the energy consumption of the network. Closely related to energy consumption is how to reduce the amount of data [40, 41]. The other challenge is how to effectively reduce the time for data collection while ensuring the long lifetime of the network [42, 43]. The decision delay refers to the interval between the time for starting data collection and the time when the collected data is sufficient for decision making. The following first discusses the research work related to the first challenge. In the wireless sensor network, since each sensor node is powered by battery, when the energy of one of the nodes is used up, the entire network is paralyzed. Due to the energy consumption imbalance in the network,

energy of some nodes can be wasted. Therefore, there has been a lot of research on the reduction of energy consumption in WSNs. In ref. [44], each node transmits a data packet in one transmission period. Since the length of the data packet may vary, the modulation rate is adjusted to ensure that a data packet is transmitted in one cycle. When the time of a transmission cycle is fixed, changing the duty cycle also changes the working time of the node, which changes the modulation rate of node. Therefore, by changing the duty cycle, the author can minimize the energy of a node to transmit a data packet, thus reducing the network energy consumption (Table 1).

Reducing the energy consumption of a single node transmission can increase the network lifetime. However, due to the large amount of data forwarded by the nodes close to the CC, it also causes the problem of energy holes. A lot of research has been done on the issue of energy holes [45–48]. There are node deployment aspects [49, 50] to avoid this problem, Chen et al. [49] studied in linear network and grid network, where each node is deployed through unequal distances, so that the transmission distance between the nodes close to CC is smaller, and the transmission distance between the nodes far from the CC is larger. By so doing, the energy consumption of the network is balanced. In the work in [50], in order to avoid the problem of energy hole, a method of hierarchical deployment is proposed. This deployment method divides a large network area into small sub-areas, and each sub-area has some common sensor nodes and an assisting node. Compared with ordinary sensor node, the assisting node has more initial energy and larger transmission radius than ordinary nodes, and it only requires one hop or a few hops to transmit to the CC. Therefore, in a sub-area, ordinary sensor nodes transmit data to the assisting node, and then assisting node transmits the data to the CC. Since the initial energy of assisting node is more, the energy hole can be

Table 1 Network parameters

Parameter	Value	Value
T_{start}	The time of start-up	$5 \cdot 10^{-6}\text{s}$
η	The drain efficiency of the amplifier	0.35
k	The exponent order	3
G_1	The gain factor	10^3
M_1	The link margin	10^4
P_{circuit}	The power of circuit	140.7 mW
P_{syn}	The power of frequency synthesizer	50 mW
σ^2	The AWGN power spectral density	$3.981 \cdot 10^{-21}$
B	The channel bandwidth	10^4
L	The number of bits in a data packet	10^3
N_f	The receiver noise figure	10

relieved. There are also aspects from the transmission path to avoid this phenomenon. In [51], an algorithm called EA-BECHA is proposed to balance the load of the nodes. In the EA-BECHA algorithm, the current node always selects the next hop node which has the highest residual energy, so some high-load nodes are prevented from dying prematurely, which can relieve the problem of energy waste.

There are also some studies on how to reduce the amount of data collected. In the collected monitoring data, since the monitoring devices are densely distributed, there is some correlation between the collected data, i.e., they can be derived from each other. Therefore, data transmitted to the base station is redundant [52]. The redundant data is often synonymous with sparsity, and matrix completion technique is one of two typical sparse representation techniques. The matrix completion technique can recover the matrix completely as long as the amount of data in the matrix meets some requirements. In [53], the author not only gives the minimum amount of data needed for matrix completion in the traditional low-rank matrix but also proves that the amount of data required for matrix completion is actually related to the coherence of the matrix. The definition of coherence gives the number of data that an arbitrary matrix can use for matrix completion technique. Of course, if there is no data in one row or one column of the matrix, the recovered matrix has a large error. Therefore, there are some studies on the data distribution model. The work in [54] proves that when the amount of known data of the matrix meets the requirements, and the known data in matrix conforms to the Bernoulli distribution, the matrix can be considered as recoverable. Matrix completion technique has many applications in sensor networks. In [55], a distribution model called UTSCS is proposed to guarantee that there is sampled known data in each row and column, and ensure that each sensor is sampled at a time interval. Thus, the data sampling rate is greatly reduced, and it can also be guaranteed to be recovered by matrix completion technique. As can be seen from the above, matrix completion technique has been studied theoretically, and has also been applied to the sampling of sensor networks [38, 56].

There is not much work on reducing decision delay. Decision delay is essentially the time it takes for the entire network to make decisions. This is an important performance indicator for WSNs, although this is not fully studied in previous studies. In previous studies, it was often studied how long a single packet was routed from the source node to the CC [57, 58]. Due to the characteristics of wireless communication channels, to guarantee reliability of data transmission, it is usually necessary to adopt a mechanism to guarantee reliability,

resulting in increase in delay. The method commonly used to guarantee the reliability of data transmission is send-wait retransmission protocol [59]. In stop-wait protocol, sender waits for the receiver to return an ACK of the received data after sending the data. If the sender receives the ACK, it starts the transmission of the next data packet. Otherwise, if the sender waits longer than the predetermined threshold, it considers the data packet lost, and then starts to retransmit the data packet until the sender receives the ACK, or discard the packet when the number of retransmissions reaches the predetermined threshold. It can be seen that the delay of data transmission in WSNs is larger. The decision delay is the time that the network experiences when it collects packets that can make decisions. Thus, decision delay depends on the amount of data collected. If each grid in the network requires data collection, the data collection takes a long time. The strategy proposed in this paper only collects part of the grid data, which can effectively reduce the decision delay. Another important factor affecting the decision delay is the loss of data packets, which can increase the number of packets that need to be retransmitted. This paper proposes a method for quickly supplementing sample data, and the idea is to route the packets that may need to be retransmitted to areas not far from the CC. By so doing, we can quickly resend the data when data loss is needed, thereby effectively reducing the decision delay.

4 The system model and problem statement

4.1 The energy consumption model

The energy consumption model of this paper is similar to ref. [44], where a transmission cycle consists of three parts, including operating time T_{on} , standby time T_{stby} , and start-up time T_{start} . Similarly, P_{start} is the power for transient mode, and its mainly equal to the frequency synthesizer power. P_{stby} is the power for standby, which is considered to be null for simplification and P_{on} is the power of active mode. Therefore, the energy consumption for transmitting a packet is as follows:

$$\begin{aligned} E_{total} &= P_{on}T_{on} + P_{start}T_{start} \\ &= (P_{Tx} + P_{circuit} + P_{PA})T_{on} + P_{circuit}T_{start} \end{aligned} \quad (1)$$

where $P_{circuit}$ is the power of the circuit, P_{PA} is the power of amplifier. The power of amplifier can be expressed as:

$$P_{PA} = \beta P_{Tx} = \left(\frac{\xi}{\eta} - 1 \right) P_{Tx} \quad (2)$$

where η represents the drain efficiency of amplifier, and ξ is peak to average ratio that can be obtained from constellation size M : $\xi = 3 \left(\frac{\sqrt{M}-1}{\sqrt{M+1}} \right)$ for M-ary Quadrature Amplitude Modulation (MQAM).

Rearranged the above, the energy consumption of transmitting a packet using MQAM modulation technique is:

$$E_{\text{total}} = (1 + \beta)P_{Tx}T_{\text{on}} + P_{\text{circuit}}T_{\text{on}} + 2P_{\text{syn}}T_{\text{start}} \quad (3)$$

P_{Tx} represent the transmit power of the sensor node. According to k th path loss model [44], it can be calculated as:

$$P_{Tx} = P_{rx}G_d \quad (4)$$

$G_d = G_1d^kM_1$ represent the power gain of factor, G_1 is the gain factor of $1m$, M_1 is the link margin, and d is the sending radius of node. The exponent order k is between 2 and 4. In this study, $k = 3$ is selected.

P_{rx} is received signal power, the relationship between the received signal power, and signal-to-noise ratio (SNR) is:

$$P_{rx} = 2BN_f\sigma^2 \cdot \text{SNR} \quad (5)$$

According to ref. [44], in MQAM technique, the relationship between SNR and BER (bit error rate) is followed:

$$P_e \approx \frac{4}{b} \left(1 - \frac{1}{\sqrt{M}}\right) e^{\left(-\frac{3}{M-1}\right)\frac{\text{SNR}}{2}} \quad (6)$$

where b is modulation rate, and $M = 2^b$ is constellation size. In this study, $b = 3$ and $b = 4$ are selected. From this, the SNR can be calculated, and then the received signal power can be obtained. Then, the transmit power can be obtained, and finally the energy consumption of transmitting a packet can be obtained.

4.2 The matrix completion model

The composition of matrix in this paper is similar to that in ref. [55]. The row of the matrix represents that the data generated by the same sensor and received by sink, while the column of the matrix shows that the data generated by different sensors in the same cycles and received by the sink. As show in (7), $x_{1,1}, x_{1,2}, \dots, x_{1,n}$ represent the data generated by sensor numbered 1 at different cycles.

$$M = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n_2} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1,1} & x_{n_1,2} & \cdots & x_{n_1,n_2} \end{bmatrix} \quad (7)$$

Matrix completion is a technique for recovering the entire matrix from a submatrix of the matrix [53]. That is, for an unknown low-rank matrix $M \in R^{n_1 \times n_2}$, the rank of the matrix satisfies $r \ll \min\{n_1, n_2\}$, and only a subset of the matrix $M_{i,j}((i,j) \in \Omega)$ is needed to recover the unknown matrix. Known subsets Ω are randomly selected,

and sampling operation $P_{\Omega} : R^{n_1 \times n_2} \rightarrow R^{n_1 \times n_2}$ can be defined as:

$$[P_{\Omega}(X)]_{i,j} = \begin{cases} X_{i,j} & (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

If the set Ω has enough data, the matrix can be recovered by solving the following rank minimization problem [53].

$$\begin{cases} \min \text{rank}(X) \\ \text{s.t. } P_{\Omega}(X) = P_{\Omega}(M) \end{cases} \quad (9)$$

where $\text{rank}(\cdot)$ represents the rank of a matrix, and X is a random matrix.

However, since the rank minimization problem (9) is NP-hard [53], it is very difficult to solve. Therefore, ref. [53] also proved that most of the matrices M with rank r can be recovered well by solving the convex optimization problem:

$$\begin{cases} \min \|X\|_* \\ \text{s.t. } P_{\Omega}(X) = P_{\Omega}(M) \end{cases}$$

where $\|X\|_*$ represent the nuclear norm of the matrix X , and is equal to the sum of the singular values.

According to ref. [53], using convex optimization to recover an unknown matrix from a random matrix requires that the number of random matrix samples meet the following requirements:

$$m \geq Cn^{5/4}r \log n \quad (10)$$

where C is a constant, $n = \max\{n_1, n_2\}$, and the correct rate of recovery is $1 - cn^{-3} \log n$.

4.3 Problem statement

Therefore, the main purpose of this paper is to design an efficient strategy to increase the network lifetime, and the network lifetime is determined by the node that consumed the most energy in network [49]. Therefore, the goal can be converted to reduce the maximum energy consumption of the network as much as possible.

1. The number of data forwarded in a node affects its energy consumption. The more data forwarded by a given node, the greater its energy consumption. Therefore, in order to increase the network lifetime, it is necessary to reduce the maximum amount of data forwarded by the nodes in the network. The amount of data forwarded by the node is usually composed of two parts: the amount of data generated by itself and the amount of data generated by other nodes. Therefore, reducing the amount of data generated by itself and the amount of data sent by other nodes can reduce the maximum energy

consumption. The set of all nodes in the network is defined as $S = \{1, 2, \dots, N\}$, the network lifetime is denoted by l , and the amount of data forwarded by each node is D_i . We have

$$\max(l) = \min \max(D_i) \tag{11}$$

- In the transmission process, the retransmission mechanism is used to ensure the reliability of the transmission. However, the retransmission mechanism cannot ensure that the transmission of each packet is successful [44], and there will still be a small number of packets lost. For these missing data packets, the sink will broadcast to notify the node which is sending these packets to resend the packet with the reliability δ . Therefore, after the data is transmitted over the network, it also requires the transmission of supplementary data. Considering that the expected number of retransmissions of node i is ζ_i , when the node needs to transmit the amount of τ_i , the delay of nodes in the network is $\tau_i \zeta_i$. In the process of supplemental data transmission, the probability of data loss is $1 - \delta$. In this process, the number of data that the node needs to resend is $(1 - \delta)\tau_i$, and therefore the delay for

each node in the network is $\tau_i \zeta_i + (1 - \delta)\tau_i \zeta_i$. Thus, the maximum delay of the network is:

$$\min(\Theta) = \min_{i \in S} \max(\tau_i \zeta_i + (1 - \delta)\tau_i \zeta_i) \tag{12}$$

5 The design of LRDC scheme

5.1 Preliminaries

First, the optimization of the grid network is studied by matrix completion, and its network topology is shown in Fig. 1. This network can be widely used in precision agriculture, precision industry, personalized healthcare, and precision medicine, where smart sensor nodes deployed in these applications to detect various physical phenomena. When an event or physical phenomenon occurs, sensor nodes generate alert and alarm to achieve the goal of smart monitoring. For agricultural planting plants, factory production lines, hospital beds, etc., these monitored objects are all regular, so sensor nodes are deployed in an equidistant grid. Therefore, in this paper, we first study the grid network and then generalize it to the general sensor nodes randomly deployed planar network.

We consider a sensor network composed of N nodes, and each node generates a data packet to transmit to sink in a round of transmission, and a total of T rounds of data are transmitted. The sink will receive $N \times T$ packets, and these data packets can be represented using a matrix X ($X \in R^{N \times T}$).

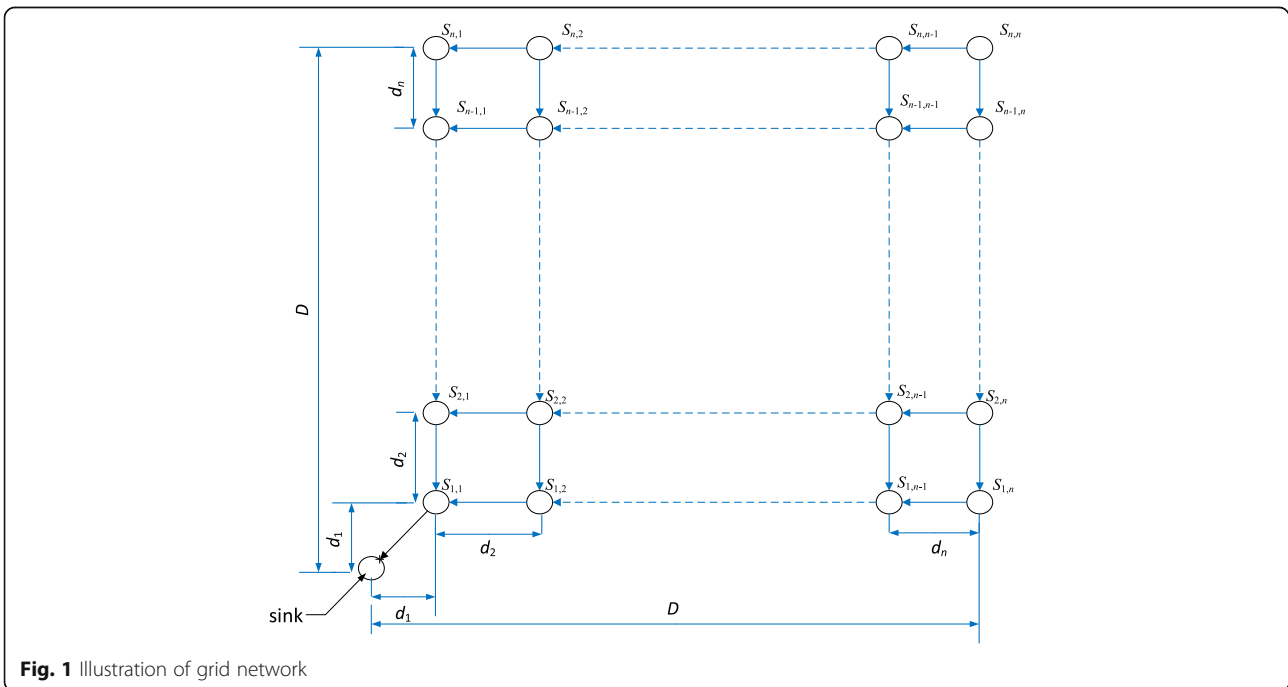


Fig. 1 Illustration of grid network

With LRDC, each sensor node will transmit T packets to the sink. The minimum amount of data required for the matrix completion technique is called the basic data set, while the excessive part of the data is called backup data set. Therefore, a matrix Q is defined as:

$$Q_{i,j} = \begin{cases} 0, X_{i,j} \in \text{the backup data set} \\ 1, X_{i,j} \in \text{the basic data set} \end{cases}$$

where i, j represents the packet transmitted by the sensor node numbered i in the j th round.

So the data matrix finally received by sink can be given as:

$$B = Q \cdot * X$$

where $\cdot, *$ represent a scalar product of two matrices.

5.2 LRDC scheme in grid network

The main idea of the LRDC scheme is illustrated in Fig. 2. First, it is necessary to determine which data belongs to basic data set and which data belongs to backup data set before data collection. Then, the basic data transmitted to the sink and the backup data transmitted to its storage location close to the sink in case of retransmission. When the basic data expected to be collected is lost, the sink can send a signal to notify the nodes with the backup data to transmit supplementary data. Therefore, the number of data required for matrix completion technology is satisfied. Finally, the matrix completion technology is used to recover the data that is not transferred to the sink.

The backup data consists of redundant data and the basic data consists of non-redundant data. To determine the locations for storing redundant data, the energy consumption of bottleneck node is need to know. Therefore, in what follows, we will analyze the energy consumption of the bottleneck node.

Energy consumption of bottleneck node: In the LRDC strategy, the redundant packets do not pass through the

bottleneck node, so the energy consumption of bottleneck nodes is only related to non-redundant packets.

Therefore, the energy consumption of the bottleneck node can be calculated. According to Eqs. (3), (4), (5), and (6), considering that the transmission radius of each node is d , the energy consumption of a packet transmitted by MQAM modulation technique is:

$$E_b = \frac{4}{3}(1 + \beta)N_f\sigma^2(M-1) \ln\left(\frac{4\left(1-\frac{1}{\sqrt{M}}\right)}{bP_e}\right)BT_{on}G_d + (P_{circuit})T_{on} + 2P_{syn}T_{start} \tag{13}$$

where P_e is the BER, and the reliability of one hop is as Eq. (14).

$$\mu = (1-P_e)^L \tag{14}$$

The retransmission mechanism is used to guarantee the success rate of transmission, so the maximum number of retransmissions of a k -hop node is given in Theorem 1.

Theorem 1: To guarantee the probability of successful transmission to the destination node is at least δ , the maximum number of retransmissions that reach the destination node after k -hop is:

$$\zeta_{k-max} = \left\lceil \frac{\log(1-\delta)}{\log(1-\mu^k)} \right\rceil \tag{15}$$

Proof: The probability of a node retransmitting ζ_{k-max} times still fail is $(1-\mu^k)^{\zeta_{k-max}}$. Therefore, the probability of successful transmission within ζ_{k-max} times retransmission is $1-(1-\mu^k)^{\zeta_{k-max}}$. The probability of successful transmission is required more than δ , that is

$$1-(1-\mu^k)^{\zeta_{k-max}} > \delta$$

Then, we can have

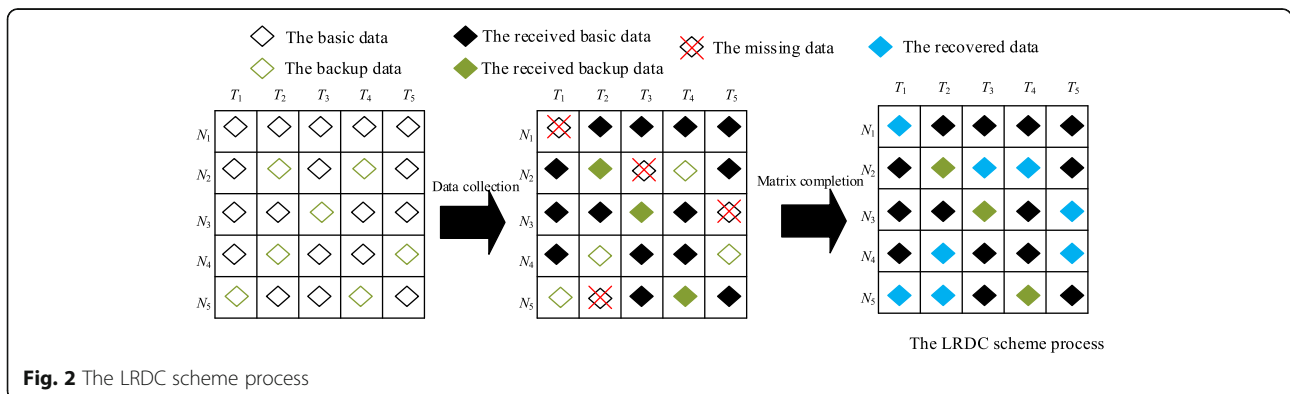


Fig. 2 The LRDC scheme process

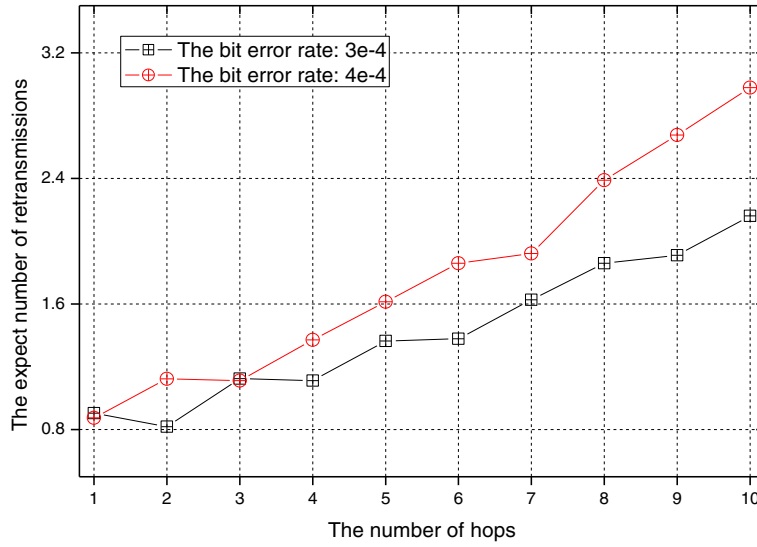


Fig. 3 The expected number of retransmissions per node

$$\zeta_{k-\max} > \frac{\log(1-\delta)}{\log(1-\mu^k)}$$

Round up the right side of the formula, we can get the formula (15).

Theorem 1 gives the maximum number of retransmissions when the successful transmission rate is at least δ . Therefore, the number of expected retransmissions that reach the destination node via k -hop can be calculated:

$$\zeta_k = \sum_{i=1}^{\zeta_{k-\max}} i \cdot \mu^k \cdot (1-\mu^k)^{i-1} \tag{16}$$

Figure 3 shows the expectation of retransmissions times. It can be seen that the more hops required to be transferred to destination node, the more expected number of retransmissions. Therefore, nodes near the destination can have reduced number of retransmissions.

Since the number of retransmissions is different for each node, the number of retransmission times can be calculated into the number of packets that each node needs to send, such as:

$$m_{i,j} = x_{i,j} \zeta_k \tag{17}$$

where k is the number of hops from $S_{i,j}$ to sink, and $x_{i,j}$ is the number of packets that need to send at $S_{i,j}$.

Therefore, the number of packets forwarded by each node is shown in Theorem 2.

Theorem 2: For grid network, in T -round data collection, the amount of data forwarding for each node in the network can be calculated as follows:

$$\begin{cases} D_{a,b} = \sum_{i=a}^n \sum_{j=b}^n \frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}} m_{i,j} & 1 < a, b \leq n \\ D_{a,b} = \sum_{k=b}^n \left(m_{1,k} + \frac{1}{2} \sum_{i=2}^n \sum_{j=k}^n \frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}} m_{i,j} \right) & a = 1 \\ D_{a,b} = \sum_{k=a}^n \left(m_{k,1} + \frac{1}{2} \sum_{i=k}^n \sum_{j=2}^n \frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}} m_{i,j} \right) & b = 1 \end{cases} \tag{18}$$

where $m_{i,j}$ is the amount of data that $S_{i,j}$ needs to send in T -round data collection.

Proof: First, the node that is not in the first row or first column is analyzed, as shown in Fig. 1. $S_{a,b}$ is any node in grid network that is not in the first row or first column. Since the nodes in grid network only transmit data to the left or down, the node that contributes the data of $S_{a,b}$ can only be from the upper right of $S_{a,b}$.

The nodes with the same number of hops to $S_{a,b}$ belong to the same layer. For any node $S_{i,j}$, the number of hops to reach $S_{a,b}$ is $i+j-a-b$, because there are only two options for each hop in grid network. It can only be transmitted down or left, therefore, the node $S_{i,j}$ has $2^{i+j-a-b}$ transmission paths in total.

In $2^{i+j-a-b}$ transmission paths to $S_{a,b}$, it is easy to know that the packet only needs to be transmitted down $i-a$ times and to the left $j-b$ times. Therefore, the path that can reach $S_{a,b}$ has $C_{i+j-a-b}^{i-a}$.

Therefore, the probability that the data of $S_{i,j}$ can be transmitted to $S_{a,b}$ is $\frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}}$ and the number of data that $S_{a,b}$ forward $S_{i,j}$ is $\frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}} m_{i,j}$.

Then, since the nodes of the first row can only be transmitted in one direction, the amount of data forwarded by the node $S_{1,b}$ in first row is $\sum_{i=b}^n m_{1,i}$. Before

that, the number of data forwarding of all nodes except the first row and the first column is calculated. Therefore, considering the second row of nodes, there is $\frac{1}{2}$ probability that the packets will be transmitted to the first row, and the amount of data forwarded is $\sum_{k=b}^n m_{1,k} + \frac{1}{2} \sum_{i=2}^n \sum_{j=k}^n \frac{C_{i+j-a-b}^{i-a}}{2^{i+j-a-b}} m_{i,j}$. Then, Eq. (18) can be obtained.

Theorem 2 gives the amount of data forwarding of all nodes in Grid network, and combined with Eq. (13), the energy consumption of the bottleneck node can be obtained.

The position of the redundant packets: In the above, the energy consumption of the bottleneck node has been obtained. Therefore, the location of redundant packets for each node as in Theorem 3.

Theorem 3: For grid network, using the residual energy of nodes in the network to transmit redundant packets, the storage layer of redundant packets is as follows:

$$\left\{ \begin{array}{l} Y = i_1 + j_1 - 1 \\ s.t. D_{i_1, j_1} + \frac{C_{a+b-i_1-j_1}^{a-i_1}}{2^{a+b-i_1-j_1}} m_{a,b} < D_{\max} \\ D_{i_2, j_2} + \frac{C_{a+b-i_2-j_2}^{a-i_2}}{2^{a+b-i_2-j_2}} m_{a,b} > D_{\max} \mid i_2 + j_2 - 1 < 1 \\ i_1 + j_1 = i_2 + j_2 - 1 \end{array} \right. \quad (19)$$

where $m_{a,b}$ is the number of packets that $S_{a,b}$ needs to send.

Proof: It is easy to know in the grid network that the layer of the node is equal to $(i+j-1)$ (The node with the same number of hops are in the same layer). The closer the node is to sink, the lower layer which is located. Therefore, when we require that redundant data be transmitted to nodes as close as possible to the sink, it is required to transmit to the node with the smallest sum of node number.

When the number of non-redundant packets per node is known, according to Eq. (18), the data forwarding amount of each node can be calculated; thus, the maximum amount of data forwarded by the node in network is equal to D_{\max} . Therefore, it is only necessary to ensure that the amount of data caused by transmitting redundant data does not exceed D_{\max} and will not reduce the network lifetime.

For the node S_{i_1, j_1} in $(i_1 + j_1 - 1)^{\text{th}}$ layer, if the redundant data of $S_{a,b}$ can be transmitted to S_{i_1, j_1} , the following conditions will inevitably be satisfied

$$D_{i_1, j_1} + \frac{C_{a+b-i_1-j_1}^{a-i_1}}{2^{a+b-i_1-j_1}} m_{a,b} < D_{\max}$$

In order to ensure the transmit to the node as close to the sink as possible, it also needs to satisfy the following:

$$D_{i_2, j_2} + \frac{C_{a+b-i_2-j_2}^{a-i_2}}{2^{a+b-i_2-j_2}} m_{a,b} > D_{\max}$$

Or the next layer of the node has already reached the sink, that is:

$$i_2 + j_2 - 1 < 1$$

Reorganizing the above, we can get (19).

Theorem 3 gives the storage position of redundant data. Therefore, the supplementary data send position in the supplementary data stage also can be known, so the final energy consumption is obtained. However, to obtain the storage location of redundant data, it is necessary to obtain the energy consumption of the bottleneck node. The energy consumption of the bottleneck node is related to the distribution of non-redundant data, so the distribution of non-redundant data needs to be known.

The distribution of redundant data: Since the matrix completion technique cannot recover data matrices in the case of empty rows or empty columns, there exist certain requirements on the number of redundant packets on each sensor.

In ref. [54], the authors have proved that when the collected data obeys Bernoulli distribution, the matrix with missing data can be recovered using matrix completion techniques. We also use the Bernoulli distribution model to determine the number of redundant data packet per node. Theorem 4 gives the number of redundant packets per node under Bernoulli distribution.

Theorem 4: In Bernoulli distribution model, the network collects T rounds of data, and the expected number of non-redundant packets that each node needs to send is:

$$m_i = \frac{NT-m}{N} \quad (20)$$

where N is the number of nodes.

Proof: In Bernoulli distribution model, the probability that each packet is a redundant packet is the same, and in the transmission of T -round data, the total number of packets that need to be collected is NT . Therefore, the probability that a given packet is redundant is $\frac{1}{NT}$.

Thus, the number of redundant packets per node is $\frac{1}{NT} \cdot T \cdot m = \frac{m}{N}$.

In the Bernoulli model, the expected number of non-redundant packets for each node is

$$m_i = \frac{NT-m}{N}$$

Theorem 4 gives the number of non-redundant packets per node in the Bernoulli distribution model, which can also get the amount of data forwarded by the bottleneck node.

Since the number of retransmissions of nodes at each layer is different, the data each node sends has different contributions to the data amount of the bottleneck node. In addition, the distribution of redundant data is uniform under the Bernoulli distribution; therefore, a method of unbalanced distribution of redundant data can be used to reduce the data amount of the bottleneck node.

In order to satisfy the requirement for the matrix completion technique to recover the matrix, the probability that empty rows or empty columns are required to be less than the Bernoulli distribution. Then, the minimum amount of non-redundant data per node is given in Theorem 5.

Theorem 5: For a data matrix, when the matrix can be recovered, the minimum amount of data that each sensor node needs to collect is:

$$x > \frac{mT + m}{NT + m} \quad (21)$$

Proof: Define event F is an event that does not collect data packets for a column, and x is the packet that is collected in a sensor node (a row in the matrix).

In the Bernoulli distribution model, the probability that a column has not collected any packets is

$$P_B = \left(1 - \frac{m}{NT}\right)^N = \left(\frac{NT-m}{NT}\right)^N$$

When there are x packets in a row, the probability that a column has no data packets is

$$P = \left(1 - \frac{C_T^{x-1}}{C_T^x}\right)^N = \left(\frac{T-2x+1}{T-x+1}\right)^N$$

To make $P < P_B$, we have

$$\frac{T-2x+1}{T-x+1} < \frac{NT-m}{NT}$$

Therefore, we can get

$$x > \frac{mT + m}{NT + m}$$

Theorem 6: In T -round data collection, the expected value of each node's redundant data packets under the Bernoulli distribution model is less than the maximum number of redundant data packets that the matrix can recover.

$$T - \frac{mT + m}{NT + m} > \frac{NT - m}{N}$$

Proof: According to Theorem 5, under the condition of recovery by matrix completion technology, the minimum number of non-redundant data packets per node is $\frac{mT+m}{NT+m}$, so the number of redundant packets is

$$R_{\max} = T - \frac{mT + m}{NT + m}$$

According to Theorem 3, the number of redundant data packets per node in Bernoulli distribution model is

$$R_{Ber} = \frac{NT - m}{N}$$

$$\text{Let } f(T) = R_{\max} - R_{Ber} = \frac{(NT)^2 - mN - NT + m}{N(NT + m)}.$$

Obviously, $N(NT + m)$ must be greater than 0, so only need to prove $(NT)^2 - mN - NT + m$, and let $g(T) = (NT)^2 - mN - NT + m$.

The derivative of $g(T)$ is

$$g'(T) = 2N^2T - N = N(2NT - 1)$$

Therefore, $g(T)$ is a monotonically increasing function in $[1, +\infty)$, and it is obviously that $g(1)$ is greater than 0, so it can be concluded that $g(T)$ is greater than 0, so $R_{\max} > R_{Ber}$.

According to Theorem 5, each sensor node needs to successfully send at least x data packets to sink to ensure that the probability of empty rows or empty columns is less than the Bernoulli distribution. Theorem 6 gives proof that the minimum required data in each row for matrix completion is less than the data of each row on Bernoulli distribution model. Therefore, the unbalanced distribution of redundant data packets of nodes can also reduce the maximum energy consumption.

Therefore, a new scheme of unbalanced of redundant data (UORD) distribution is proposed, and in the UORD distribution, the number of redundant data packets per node is determined by the amount of data forwarded by each node, as given in Algorithm 1.

Figure 4 shows the number of non-redundant packets per node with the LRDC scheme. The number of nodes in the network is 100, and the sum of non-redundant packets at the same hop to the sink is counted. It can be clearly seen that with the Bernoulli distribution, the amount of non-redundant data reduced by each node is uniform. However, with the UORD distribution, the non-redundant data for the node with a small number of hops from the sink will not be decreased, but the

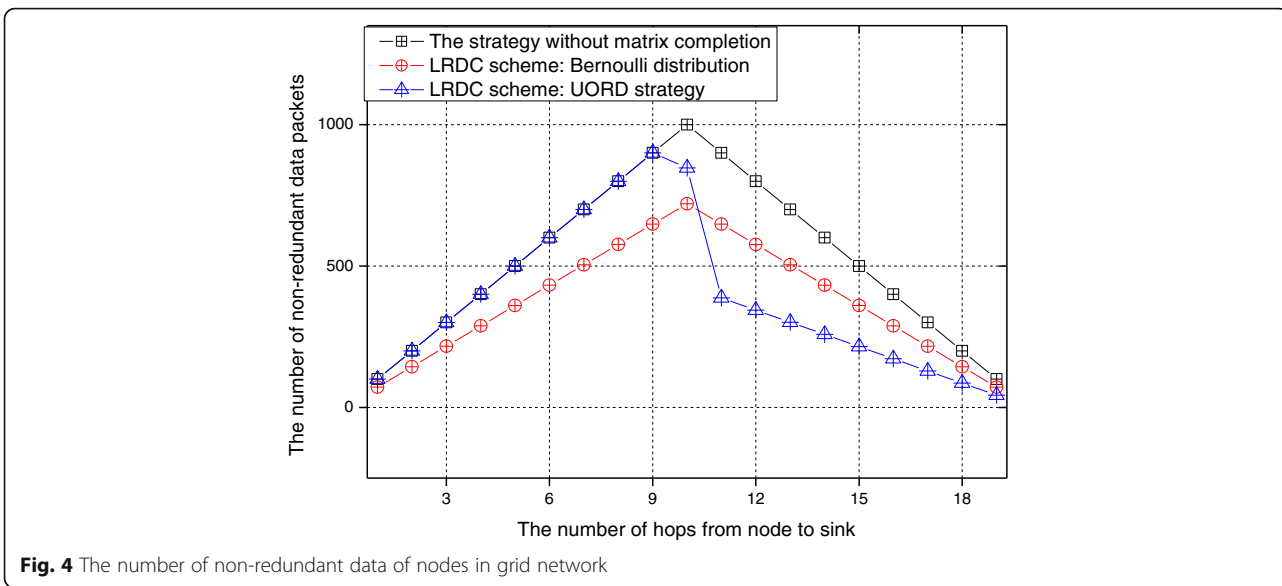


Fig. 4 The number of non-redundant data of nodes in grid network

non-redundant data for the node with a large number of hops from the sink will be greatly reduced. This is because in the grid network, there is only one node that has only one hop from the sink. In other words, the data packets of all the nodes will be forwarded through this node. Therefore, the data load of this node is the largest, and thus, it is better to reduce the non-redundant data of the nodes with more hops because the number of retransmission times of a data packet sent by the nodes is large.

Figure 5 shows the storage location of redundant data. It can be seen that if matrix completion technique is not used, supplementary data only can be sent from itself. In

the network with LRDC scheme, redundant data can be transmitted to the node with only two hops from sink when the redundant data meets the Bernoulli distribution. In the UORD distribution, redundant data packets are also transmitted to nodes that are only two hops from the sink. But, since the node with a small hop does not have redundant data packets, these nodes only can send supplementary data from itself.

Since the success rate of the transmission does not reach 100% when the non-redundant data is transmitted, there is still some data loss. Therefore, sink broadcast informs the corresponding node to send

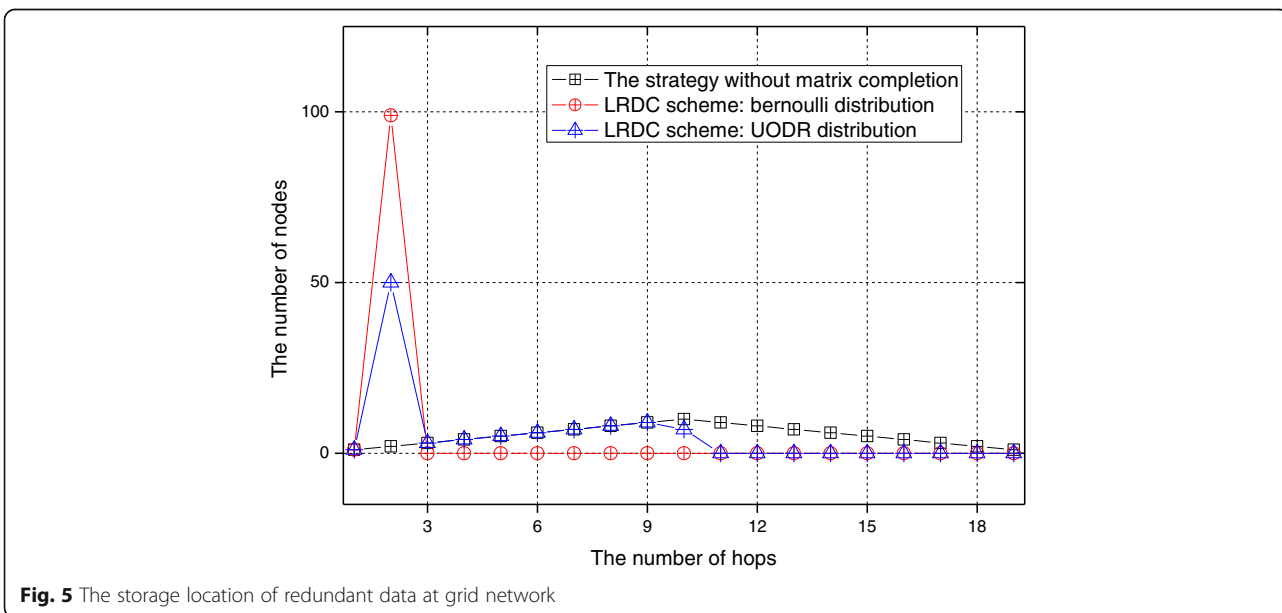


Fig. 5 The storage location of redundant data at grid network

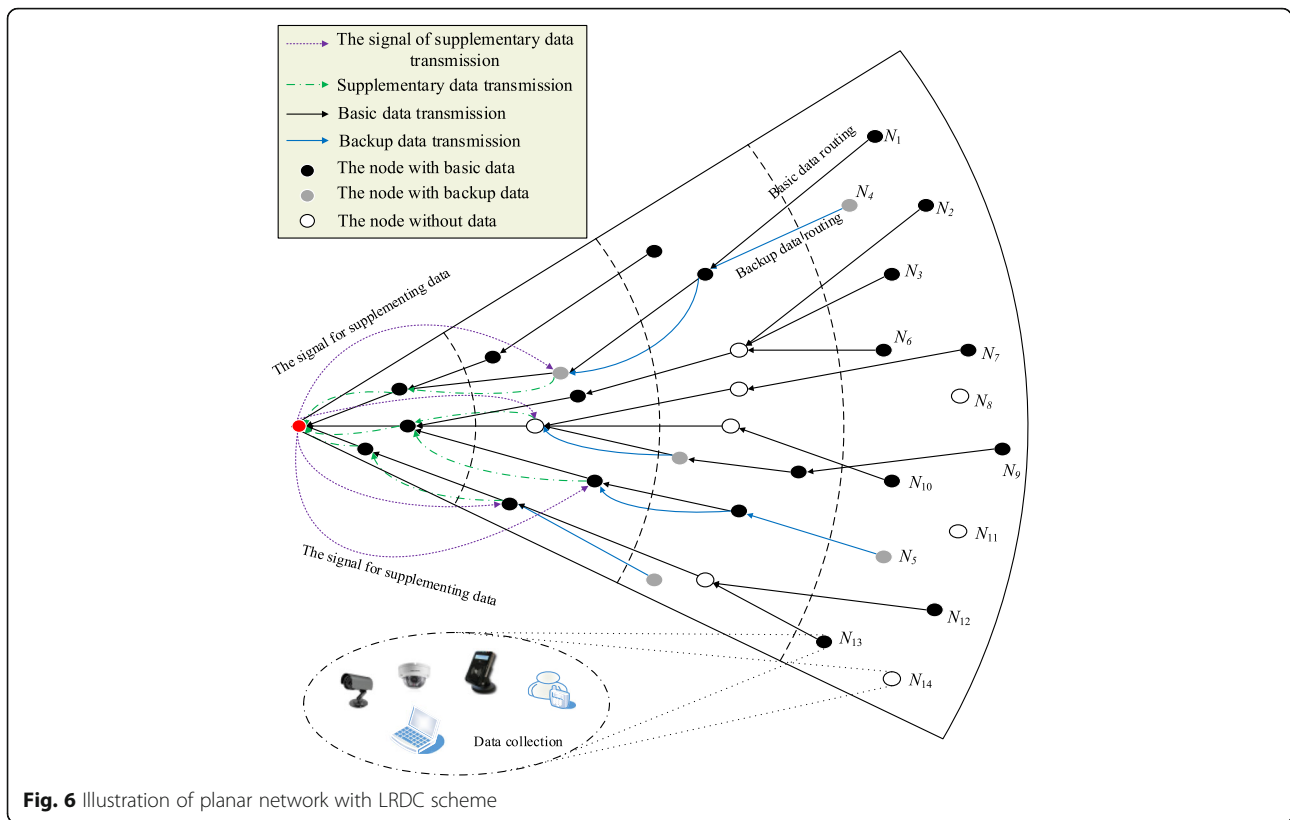


Fig. 6 Illustration of planar network with LRDC scheme

data again. However, In LRDC scheme, we can directly transmit the redundant data stored in the nodes near to sink as supplementary data, so the energy consumption of each node under the LRDC scheme is as theorem 7.

Theorem 7: For grid network under the LRDC scheme, when the network transmits T -round of data, the energy consumption of each node is:

$$E_{i,j} = \begin{cases} E_b(D_{i,j}^{non} + D_{i,j}^r + D_{i,j}^s) & i + j - 1 \geq Y \\ E_b(D_{i,j}^{non} + D_{i,j}^s) & i + j - 1 < Y \end{cases} \quad (22)$$

where Y is the storage location of $S_{i,j}$ redundant data.

Proof: In the transmission of T -round of data, the energy consumption of the LRDC scheme may consist of three parts, including the energy consumption for transmitting non-redundant data packets, the energy consumption for transmitting redundant data packets, and the energy consumption for supplementary data transmission.

It is easy to know that the number of non-redundant packets need to send is $x_{i,j}$. According to formula (17), adopting retransmission mechanism, the number of data packets that each node needs to send is $m_{i,j}^{non}$, so from

Theorem 1, the amount of data forwarded by each node is $D_{i,j}^{non}$.

The transmission of redundant data packets also adopts the retransmission mechanism. Therefore, the number of redundant data packets that each node needs to send is $m_{i,j}^r$. However, the transmission of redundant data packets cannot pass through the bottleneck node. Thus, for node with $i + j - 1 \geq Y$, Theorem 1 can be used to obtain the amount of data forwarded by each node $D_{i,j}^r$, and for the remaining nodes, the amount of data forwarded is 0.

For supplementary data transmission, because the retransmission mechanism only guarantees the probability of successful transmitted is δ , so the probability of loss when transmitting non-redundant packets is $(1 - \delta)$; therefore, the number of supplementary data packets of nodes need to send is $\lceil (1 - \delta)x_{i,j} \rceil$. The node with redundant data send supplementary data from nodes that store redundant data, and the node without redundant data send supplementary data from itself. Considering node $S_{i',j'}$ has stored redundant data for node $S_{i,j}$, so the number of packets that node needs to send is $m_{i',j'}^s$, then according to Theorem 1 can get the amount of data forwarded by each node is $D_{i,j}^s$.

Reorganizing the above, we can get the energy consumption of the network as below

$$E_{i,j} = \begin{cases} E_b \left(D_{i,j}^{non} + D_{i,j}^r + D_{i,j}^s \right) & i + j - 1 \geq Y \\ E_b \left(D_{i,j}^{non} + D_{i,j}^s \right) & i + j - 1 < Y \end{cases}$$

Theorem 7 gives the energy consumption of each node in grid network. Therefore, the maximum energy consumption also can be obtained.

Because in the LRDC scheme, redundant data is transmitted to nodes near the sink, so the number of hops to transmit supplementary data is greatly reduced, thereby reducing the delay of supplementary data transmission. Under this condition, the delay of each node in the grid network is calculated as Theorem 8.

Theorem 8: For grid network, with the LRDC scheme, the delay for each node is:

$$\phi_{i,j} = m_{i,j} \zeta_{k_1} + (T - m_{i,j}) \zeta_{k_2} + \lceil (1 - \delta) m_{i,j} \rceil \zeta_{k_3} \quad (23)$$

Proof: In grid network, considering the delay of $S_{i,j}$, the delay also can be divided into three parts, including the delay caused by transmitting non-redundant data packets, the delay caused by transmitting redundant data packets, and the delay caused by supplementary data transmission.

For the delay caused by the transmission of non-redundant data packets, it is considered that $S_{i,j}$ transmitted to the sink through k_1 hops. Therefore, under the retransmission mechanism, the delay for transmitting this part of data is $m_{i,j} \zeta_{k_1}$.

When transmitting redundant data, it only need to ensure that redundant data is transmitted to the node where it is stored, and the storage location can be obtained by Theorem 3. Therefore, the number of hops from $S_{i,j}$ to the storage location, which is considered k_2 , so the delay of this part is $(T - m_{i,j}) \zeta_{k_2}$.

Finally, in the LRDC scheme, the supplemental transmission process requires the transmission of $\lceil (1 - \delta) m_{i,j} \rceil$ supplementary data packets, and transmission can start from the node that stores the redundant data packet. Considering that the number of hops from node for storing redundant packets to sink is k_3 , the delay is $\lceil (1 - \delta) m_{i,j} \rceil \zeta_{k_3}$.

Then, we can have

$$\phi_{i,j} = m_{i,j} \zeta_{k_1} + (T - m_{i,j}) \zeta_{k_2} + \lceil (1 - \delta) m_{i,j} \rceil \zeta_{k_3}$$

Algorithm 1: The UORD distribution for maximizing lifetime in Grid network

Initialize: Considering the set of all nodes N , the length of row in grid network n

Input: the number of data packets that each node needs to transmit $x_{i,j}$.

Output: the number of data packets forwarded by each node $D_{i,j} \mid (i,j) \in N$.

1: the number of redundant packets m can be calculated by Eq. (10).

2: the amount of data $D_{i,j}$ forwarded by each node is obtained by Eq. (18).

3: **For** $k=1$ to m

4: Defined a value (< 0) α and a number (i', j') .

5: **For each node** (i, j) in N

6: **For** $ki=1$ to i and $kj=1$ to j

7: **If** $i==1$ or $j==1$

8: $Q_{ki,kj}^{i,j} = x_{i,j}$

9: **else**

10: $Q_{ki,kj}^{i,j} = \frac{c_{i+j-ki-kj}^{i-ki}}{2^{i+j-ki-kj}} x_{i,j}$

11: **End if**

// to get the contribution with the amount of data of node (i, j) to node (ki, kj)

12: $D'_{ki,kj} = D_{ki,kj}$, $D'_{ki,kj} = D'_{ki,kj} - Q_{ki,kj}^{i,j}$

13: $D_{max} = \max(D'_{ki,kj})$

14: **End for**

15: **If** $D_{max} < \alpha$

16: $\alpha = E_{max}$, $(i', j') = (i, j)$.

17: **End if**

18: **End for**

19: The number of redundant packets $R_{i',j'} = 1$

20: **For** $ki=1$ to i' and $kj=1$ to j'

21: **If** $i'==1$ or $j'==1$

22: $Q_{ki,kj}^{i',j'} = x_{i',j'}$.

23: **else**

24: $Q_{ki,kj}^{i',j'} = \frac{c_{i'+j'-i'-j'}^{i'-i'}}$

25: **End if**

26: $D_{ki,kj} = D_{ki,kj} - Q_{ki,kj}^{i',j'}$

27: **End for**

28: **End for**

5.3 LRDC scheme in grid network

In the following, a more common planar random network model [16] will be studied, as shown in Fig. 6.

The sensor nodes in the network are randomly and evenly deployed in a circle with radius R , and the transmission radius of each node is considered to be d . From Fig. 6, it can be seen that the backup data will only be transmitted to its storage location and will not pass through the bottleneck node. Only when the sink sends a signal to transmit supplementary data, a small amount of backup data will be transmitted to the sink as supplementary data.

Similarly, the energy consumption of the bottleneck node in LRDC scheme needs to be analyzed. By using the shortest path method [16] to transmit data, at a

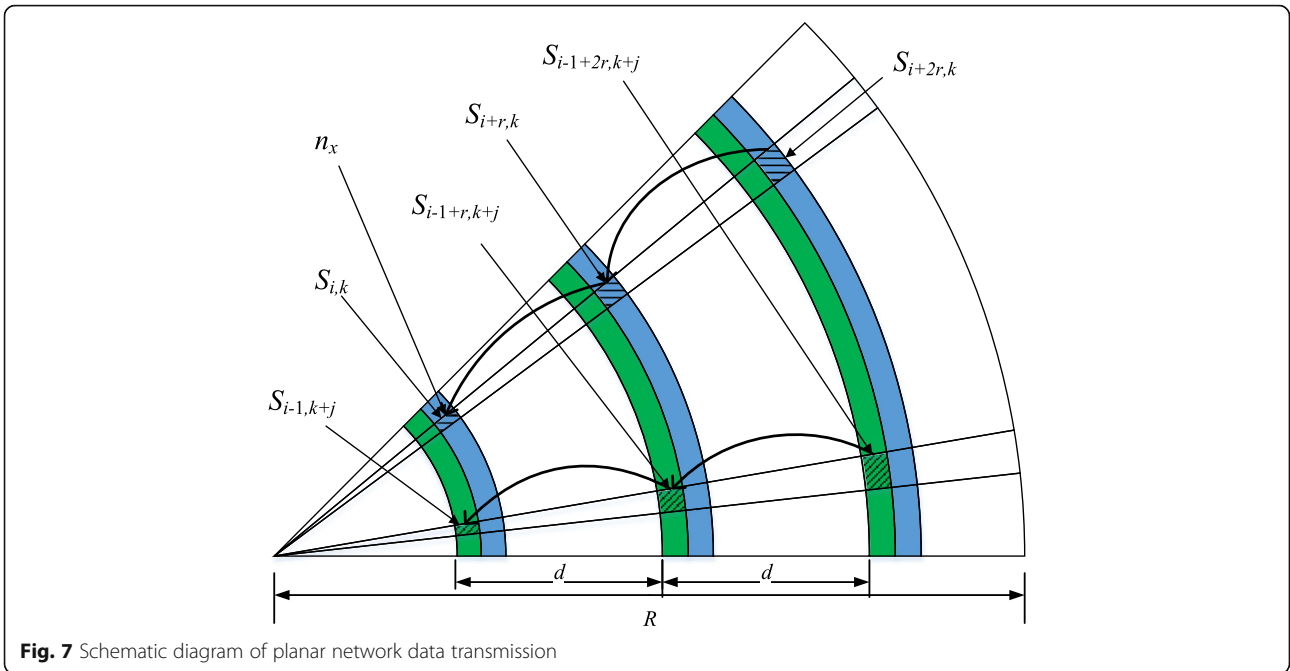


Fig. 7 Schematic diagram of planar network data transmission

distance of x meters from the sink, the number of data forwarding is:

$$D_x = (z + 1) + \frac{z(1 + z)d}{2x} \tag{24}$$

where z is an integer that makes $x + zd$ just smaller than R .

Energy consumption of bottleneck node: According to Eq. (24), when each node sends a data packet, the number of forwarding data of each node of the network can

be calculated. However, since the number of data that each node needs to send may not be the same, the number of data that each node forwards needs to be recalculated as Corollary 1.

Corollary 1: In the T -round of data collection, the number of data forwarded by the node with the distance x from sink is:

$$D_x = \lambda_x + \frac{\lambda_{x+d}(x + d) + \dots + \lambda_{x+zd}(x + zd)}{x} \tag{25}$$

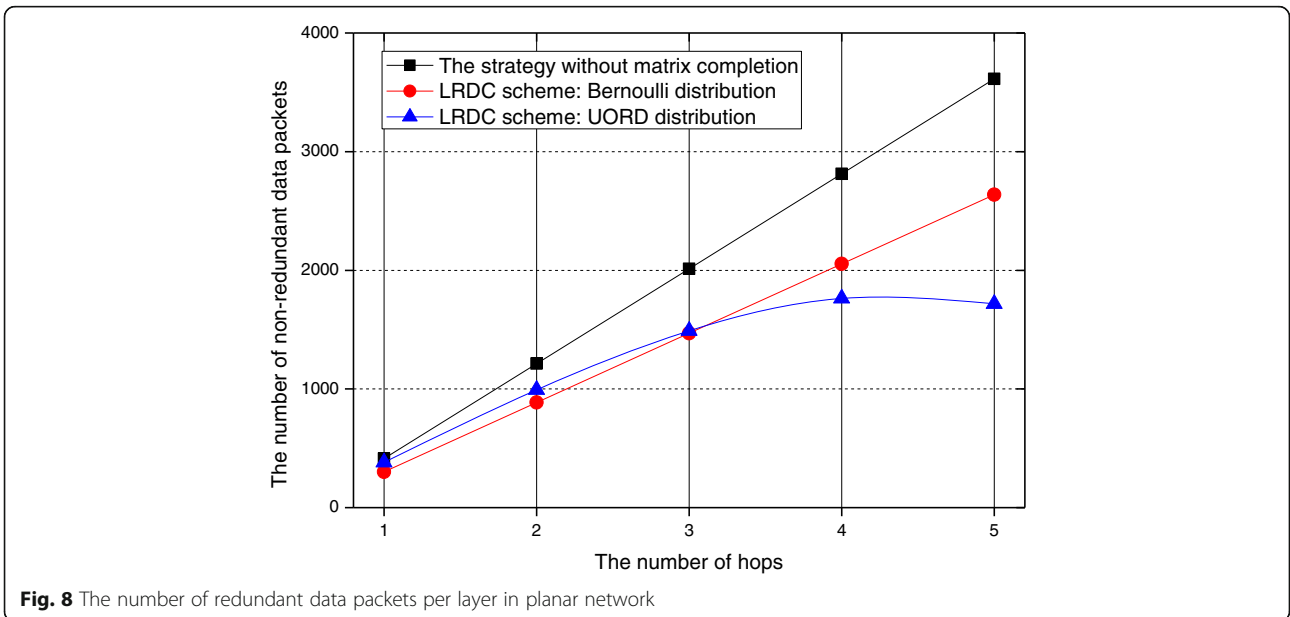


Fig. 8 The number of redundant data packets per layer in planar network

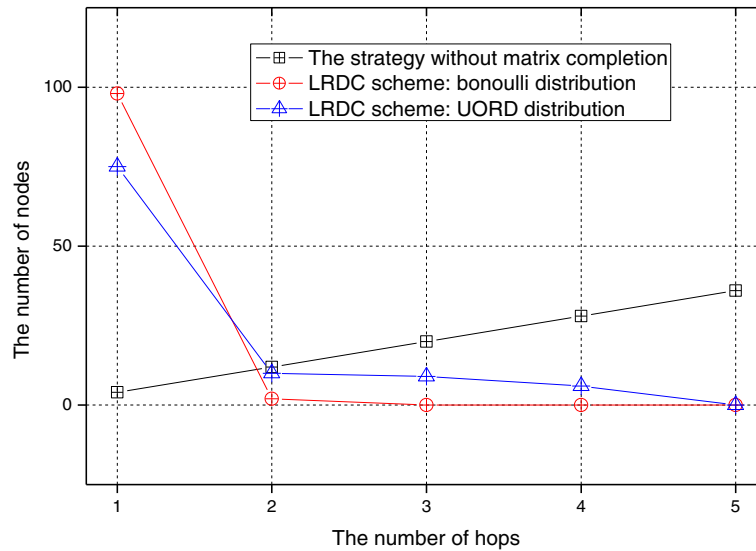


Fig. 9 The storage location of redundant data in planar network

where λ_{x+kd} represents the data amount of the node in the region between $x+kd$ and $x+(k+1)d$.

Proof: As shown in Fig. 7, considering the node n_x is located in $S_{i,k}$ and the width of the ring is d_x . When d_x is very small, the area where n_x is located can be approximated as a rectangle. At the same time, due to d_x is very small, the amount of data that the nodes forward in the same rectangle can be considered the same. Therefore, the area where n_x is located is

$$S_{i,k} = axd_x$$

And the number of nodes in this area is

$$N_x = S_{i,k}\rho = axd_x\rho$$

This area needs to transmit λ_x data in T -round data collection, so the amount of data in T -round data collection is

$$D_x = \lambda_x N_x$$

It can be seen in Fig. 7 that the area where n_x is located will forward the data in $S_{x+r,k}, S_{x+2r,k}, \dots, S_{x+zr,k}$ (z is the maximum value that satisfy $x+zr \leq R$), so the total amount of data forwarded is

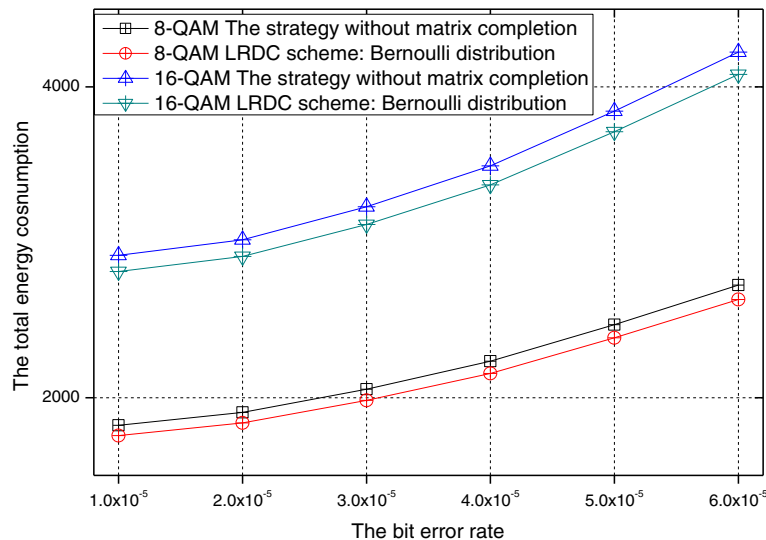


Fig. 10 The total energy consumption of the network under the Bernoulli distribution model in grid network

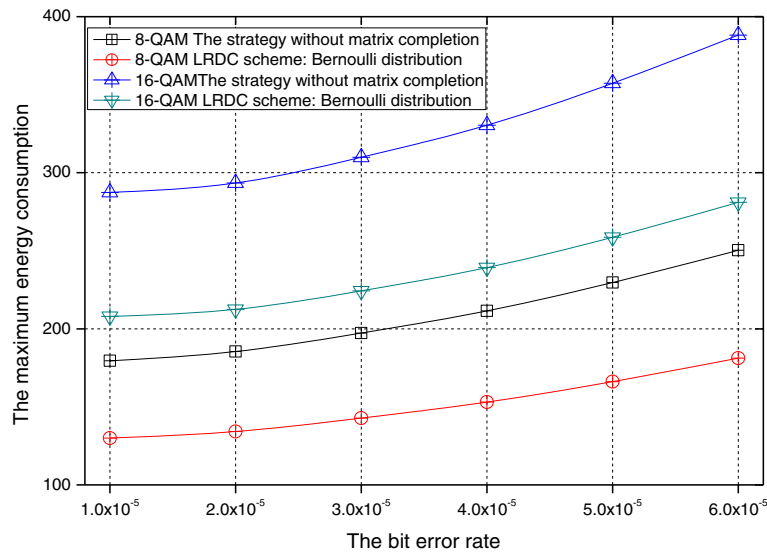


Fig. 11 The maximum energy consumption of the network under the Bernoulli distribution model

$$\lambda_x N_x + \lambda_{x+r} N_{x+r} + \dots + \lambda_{x+zr} N_{x+zr}$$

Therefore, the amount of forwarding data for each fan-shaped ring where n_x is located is

$$D_x = (\lambda_x N_x + \lambda_{x+r} N_{x+r} + \dots + \lambda_{x+zr} N_{x+zr}) / N_x$$

$$= \lambda_x + \frac{\lambda_{x+r}(x+r) + \dots + \lambda_{x+zr}(x+zr)}{x}$$

Corollary 1 gives the amount of data to be forwarded by each node when each node sends different amounts of data in the planar network. Therefore, the energy consumption of the bottleneck node in LRDC scheme can be obtained.

The position of the bottleneck node: Similar to the grid network, considering that nodes with the same hops are in the same layer, the position of redundant data storage also can be obtained.

Theorem 9: In the planar network, for a node with distance x from sink, it stores the redundant data in the Y th layer, and Y is calculated as follows:

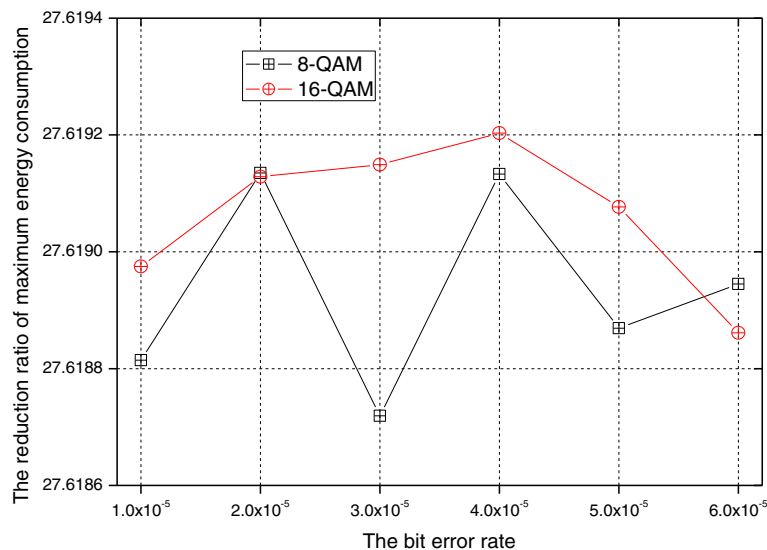


Fig. 12 The reduction ratio of maximum energy consumption under Bernoulli distribution

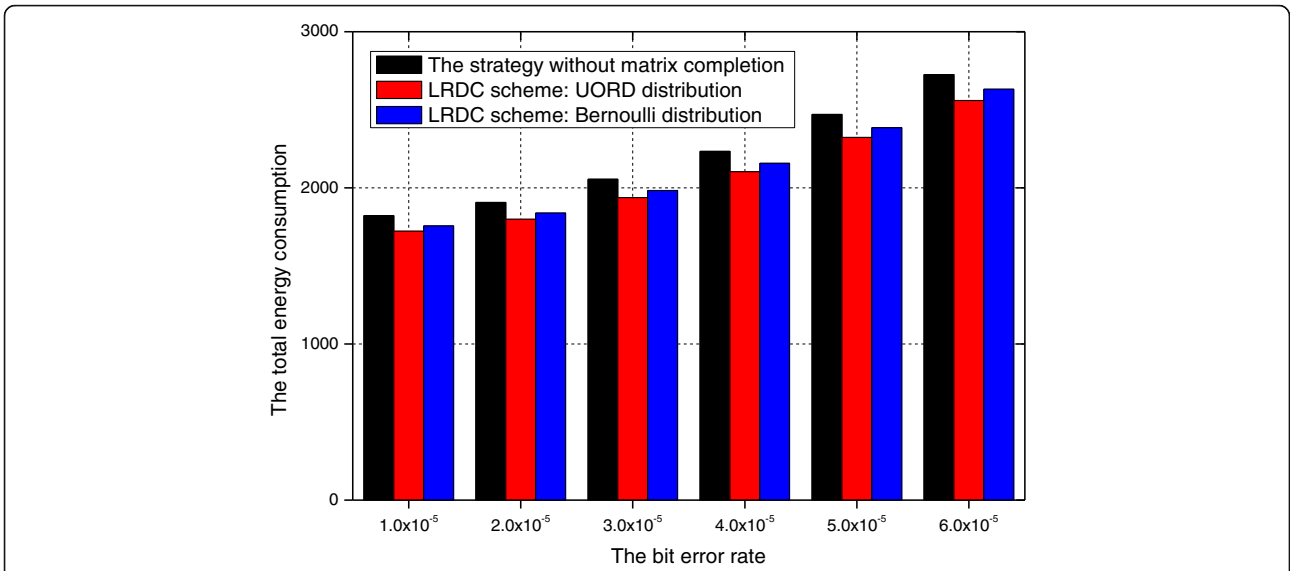


Fig. 13 The total energy consumption of networks under UORD distribution

$$\begin{cases} Y = \frac{x-zd-1}{d} + 1 \\ s.t. D_{x-zd} + \frac{\lambda_x x}{x-zd} < D_{max} \\ D_{x-(z+1)d} + \frac{\lambda_x x}{x-(z+1)d} > D_{max} \mid x-(z+1)d \leq 0 \end{cases} \quad (26)$$

Proof: When the amount of non-redundant data of each node in the network is known, the amount of data D_x forwarded by each node can be obtained by Corollary 1 so as to obtain the maximum amount of data D_{max} to be forwarded.

Considering that node A located at the distance of sink $(x - zd)$ meter, and node B located at the distance of sink x

meter. By Corollary 1, it can be obtained that the node A will forward the amount of data transmitted by node B is $\frac{\lambda_x x}{x-zd}$. Therefore, if the redundant data can be transmitted to the node where the sink distance is $(x - zd)$, must be satisfied

$$D_{x-zd} + \frac{\lambda_x x}{x-zd} < D_{max}$$

In addition, the redundant data required to be transmitted to the area closest to the sink. Therefore, to satisfy the above conditions, the following should hold:

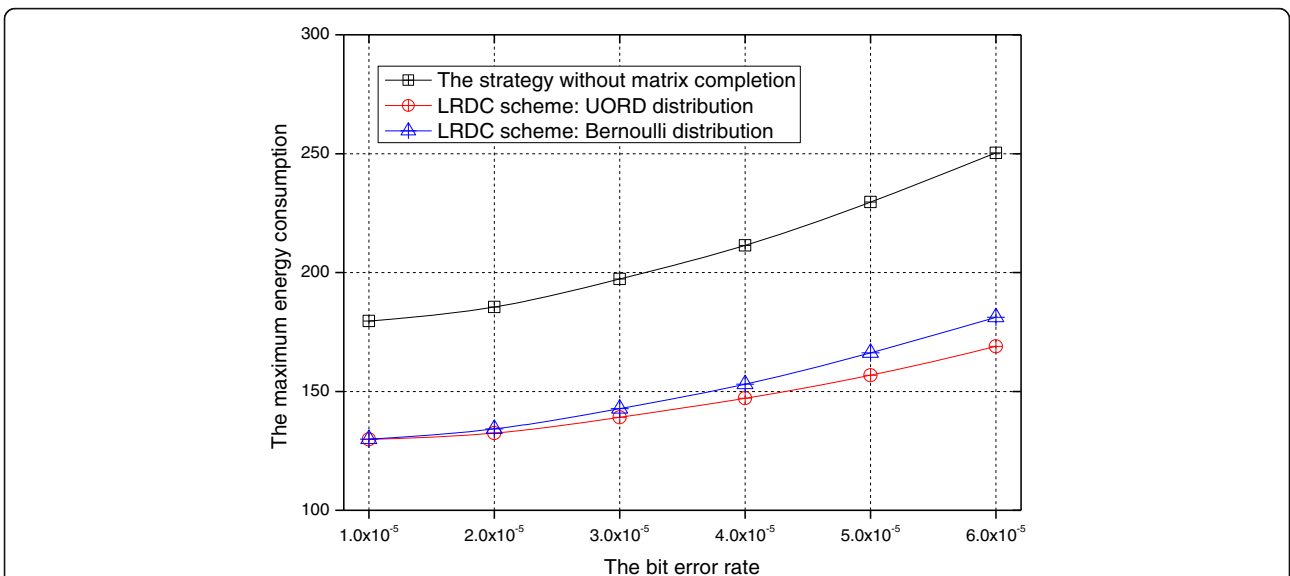


Fig. 14 The maximum energy consumption of networks under UORD distribution

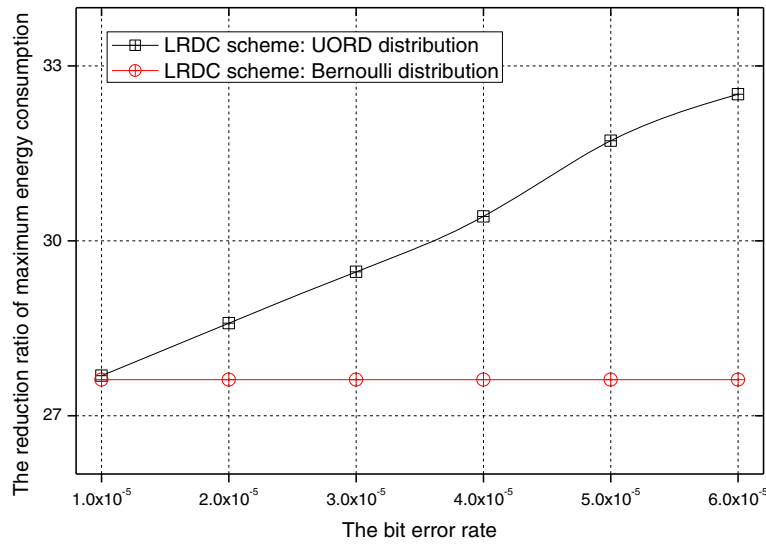


Fig. 15 The maximum energy consumption reduction ratio under UORD distribution

$$D_{x-(z+1)d} + \frac{\lambda_x x}{x-(z+1)d} > D_{\max}$$

Or the next hop is arrived at the sink, that is

$$x-(z+1)d \leq 0$$

Therefore, the nearest location to which redundant data packets can be transmitted is obtained, so it is possible to calculate how many hops are stored in the node that is away from sink.

Theorem 9 shows the storage location of redundant data. It also needs to know the distribution of redundant

data packets to get the storage position of redundant data.

In the planar network, similar to the grid network, according to Theorem 4, redundant data packets can be evenly distributed on each sensor node. Similarly, according to Theorem 5 and Theorem 6, the minimum required non-redundant data for each node can be known and it is less than that in the Bernoulli distribution. Therefore, under UORD distribution, the energy consumption of the bottleneck node is smaller. How to distribute the redundant data packets is shown in Algorithm 2.

Figure 8 shows the sum of the number of non-redundant packets of nodes with the same number of hops in a planar

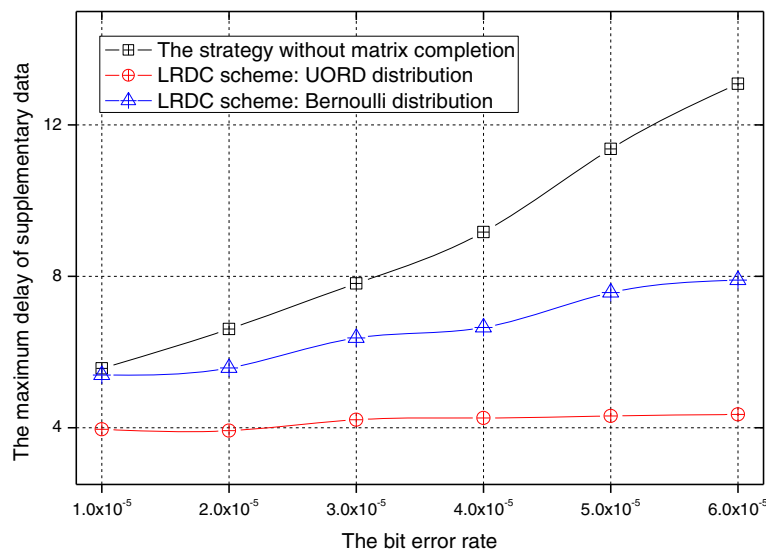
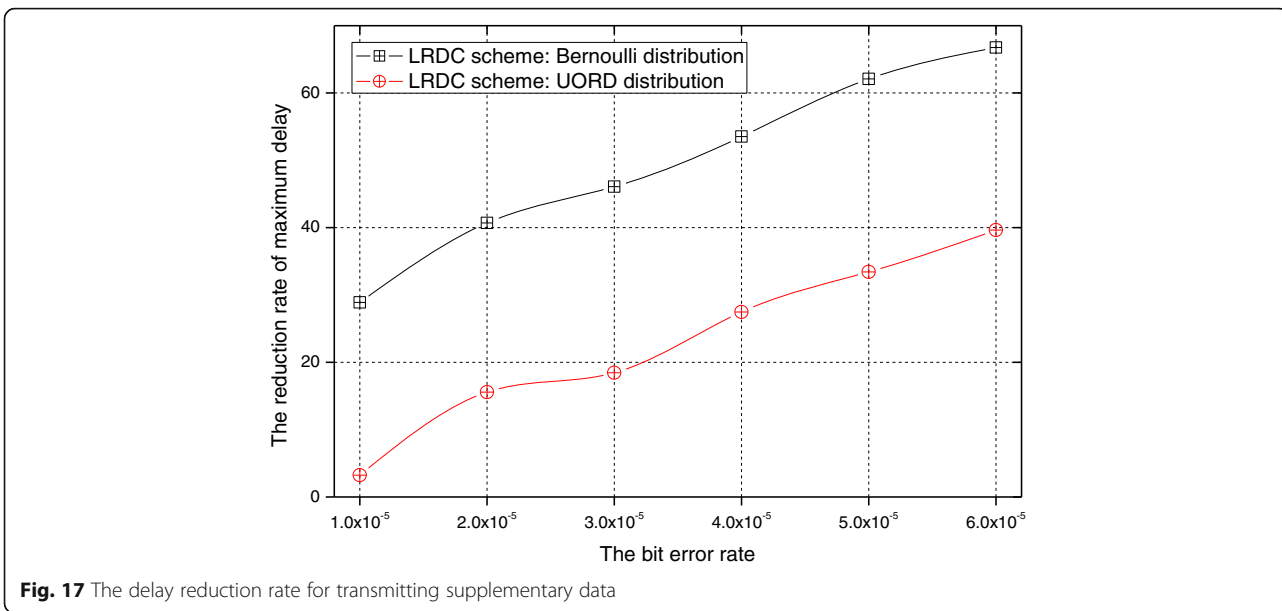


Fig. 16 The maximum delay of transmitting supplementary data



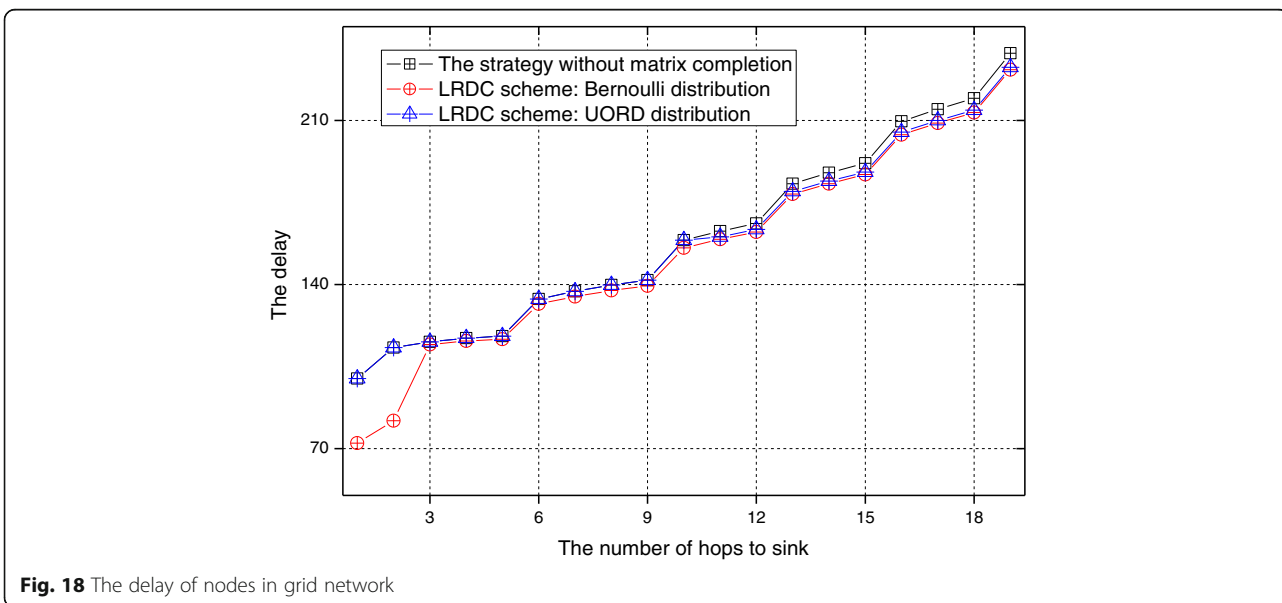
network. It can be seen that the area of each layer gradually increases, and the number of nodes in each layer gradually increases. Therefore, when matrix completion technique is not used, the total number of data packets to be sent per layer increases linearly. Because in the Bernoulli distribution, redundant data packets are uniformly distributed on each node, the total number of data packets to be sent per layer is also linearly increasing. In UORD distribution, redundant data is mainly distributed in high-layer nodes due to the higher number of retransmissions in high-layer node.

Figure 9 shows the location of redundant data packets stored in each node of planar network in LRDC scheme. It can be seen that in a network without using matrix completion technique, each node does not have redundant

data, so it is necessary to send data from the original node. In the network using the LRDC scheme, it can be seen that under the Bernoulli distribution, almost all redundant data of nodes are stored at the node that is from one hop from away from sink. Additionally, in the UORD distribution, some nodes still need to send data from itself, because they do not have redundant data packets.

Therefore, the storage location of redundant data has been obtained, and the amount of data forwarded by each node can be obtained, so the energy consumption of each node can also be obtained as in Theorem 10.

Theorem 10: For a planar network, using the LRDC scheme, the energy consumption of each node in the network is



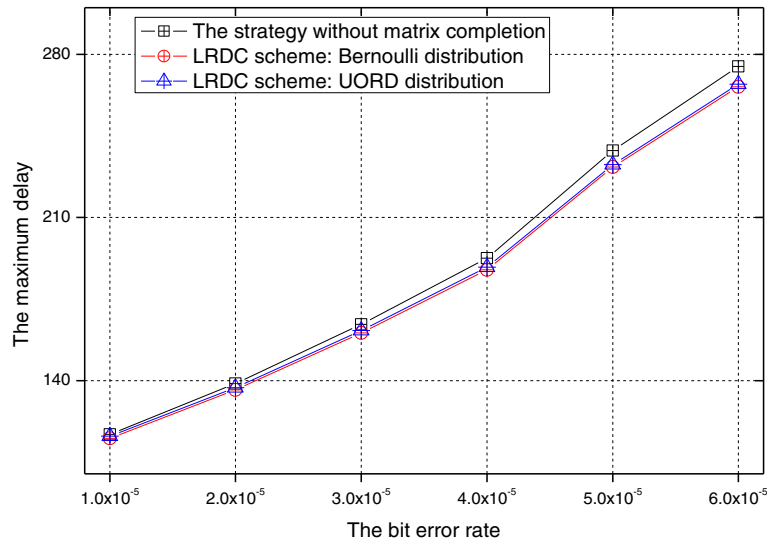


Fig. 19 The maximum delay in grid network

$$E_x = \begin{cases} E_b(D_x^{non} + D_x^r + D_x^s) \frac{x-1}{d} + 1 \geq Y \\ E_b(D_x^{non} + D_x^s) \frac{x-1}{d} + 1 < Y \end{cases} \quad (27)$$

where Y is the location of redundant packets for the node that distance from sink is x .

Proof: Similar to the grid network, the energy consumption of transmission can also be divided into three parts.

For the energy consumption for transmitting non-redundant data, considering that the node's non-redundant data is m_x . Similar to Theorem 7, the amount of data that uses the retransmission mechanism can be

obtained, and the amount of data forwarded by each node can be obtained according to (27) as D_x^{non} .

For redundant data packets, according to Theorem 8, the location to store redundant data packets can be obtained. The amount of redundant data of the node is $T - m_x$. Similar to the grid network, when the number of layers ($\frac{x-1}{d} + 1$) of the node is greater than the number of layers of the bottleneck node, according to Eq. (27), the amount of data forwarded by the node is D_x^r , while for the remaining nodes, the amount of data forwarded by the node is 0.

The supplementary data transmission is the same as in grid network. Theorem 8 can obtain the location of redundant data. When the redundant data of nodes is stored by other nodes, the supplementary data of the nodes is sent by

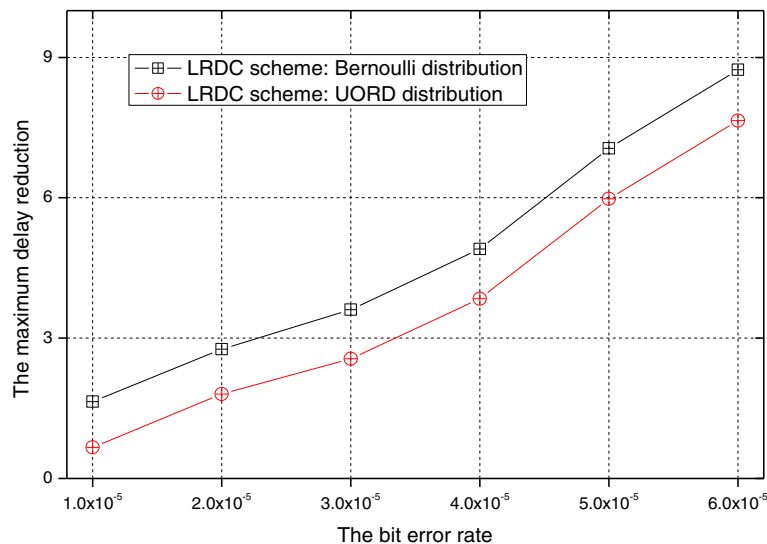


Fig. 20 The maximum delay reduction ratio in grid network

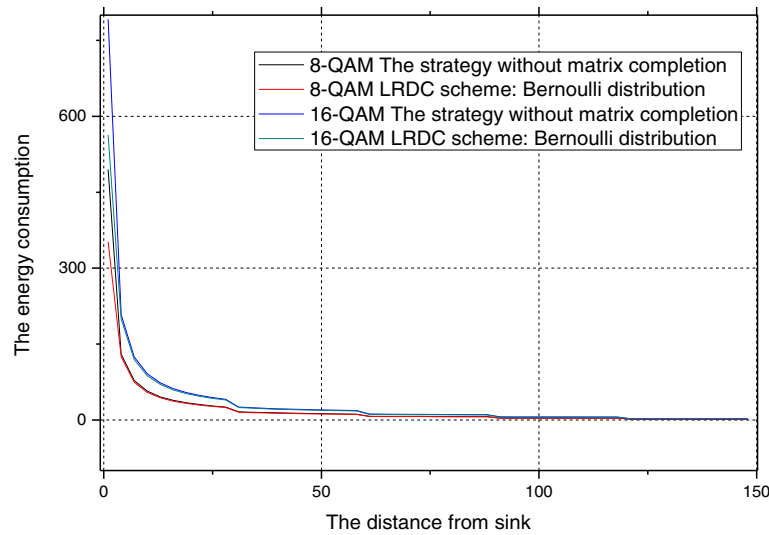


Fig. 21 The energy consumption of nodes in planar network under Bernoulli distribution

other nodes. When the node's redundant data is still in the original node or there is no redundant data, the supplementary is sent by itself. Therefore, it can also obtain the amount of data D_x^s by each node according to Corollary 1.

Reorganized the above, the energy consumption can be given as:

$$E_x = \begin{cases} E_b(D_x^{non} + D_x^r + D_x^s) \frac{x-1}{d} + 1 \geq Y \\ E_b(D_x^{non} + D_x^s) \frac{x-1}{d} + 1 < Y \end{cases}$$

Theorem 10 gives the energy consumption of various places in the planar network under LRDC scheme. With this theorem, the maximum energy consumption of the

network under different distributions of redundant data can be obtained.

Similarly, the LRDC scheme stores redundant data to nodes near the sink, so it also reduces the delay of network and the delay is as given in Theorem 11.

Theorem 11: For a planar network, the delay of each node under the LRDC scheme is as follows:

$$\phi_x = m_x \zeta_{k_1} + (T - m_x) \zeta_{k_2} + [(1 - \delta) m_x] \zeta_{k_3} \quad (28)$$

Proof: For a planar network, considering the node with the distance x from the sink, similar to grid network, the delay is also divided into three parts.

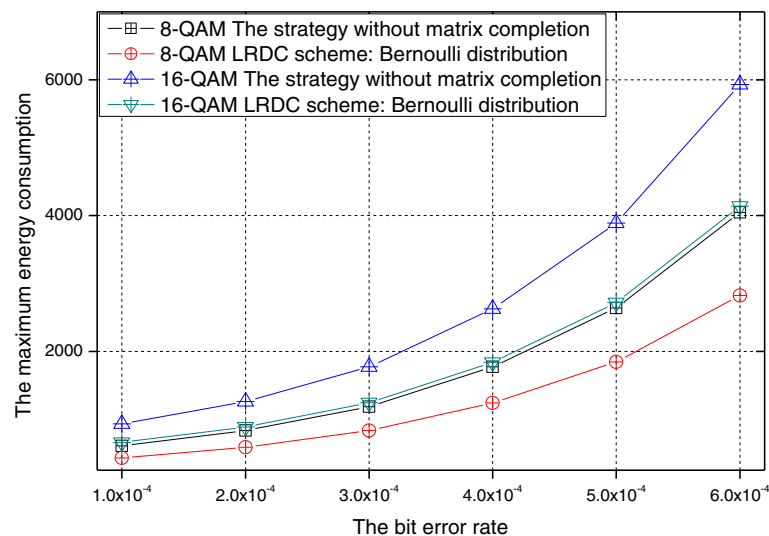


Fig. 22 The maximum energy consumption of planar network under Bernoulli distribution

For the delay caused by the transmission of non-redundant data packets, it is considered that the nodes in the S_x area need to pass k_1 hops to sink. Therefore, under the retransmission mechanism, the delay for transmitting this part of the data is $m_x c_{k_1}$.

When transmitting redundant data, it is sufficient to ensure that the redundant data arrives at the node where the redundant data is stored, and the storage location can be obtained by Theorem 8. Therefore, the number of hops from the S_x area to the storage location can be obtained. Considering that is k_2 , this part of the delay is $(T - m_x) c_{k_2}$.

Finally, in the LRDC scheme, the S_x area needs to be supplemented with $\lceil (1 - \delta) m_x \rceil$ data packets in the supplementary transmission, and the transmission can be started directly from the node where redundant data packet is stored. Supposing that the number of hops from the node that stores the redundant data packet to sink is k_3 , the delay is $\lceil (1 - \delta) m_x \rceil c_{k_3}$.

Algorithm 2: The UORD scheme for maximizing lifetime in planar network

Initialize: Considering the number of sensor nodes N , and the data collection rounds T .

Input: the number of sensor nodes N , the maximum number of redundant data packets c .

Output: the amount of data forwarded by each area of the network D_x .

```

1: the number of redundant packets  $m$  can be calculated
   by Eq. (10).
2: the amount of data  $D_x$  forwarded by each node is
   obtained by Eq. (26).
3: while  $m > 0$ 
4:   Defined a value ( $< 0$ )  $\alpha$  and a mark  $\varepsilon$ 
5:   For  $i = 1$  to  $d$ 
6:     If  $D_i > \alpha$ 
7:        $\alpha = D_i$ ,  $\varepsilon = i$ 
8:     End if
9:   End for
10:   $\alpha =$  a value ( $< 0$ ) and defined a mark  $\tau$ 
11:   $i = \varepsilon$ 
12:  while  $i < R$ 
13:     $Q_\varepsilon^i = \frac{i}{\varepsilon} \lambda_i$ 
        // the contribution of the distance from the sink
        // is  $i$  to the amount of data at the distance  $\varepsilon$ 
14:    If  $L_i < c$  and  $D_i - Q_\varepsilon^i > \alpha$ 
15:       $\alpha = D_i - Q_\varepsilon^i$ ,  $\tau = i$ 
16:    End if
17:     $i = i + d$ 
18:  End while
19:   $L_\tau ++$ 
20:   $k = \tau$ 
21:  while  $k \geq \varepsilon$ 
22:     $Q_k^\tau = \frac{\tau}{k} \lambda_\tau$ ,  $D_k = D_k - Q_k^\tau$ 
23:     $k = k - d$ 
24:  End while
25:   $m = m - \frac{2\pi\tau}{\pi R^2} N$ 
        // the number of nodes away from sink distance  $\tau$ 
26: End while

```

6 The experimental results and discussion

In this section, simulation results for LRDC scheme are provided. In the following simulations, the performance in grid networks is first evaluated. Due to the large amount of similarities in the collected data, it is considered that the rank of data matrix is 5, the transmission radius of each node is 30 m, and the total number of sensor nodes on the network is 100.

6.1 Performance evaluation in grid network

First, the performance of the network is evaluated for the LRDC scheme. When redundant data obeys Bernoulli distribution, the data packets on each sensor node is uniform. Therefore, the energy consumption of each node in the network can be calculated according to Eq. (23). The total energy consumption of the network is shown in Fig. 10.

Figure 10 shows the total energy consumption at different bit error rates. It can be seen that the total energy consumption of the network increases with the increase of the bit error rate. It can also be seen that the total energy consumption using the matrix completion technique and without the matrix completion technique has no obvious improvement. This is because the LRDC scheme uses the residual energy in the node to transmit redundant data to the node close to sink.

Figure 11 is the maximum energy consumption of the network. The maximum energy consumption using the LRDC scheme is greatly reduced compared with the scheme without using matrix completion. Combined with Fig. 12, it can be seen that the maximum energy consumption of the network is reduced by about 27.6%, and the improvement is similar under different bit error rates. Therefore, the LRDC scheme has a significant improvement in network lifetime.

From the above, it can be seen that the LRDC scheme can improve the network lifetime under Bernoulli distribution model, but the network lifetime can be further optimized. Next, the effect of UORD distribution on network performance will be studied.

Figure 13 shows the total energy consumption of the LRDC scheme under the UORD distribution. It can be seen that the total energy consumption is slightly reduced under the UORD distribution. The improvement is the same as Bernoulli distribution, and UORD distribution also has little effect on reducing the total energy consumption.

Figure 14 shows the maximum energy consumption of the network. It can be clearly seen that the maximum energy consumption of UORD distribution is the lowest. Combined with Fig. 15, the maximum energy consumption of the LRDC scheme under two distributions can be reduced by more than 27.6%, compared with the network without using matrix completion technique. Moreover, as

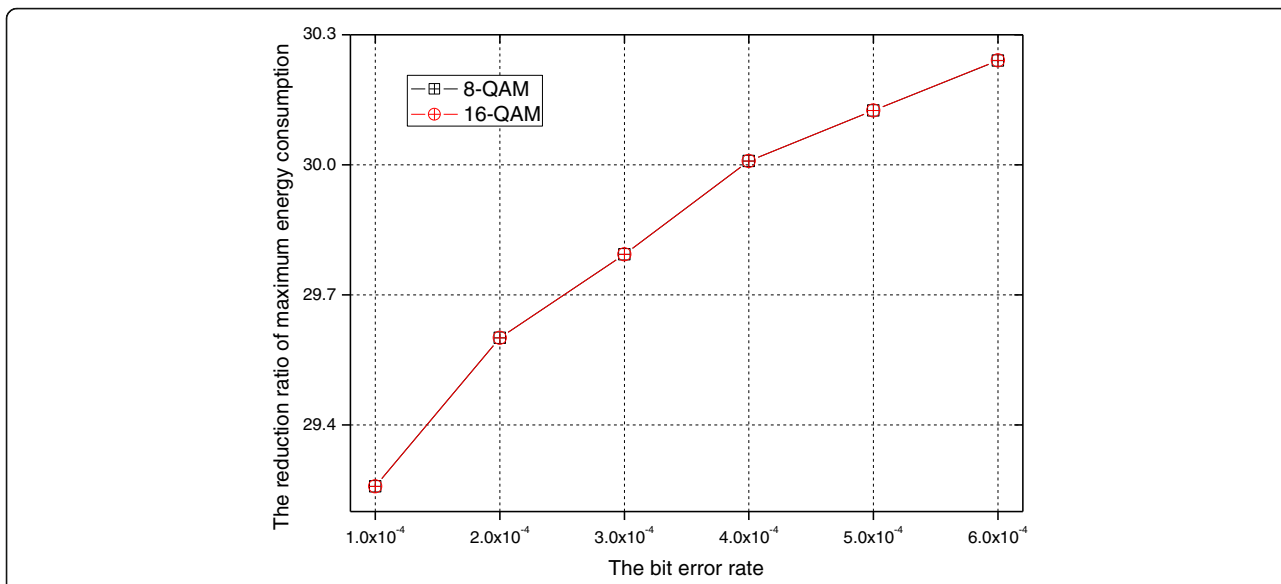


Fig. 23 The maximum energy consumption reduction ratio

the bit error rate increases, the improvement under the UORD distribution will increase, the optimized effect can exceed 32.5%.

Since redundant data is stored in the near-sink node in LRDC scheme, delay can be reduced during the data retransmission.

Figure 16 is the maximum delay for transmitting supplementary data in grid network. Due to the supplementary data is stored in the near sink node in UORD distribution so the optimization effect of the UORD distribution is better. From Fig. 17, under the Bernoulli distribution, the delay of the supplementary data transmission is reduced by at least 28.9%, and under the UORD distribution, there

is an obvious reduction in high bit error rate, which can be reduced to 39.6%.

Figure 18 shows the delay of nodes in grid network with different hops to sink. It can be seen that in the Bernoulli distribution model, the delay of all nodes in network is reduced, but under the UORD distribution, the delay of near sink node is not improved. This is because the number of retransmissions away from sink is high, and reducing the amount of data can reduce more energy consumption.

Figure 19 shows the maximum delay of the network. It can be seen that the difference in maximum delays between the Bernoulli distribution and UORD distribution

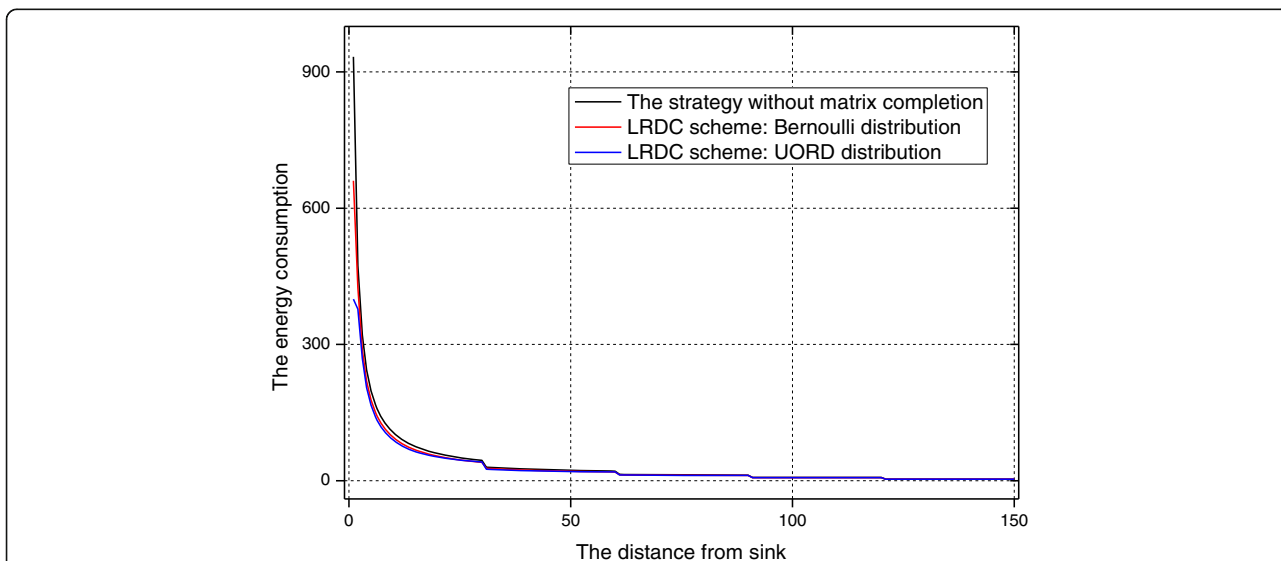


Fig. 24 The energy consumption of nodes in planar network under UORD distribution

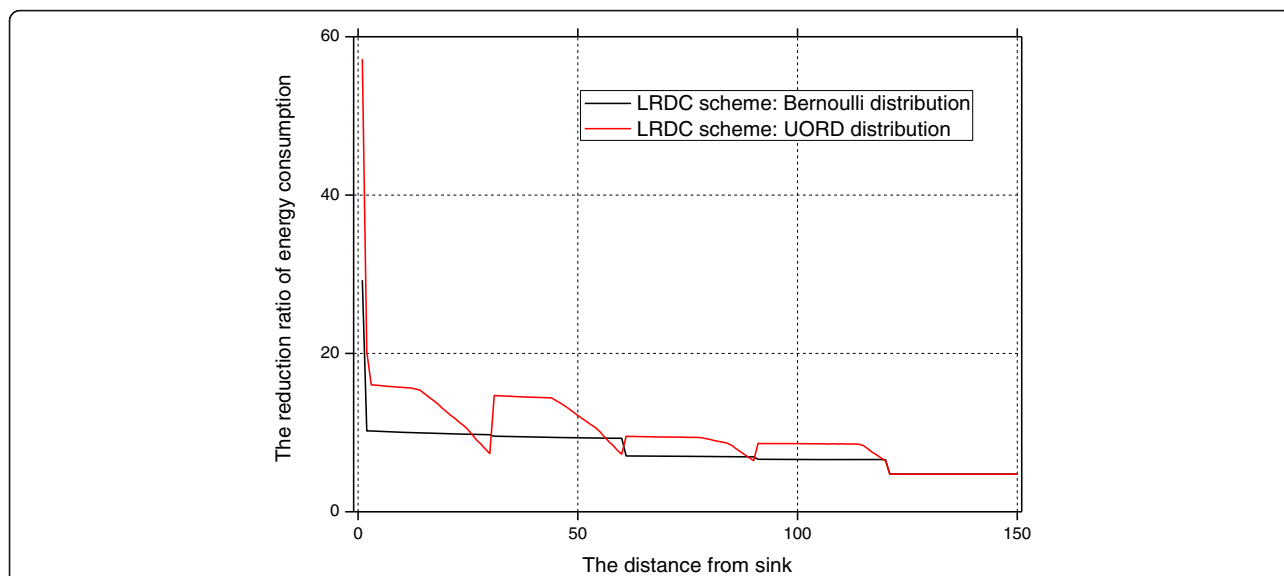


Fig. 25 The energy consumption reduction ratio in planar network under UORD distribution

is small. Combined with Fig. 20, the network that is using LRDC scheme has better delay compared with the network without using the LRDC scheme, but the gain is not large. The maximum delay of the entire network can only be reduced by about 8.7% at most, and as the bit error rate increases, the gain will also increase.

6.2 Performance evaluation in planar network

In the following experiments, a more common planar network is considered. There are total of 100 nodes in the network, and each node transmits 100 rounds of data. The transmission radius of node is 30 m, and the radius of the network is 150 m.

Figure 21 shows the energy consumption of various regions in the network. It can be seen that in the near-sink region, the energy consumption of the nodes is obviously reduced. In the area closest to sink, the energy consumption can be reduced by approximately 37.5%, and there is no obvious improvement under different modulation rates.

First, the maximum energy consumption of a planar network under Bernoulli distribution is studied. As shown in Fig. 22, the energy consumption of the LRDC scheme is lower than that of not using matrix completion technique. Combined with Fig. 23, it can be seen that the maximum energy consumption is reduced by

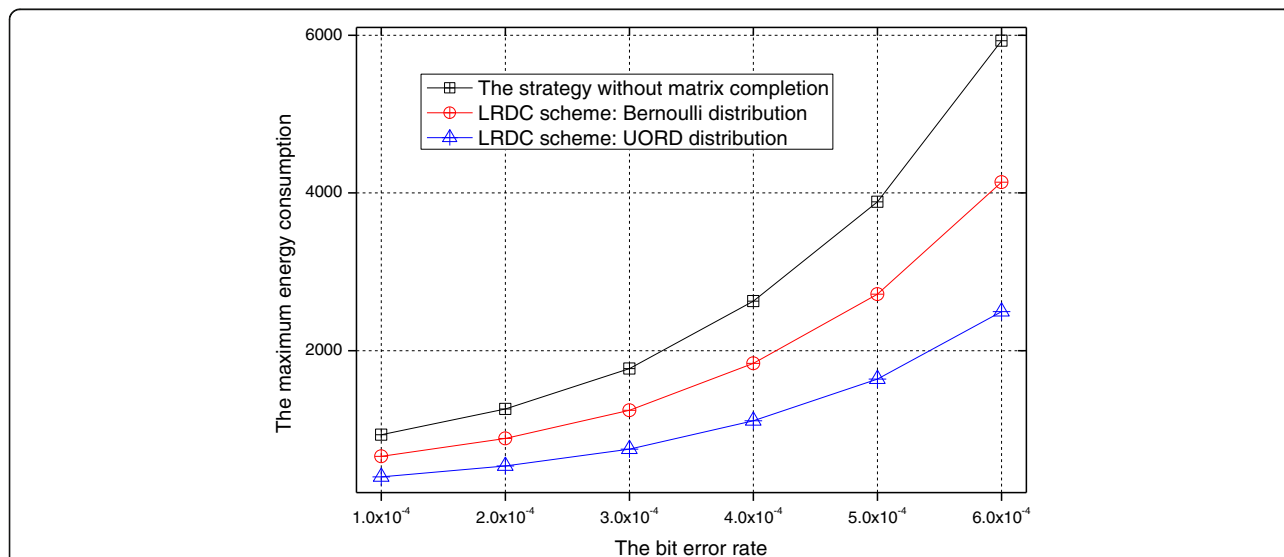


Fig. 26 The maximum energy consumption of planar network under UORD distribution

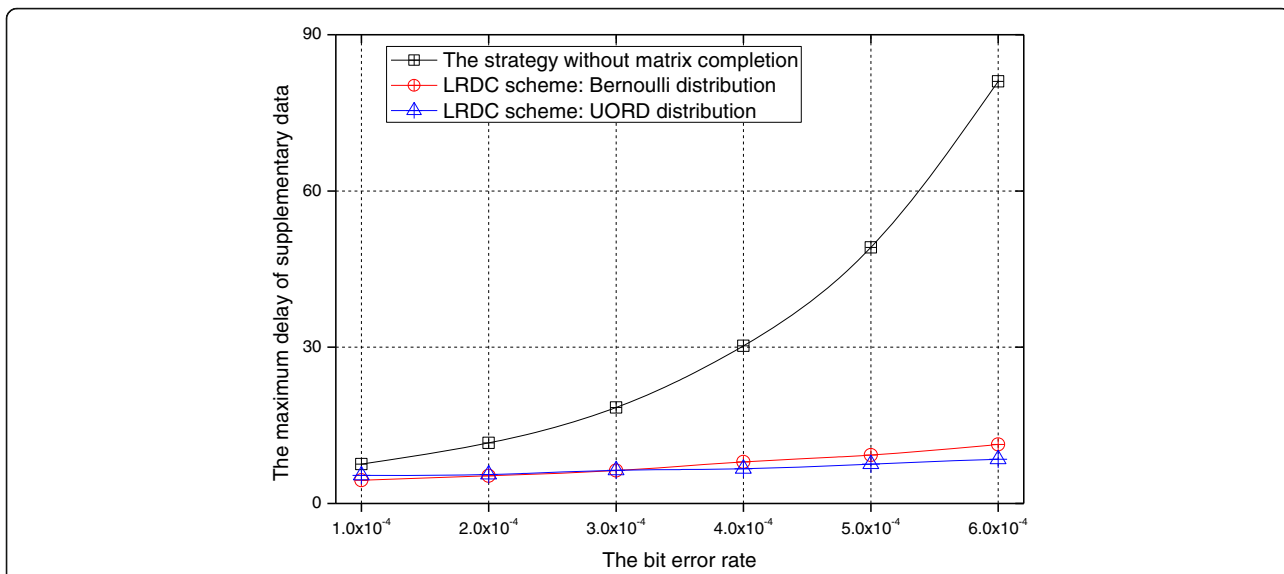


Fig. 27 The delay for transmitting supplementary data in planar network

more than 29.4% compared to a network without matrix completion technology, and with the increase of the bit error rate.

Figure 24 shows the energy consumption of the network under the UORD distribution of the LRDC scheme. It can be seen that the improvement is very small in the region of the far sink node, as in the Bernoulli distribution. However, as can be seen in Fig. 25, in the near-sink region, the energy consumption of the network under UORD distribution is reduced by up to approximately 57.2%, and the near-sink region is the region with the highest energy consumption in the planar network, so the UORD distribution improves network lifetime well.

Figure 26 shows the maximum energy consumption of the network under both distributions. It is clear that the maximum energy consumption of the two distributions using the LRDC scheme is smaller than that without the matrix completion technique. The maximum energy consumption of the network is reduced by 57% under the UORD distribution.

Figure 27 shows the maximum delay of the network during the supplementary data transmission. It can be seen that the LRDC scheme can significantly reduce the delay when transmitting supplementary data. As shown in Fig. 28, it can be seen that as the bit error rate increases, the gain of LRDC scheme increases. The maximum delay

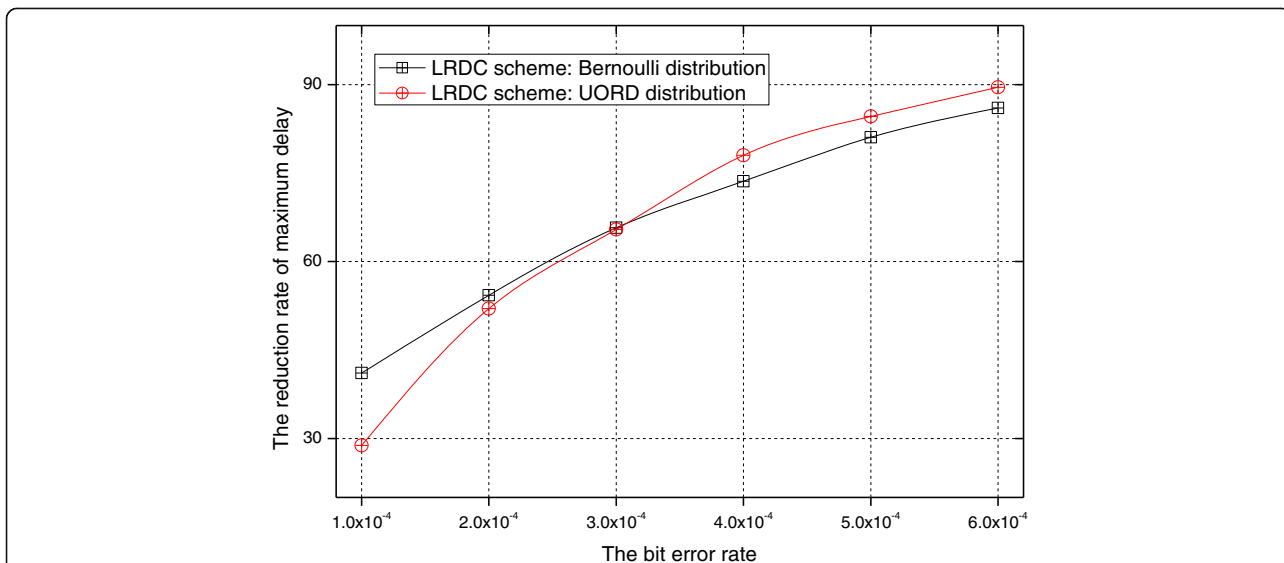


Fig. 28 The delay reduction ratio of transmitted supplementary data

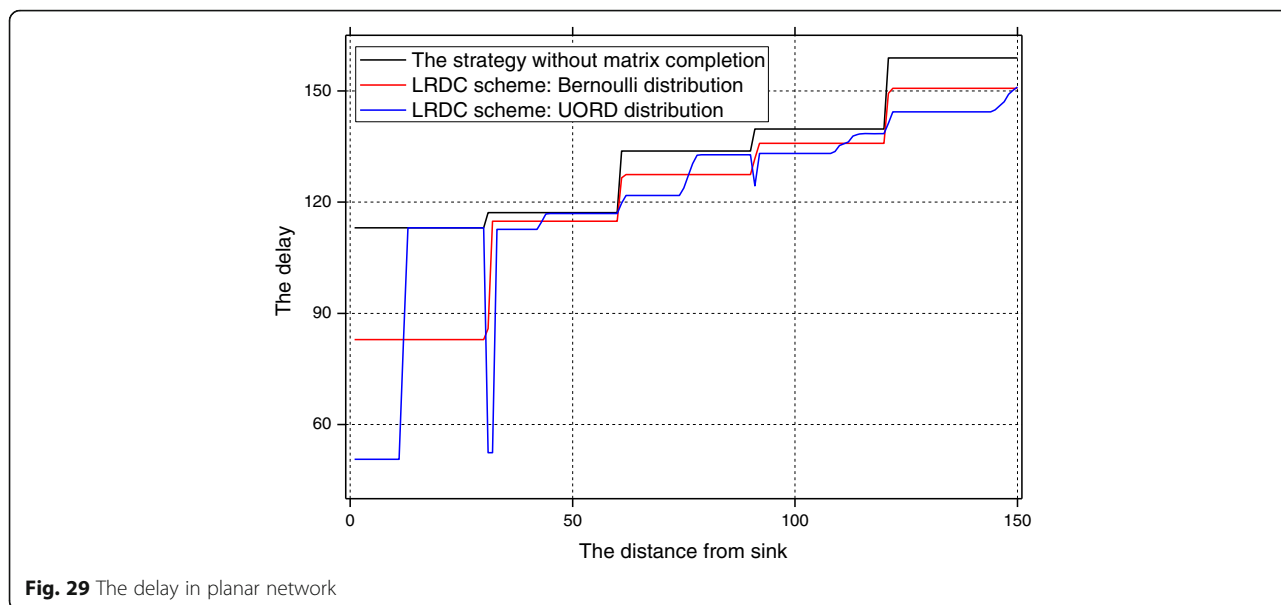


Fig. 29 The delay in planar network

of supplementary data transmission is reduced by more than 80% at high bit error rates.

Figure 29 shows the delay of each node in the network. It can be seen that under the Bernoulli distribution, the delay of the nodes in each area will have the improvement, but under the UORD distribution, it only has a larger improvement for delay in some regions. This is because in the UORD distribution, redundant data is concentrated on some nodes that contribute to the node with the highest energy consumption, and the total amount of redundant data is limited, resulting in many areas without delay improvement.

Figure 30 shows the maximum delay of the network under different bit error rates. The maximum delay of the

network will increase with the increase of bit error rate. As can be seen in Fig. 31, the UORD distribution is better. Similar to the Bernoulli distribution, with the increase of bit error rate, we can have more improvements. The reduction ratio of delay can exceed 17.9% at most.

7 Conclusion

In this paper, we propose an LRDC strategy based on matrix completion technique to optimize the performance in terms of network lifetime and delay. Different from existing works, the proposed scheme makes efficient use of the correlation of the data collected by the sensor nodes. By so doing, only a part of the data is collected, all the data can be recovered using the matrix completion

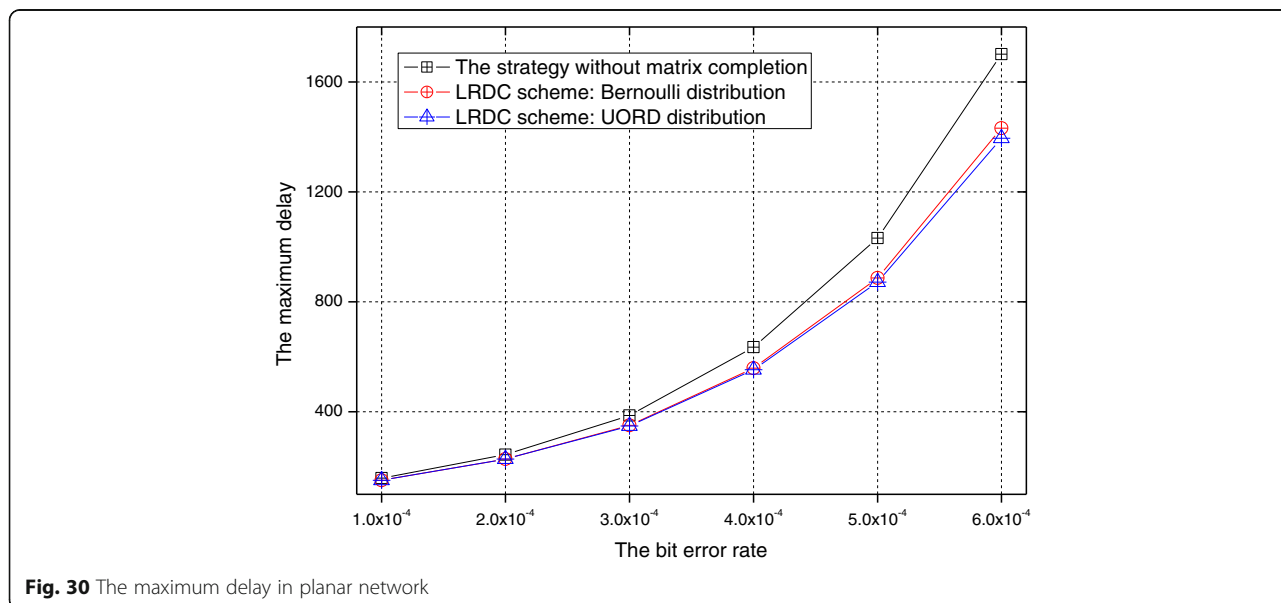


Fig. 30 The maximum delay in planar network

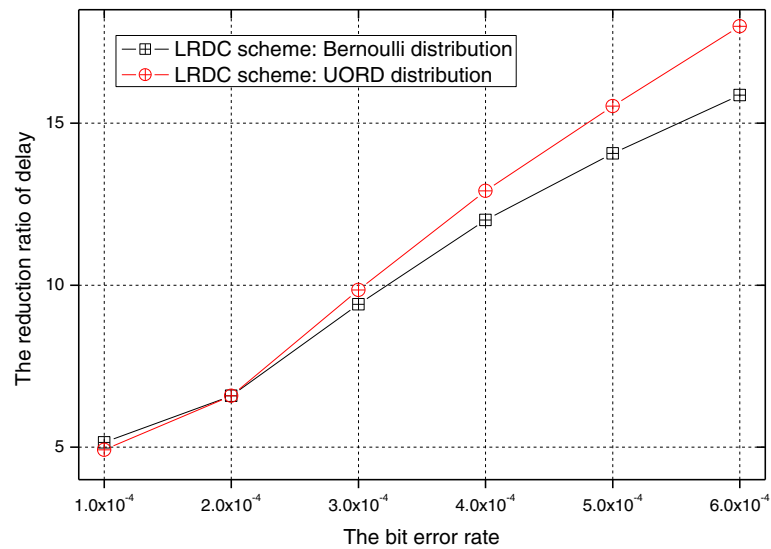


Fig. 31 The maximum delay reduction ratio of planar network

technique, thereby reducing the energy consumption of the transmission and increasing the network lifetime. At the same time, simply reducing the number of data packets sent by each node cannot effectively improve the energy efficiency of the network. There is still residual energy in the area far away from CC, so we can use this part of energy to transfer the backup data set of each node in the network to the area near the CC. Once the data is lost due to the unreliability of data transmission, the data can be supplemented directly from the nodes near CC to satisfy the amount of data required by the matrix completion technique, which can reduce delay, while not drastically affecting the network lifetime.

Abbreviations

CC: Control center; DCs: Data centers; EPA: The U.S. environmental protection agency; IOT: Internet of thing; LRDC: Low redundancy data collection; UORD: Unbalanced of redundant data; WSN: Wireless sensor network

Acknowledgements

The authors thank the editors and referees very much for elaborate and valuable suggestions which helped to improve the paper.

Funding

This work was supported in part by the National Natural Science Foundation of China (61772554, 61572526, 61572528), and the Natural Science Foundation of Zhejiang Province (No. LY17F020032).

Availability of data and materials

Not applicable.

Authors' contributions

Jiawei Tan is the main author of the current paper. Anfeng Liu contributed to the conception and design of the study. Wei Liu, Mande Xie, Houbing Song, Ming Zhao, and Guoping Zhang commented the work. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹School of Information Science and Engineering, Central South University, Changsha 410083, China. ²School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China. ³School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China. ⁴Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA. ⁵School of Software, Central South University, Changsha 410075, China. ⁶Faculty of Informatics & Electronics, Zhejiang SCI-Tech University, Zhejiang 310018, Hangzhou, China.

Received: 13 September 2018 Accepted: 30 November 2018

Published online: 08 January 2019

References

- S. Sarkar, S. Chatterjee, S. Misra, Assessment of the suitability of fog computing in the context of internet of things. *IEEE Trans. Cloud Comput* **6**(1), 46–59 (2018)
- M. Wu, Y. Wu, C. Liu, Z. Cai, N. Xiong, A. Liu, M. Ma, An effective delay reduction approach through portion of nodes with larger duty cycle for industrial WSNs. *Sensors* **18**(5), 1535 (2018). <https://doi.org/10.3390/s18051535>
- Y. Ren, W. Liu, Y. Liu, N. Xiong, A. Liu, X. Liu, An effective crowdsourcing data reporting scheme to compose cloud-based services in mobile robotic systems. *IEEE Access* **6**(1), 54683–54700 (2018)
- M. Zhang, P. Yang, C. Tian, S. Tang, X. Gao, B. Wang, F. Xiao, Quality-aware sensing coverage in budget-constrained mobile crowdsensing networks. *IEEE Trans. Veh. Technol.* **65**(9), 7698–7707 (2016)
- S. Yu, X. Liu, A. Liu, N. Xiong, Z. Cai, T. Wang, Adaption broadcast radius based code dissemination scheme for low energy wireless sensor networks. *Sensors* **18**(5), 1509 (2018). <https://doi.org/10.3390/s18051509>
- Z. Li, Y. Liu, M. Ma, A. Liu, X. Zhang, G. Luo, MSDG: A novel green data gathering scheme for wireless sensor networks. *Comput. Netw.* **142**(4), 223–239 (2018)

7. Y. Li, C. Ai, C. Vu, Y. Pan, R. Beyah, Delay-bounded and energy-efficient composite event monitoring in heterogeneous wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **21**(9), 1373–1385 (2010)
8. X. Liu, Y. Liu, N. Xiong, N. Zhang, A. Liu, H. Shen, C. Huang, Construction of large-scale low cost deliver infrastructure using vehicular networks. *IEEE Access* (2018). <https://doi.org/10.1109/ACCESS.2018.2825250>
9. X. Liu, W. Liu, Y. Liu, H. Song, A. Liu, X. Liu, A trust and priority based code updated approach to guarantee security for vehicles network. *IEEE Access* (2018). <https://doi.org/10.1109/ACCESS.2018.2872787>
10. P. Yang, Y. Yan, X.Y. Li, Y. Zhang, Y. Tao, L. You, Taming cross-technology interference for Wi-Fi and ZigBee coexistence networks. *IEEE Trans. Mob. Comput.* **15**(4), 1009–1021 (2016)
11. M.Z.A. Bhuiyan, G. Wang, J. Wu, J. Cao, et al., Dependable structural health monitoring using wireless sensor networks. *IEEE Trans. Dependable Secure Comput.* **14**(4), 363–376 (2017)
12. J. Li, Z. Liu, X. Chen, F. Xhafa, X. Tan, D. Wong, L-EncDB: a lightweight framework for privacy-preserving data queries in cloud computing. *Knowl.-Based Syst.* **79**, 18–26 (2015)
13. X. Wang, Z. Ning, L. Wang, Offloading in internet of vehicles: a fog-enabled real-time traffic management system. *IEEE Trans. Ind. Inf.* (2018). <https://doi.org/10.1109/TII.2018.2816590>
14. Z. Ding, K. Ota, Y. Liu, N. Zhang, M. Zhao, H. Song, A. Liu, Z. Cai, Orchestrating data as services based computing and communication model for information-centric internet of things. *IEEE Access* **6**(1), 38900–38920 (2018)
15. T. Han, N. Ansari, Network utility aware traffic loading balancing in backhaul-constrained cache-enabled small cell networks with hybrid power supplies. *IEEE Trans. Mob. Comput. (TMC)* **16**(10), 2819–2832 (2017)
16. M. Huang, A. Liu, M. Zhao, T. Wang, Multi working sets alternate covering scheme for continuous partial coverage in WSNs. *Peer-to-Peer Netw. Appl.* (2018). <https://doi.org/10.1007/s12083-018-0647-z>
17. L. Guo, Z. Ning, W. Hou, B. Hu, P. Guo, Quick answer for big data in sharing economy: innovative computer architecture design facilitating optimal service-demand matching. *IEEE Trans. Autom. Sci. Eng.* (2018). <https://doi.org/10.1109/TASE.2018.2838340>
18. K. Ota, M.S. Dao, V. Mezaris, F.G.B. De Natale, Deep learning for mobile multimedia: a survey. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **13**(3s), 34 (2017)
19. Y. Ren, Y. Liu, N. Zhang, A. Liu, N. Xiong, Z. Cai, Minimum-cost mobile crowdsourcing with QoS guarantee using matrix completion technique. *Pervasive Mob. Comput.* **49**, 23–44 (2018)
20. S. Cheng, Z. Cai, J. Li, H. Gao, Extracting kernel dataset from big sensory data in wireless sensor networks. *IEEE Trans. Knowl. Data Eng.* **29**(4), 813–827 (2017)
21. X. Liu, M. Dong, K. Ota, L.T. Yang, A. Liu, Trace malicious source to guarantee cyber security for mass monitor critical infrastructure. *J. Comput. Syst. Sci.* (2016). <https://doi.org/10.1016/j.jcss.2016.09.008>
22. Y. Li, C. Vu, C. Ai, G. Chen, Y. Zhao, Transforming complete coverage algorithms to partial coverage algorithms for wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **22**(4), 695–703 (2011)
23. Z. He, Z. Cai, S. Cheng, X. Wang, Approximate aggregation for tracking quantiles and range Countings in wireless sensor networks. *Theor. Comput. Sci.* **607**(3), 381–390 (2015)
24. M. Chen, Y. Li, X. Luo, W. Wang, L. Wang, W. Zhao, A novel human activity recognition scheme for smart health using multilayer extreme learning machine. *IEEE Internet Things J.* (2018). <https://doi.org/10.1109/JIOT.2018.2856241>
25. W. Jiang, G. Wang, M.Z.A. Bhuiyan, J. Wu, Understanding graph-based trust evaluation in online social networks: methodologies and challenges. *ACM Comput. Surv.* **49**(1), 10 (2016)
26. C. Zhou, Y. Gu, S. He, et al., A robust and efficient algorithm for coprime array adaptive beamforming. *IEEE Trans. Veh. Technol.* **67**(2), 1099–1112 (2018)
27. K. Xie, J. Cao, X. Wang, J. Wen, Optimal resource allocation for reliable and energy efficient cooperative communications. *IEEE Trans. Wirel. Commun.* **12**(10), 4994–5007 (2013)
28. X. Liu, Y. Liu, A. Liu, L. Yang, Defending on-off attacks using light probing messages in smart sensors for industrial communication systems. *IEEE Trans. Ind. Inf.* **14**(9), 3801–3811 (2018)
29. B. Huang, A. Liu, C. Zhang, N. Xiong, Z. Zeng, Z. Cai, Caching joint shortcut routing to improve quality of experiments of users for information-centric networking. *Sensors* **18**(6), 1750 (2018). <https://doi.org/10.3390/s18061750>
30. X. Ju, W. Liu, C. Zhang, A. Liu, T. Wang, N. Xiong, Z. Cai, An energy conserving and transmission radius adaptive scheme to optimize performance of energy harvesting sensor networks. *Sensors* **18**(9), 2885 (2018). <https://doi.org/10.3390/s18092885>
31. J. Zhang, X. Hu, Z. Ning, E. Ngai, L. Zhou, J. Wei, J. Cheng, B. Hu, Energy-latency trade-off for energy-aware offloading in mobile edge computing networks. *IEEE Internet Things J.* **5**(4), 2633–2645 (2018)
32. J. He, S. Ji, Y. Pan, Y. Li, Constructing load-balanced data aggregation trees in probabilistic wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **25**(7), 1681–1690 (2014)
33. X. Xu, N. Zhang, H. Song, A. Liu, M. Zhao, Z. Zeng, Adaptive beaconing based MAC protocol for sensor based wearable system. *IEEE Access* **6**, 29700–29714 (2018)
34. T. Li, N. Xiong, J. Gao, H. Song, A. Liu, Z. Zeng, Reliable code disseminations through opportunistic communication in vehicular wireless networks. *IEEE Access* **6**(1), 55509–55527 (2018)
35. A. Liu, Q. Liu, On the hybrid using of unicast-broadcast in wireless sensor networks. *Comput. Electr. Eng.* (2017). DoI: <https://doi.org/10.1016/j.compeleceng.2017.03.004>.
36. T. Li, Y. Liu, N. Xiong, A. Liu, Z. Cai, H. Song, Privacy-preserving protocol of sink node location in telemedicine networks. *IEEE Access* **6**(1), 42886–42903 (2018)
37. A. Liu, S. Zhao, High performance target tracking scheme with low prediction precision requirement in WSNs. *Int. J. Ad Hoc Ubiquitous Comput.* **29**(4), 270–289 (2018)
38. M.Z.A. Bhuiyan, J. Wu, G. Wang, T. Wang, et al., E-sampling: event-sensitive autonomous adaptive sensing and low-cost monitoring in networked sensing systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* **12**(1), 1 (2017)
39. Z. Cai, T. Zhang, X. Wan, A computational framework for influenza antigenic cartography. *PLoS. Comput. Biol.* **6**(10), e1000949 (2010)
40. C. Yang, Z. Shi, K. Han, J. Zhang, Y. Gu, Z. Qin, Optimization of particle CBMeMber filters for hardware implementation," *IEEE Trans. Veh. Technol.*, DOI: 10.1109/TVT.2018.2853120, (2018).
41. X. Luo, J. Deng, J. Liu, W. Wang, X. Ban, J.H. Wang, A quantized kernel least mean square scheme with entropy-guided learning for intelligent data analysis. *China Communications* **14**(7), 127–136 (2017)
42. H. Teng, K. Zhang, M. Dong, K. Ota, A. Liu, M. Zhao, T. Wang, Adaptive transmission range based topology control scheme for fast and reliable data collection. *Wirel. Commun. Mob. Comput.*, 2018, 4172049, (2018). DoI: <https://doi.org/10.1155/2018/4172049>.
43. Y. Liu, M. Dong, K. Ota, A. Liu, ActiveTrust: Secure and trustable routing in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **11**(9), 2013–2027 (2016)
44. R. Anane, K. Raoof, R. Bouallegue, Minimization of wireless sensor network energy consumption through optimal modulation scheme and channel coding strategy. *J. Signal Proces. Syst.* **83**(1), 65–81 (2016)
45. Z. Li, Y. Liu, A. Liu, S. Wang, H. Liu, Minimizing convergencast time and energy consumption in green internet of things. *IEEE Trans. Emerg. Top. Comput.* (2018). <https://doi.org/10.1109/TETC.2018.2844282>.
46. P. Yang, Q. Li, Y. Yan, X.Y. Li, Y. Xiong, B. Wang, X. Sun, "Friend is treasure": exploring and exploiting mobile social contacts for efficient task offloading. *IEEE Trans. Veh. Technol.* **65**(7), 5485–5496 (2016)
47. P. Le, Y. Nguyen, Z. Ji, H.V. Liu, K.V. Nguyen, Distributed hole-bypassing protocol in WSNs with constant stretch and load balancing. *Comput. Netw.* **129**, 232–250 (2017)
48. Z. Liu, T. Tsuda, H. Watanabe, S. Ryo, N. Iwasawa, Data driven cyber-physical system for landslide detection. *ACM/springer Mob. Netw. Appl.* (2018). <https://doi.org/10.1007/s11036-018-1031-1>
49. X. Chen, Y. Hu, A. Liu, Z. Chen, Cross layer optimal design with guaranteed reliability under Rayleigh block fading channels. *KSII Trans. Internet Inf. Syst.* **7**(12), 3017–3095 (2013)
50. J. Li, P. Mohapatra, Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. *Pervasive Mob. Comput.* **3**(3), 233–254 (2007)
51. N. Jan, N. Javaid, Q. Javaid, N. Alrajeh, M. Alam, Z.A. Khan, I.A. Niaz, A balanced energy-consuming and hole-alleviating algorithm for wireless sensor networks. *IEEE Access* **5**, 6134–6150 (2017)
52. S. Cheng, Z. Cai, J. Li, X. Fang, *IEEE Conference on computer communication. Drawing dominant dataset from big sensory data in wireless sensor networks (IEEEGlasgow, 2015)* (2015), pp. 531–539

53. E.J. Candès, B. Recht, Exact matrix completion via convex optimization. *Found. Comput. Math.* **9**(6), 717 (2009)
54. E.J. Candès, T. Tao, The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56**(5), 2053–2080 (2010)
55. K. Xie, L. Wang, X. Wang, G. Xie, J. Wen, Low cost and high accuracy data gathering in WSNs with matrix completion. *IEEE Trans. Mob. Comput.* **17**(7), 1595–1608 (2018)
56. W. Qi, W. Liu, X. Liu, A. Liu, T. Wang, N. Xiong, Z. Cai, Minimizing delay and transmission times with long lifetime in code dissemination scheme for high loss ratio and low duty cycle WSNs. *Sensors* **18**(10), 3516 (2018)
57. S. Fang, Z. Cai, W. Sun, A. Liu, F. Liu, Z. Liang, G. Wang, Feature selection method based on class discriminative degree for intelligent medical diagnosis. *CMC: Computers, Materials & Continua* **55**(3), 419–433 (2018)
58. T. Li, S. Tian, A. Liu, H. Liu, T. Pei, DDSV: optimizing delay and delivery ratio for multimedia big data collection in Mobile sensing vehicles. *IEEE Internet Things. J.* (2018). <https://doi.org/10.1109/JIOT.2018.2847243>.
59. X. Li, W. Liu, M. Xie, A. Liu, M. Zhao, N. Xiong, M. Zhao, W. Dai, Differentiated data aggregation routing scheme for energy conserving and delay sensitive wireless sensor networks. *Sensors* **18**(7), 2349 (2018). <https://doi.org/10.3390/s18072349>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
