IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 15, 2018, accepted December 26, 2018, date of publication January 4, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2890862

An Adaptive Collection Scheme-Based Matrix Completion for Data Gathering in Energy-Harvesting Wireless Sensor Networks

JIAWEI TAN[1], WEI LIU[2], TIAN WANG[3], NEAL N. XIONG[4], HOUBING SONG[5], (Senior Member, IEEE), ANFENG LIU[1,6], AND ZHIWEN ZENG[1]

[1]School of Information Science and Engineering, Central South University, Changsha 410083, China
[2]School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China
[3]Department of Computer Science and Technology, Huaqiao University, Xiamen 361021, China
[4]Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA
[5]Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA
[6]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

Corresponding author: Anfeng Liu (afengliu@mail.csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572528, Grant 61772554, and Grant 6157256, and in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China, under Grant ICT1800391.

ABSTRACT Advanced communications and networks greatly enhance the user experience and have a major impact on all aspects of people's lifestyles. Widely deployed sensor nodes provide support for these services. However, although energy harvesting and transfer technology provides a solution to allow the long-term survival of wireless sensor nodes for wireless sensor networks, the single collection scheme causes a lot of energy waste. Thus, efficient energy utilization and fast data collection are still serious challenges for energy harvesting wireless sensor networks. To overcome these challenges, an adaptive collection scheme based on matrix completion (ACMC) is proposed to reduce delay and to improve the energy utilization of the network. In the ACMC scheme, compared with traditional data collection schemes, the data collection schemes vary with the available energy, collecting large amounts of data when the available energy is sufficient to obtain high-quality data-based applications. Otherwise, adaptive selecting the collected data based on previously collected data, the amount of data collected can be effectively reduced based on the application requirements, thereby improving the energy utilization of the network. The ACMC scheme also proposes a method for reducing the delay by increasing the duty cycle of the nodes that are far from the CC. At the same time, the transmission reliability of these nodes increases due to the increase in the transmission frequency. Thus, the ACMC scheme can also further reduce the delay of the network. The experimental results of the ACMC scheme in planar networks show better performance than the traditional data collection schemes and can improve the energy utilization of the network by 4.26%–6.68% while reducing the maximum delay by 9.4%.

INDEX TERMS Energy harvesting, energy utilization, matrix completion technique, data recovery, delay, adaptive collection.

I. INTRODUCTION

With the rapid development of communication and network technologies, novel information services and applications are rapidly growing worldwide [1]–[6]. The ubiquity of wireless sensor-equipped devices [7]–[9] can used to collect data at a low cost [10]–[12], establishes the foundation for numerous data-based applications, such as environmental monitoring [13], [14], smart transportation [15], [16], smart health monitoring [17]–[19], and industrial applications [20]–[22]. With the development of microelectronic technology, wireless sensor devices have become increasingly smaller, and their function has become increasingly powerful, which has led to extensive application and deployment [23]–[27]. According to [1], sensor-based devices connected to the Internet of Things (IoTs) have reached 9 billion and will reach 24 billion by 2020 [28]. Although advanced techniques

have extensively improved user's quality of experience [29], they are not adequate to meet the various requirements of seamless wide-area coverage, high-capacity hot-spots, low-power massive-connections, low-latency and high-reliability, as well as other scenarios [30]. Therefore, it is a great challenge to develop smart communications and networks that support adaptive adjustment technique [31]. Wireless sensor networks (WSNs) are an important part of the Internet of Things and can collect data at a lower cost [32]–[34]. The sensor nodes are small, easy to deploy, and can be deployed in complex environments. However, because the sensor terminal is small, the battery capacity is also limited [35], [36]. Therefore, the durable problem of WSNs has attracted widespread attention from academia and industry. Wireless energy harvesting and transfer technology is a way to effectively solve this issue [36]. The energy-harvesting wireless sensor network (EHWSN) is a network whose sensor nodes can absorb and replenish energy from the surrounding environment (solar, wind, thermal, etc.) [36]. Therefore, EHWSN can be applied to scenarios that require long-term monitoring [36]. There have been many studies that have reduced costs and improved energy efficiency. However, there are still two key issues that are not well resolved in EHWSN.

One of the important issues is how to improve the energy utilization of the network [37]. In the EHWSN, the energy provided by the surrounding environment is not stable. For example, the energy generated by solar radiation during the daytime is much greater than the energy generated at night. Therefore, the available energy of the sensor node in EHWSN is different at each moment. When the energy is sufficient, if only a small amount of energy is consumed, considerable energy will remain, and when the energy stored in the battery is full, it will waste considerable energy. In the case of insufficient energy, if considerable energy is consumed, the energy stored in the battery will be used. When this state continues for a long time, there is a possibility that the energy in the battery will be completely consumed. Therefore, reasonably allocating available energy by appropriately increasing or decreasing the energy consumption at different times is critical to improving the energy utilization of the network.

The second important issue is the delay in collecting data [13], [38]. In the data collection network, sufficient data needs to be collected. Collecting sufficient data will result in a longer delay. Due to the characteristics of the wireless transmission link, data packets may be lost. If a data packet is lost, the data packet needs to be retransmitted, which also causes a long delay. In some monitoring scenarios, such as forest fires, a long delay will postpone the timing of obtaining information, thus postponing the timing of firefighting causing greater losses. Moreover, some sensor nodes will be burned by the fire, causing the data transmission path to be cut off. Therefore, it is also important that a system can make a decision when the data are incomplete, that is, it can tolerate a portion of missing data. However, tolerating the lack of data can reduce the amount of data collected, which reduces the congestion of network transmission and accelerates network transmission. Therefore, reducing the transmission delay and tolerating the collection of missing data has important research significance.

To overcome these challenges, an adaptive collection scheme-based matrix completion (ACMC) is proposed to reduce the delay of network transmission and improve the energy utilization of the network. The main innovations of our work are as follows:

(1) In the ACMC scheme, only part of the data needs to be collected, and the remaining data are recovered by matrix completion technology so that all the data can still be obtained while reducing the amount of data collected. Since the collected environmental data have a strong correlation, the part of the data that is not collected can be recovered by matrix completion technology. Compared to the traditional data collection scheme, the amount of data that needs to be collected is reduced, thus reducing the energy consumption required to collect the data.

(2) A data collection scheme that varies with available energy is proposed in this paper, in which more data is collected under the condition that the available energy is sufficient and less data is collected when the available energy is insufficient, thereby improving energy utilization. In EHWSN, the sensor node can absorb solar radiation and convert it into usable energy. Therefore, when solar radiation is sufficient during the daytime, the available energy of the sensor node is very large, and the battery capacity of the sensor is limited. When energy can no longer be stored, considerable energy is wasted. There is no solar radiation at night so that only the energy stored in the battery can be used, and how long this situation can last cannot be accurately predicted. Therefore, the ACMC scheme adjusts the data collection scheme according to the residual energy and solar radiation in the battery. When the energy of solar radiation conversion is sufficient, more data can be collected. When the energy converted by solar radiation is very low, and there is insufficient energy in the node battery, an adaptive collection algorithm is used to collect as little data as possible while ensuring low recovery error. The scheme of this paper is different from the previous scheme; in the ACMC scheme, the data collected each time is dynamically changed, and the amount of data collected each time may be different so that the energy in the nodes at different times is fully utilized. At the same time, because as little data as possible is collected in the absence of solar radiation, it also ensures that the energy of the battery in the node is not exhausted.

(3) A way to adjust the duty cycle is proposed. The maximum delay of the network is reduced, and the energy utilization of the network is improved by adjusting the duty cycle of each node in the network.

(4) The ACMC scheme is applied to common planar networks. Experimental research and theoretical evidence show that compared with traditional scheme-based matrix completion, the energy utilization of the ACMC is increased

by 4.26%-6.68%, and the maximum delay of the network is reduced by 9.2%.

The rest of this paper is organized as follows. Section II reviews the related work. The problem description and system model are presented in Section III. In Section IV, we propose an adaptive collection scheme in detail. The theoretical analysis and simulation results are presented in Section V. Section VI concludes the paper.

II. BACKGROUND AND RELATED WORK

In the energy-harvesting wireless sensor network, which differs from traditional sensor networks in that the sensor nodes in the EHWSN can extract energy from the surrounding environment, one of the important issues is how to improve the energy utilization of the network. However, the energy provided in the environment is unstable, and there is little or no available energy at some times. Therefore, reducing the energy consumption of the network in EHWSN is also a problem. There have been many studies on how to reduce the energy consumption of network transmission [10], [39]. In [39], each node transmits a data packet in one transmission cycle, but the size of the data packet is different. By adjusting the modulation rate to ensure that one data packet is transmitted in one cycle when the size of the data packet transmitted by the network is fixed, the duty cycle can be adjusted to change the modulation rate, thereby reducing the energy consumption of the transmission.

Closely related to reducing network energy consumption is reducing the amount of data collected. Matrix completion technology is one of two classic sparsity representation techniques that can recover a complete matrix from a partial low-rank matrix. There has been considerable research on wireless sensor networks [40], [41] in which matrix completion technology is used. Recovery error is a very important issue. A row or column in the matrix without any data will result in a large number of recovery errors [42]. The more data collected, the lower the probability of empty rows or columns in the matrix. Candès and Tao [42] presented the minimum amount of data to be collected using the matrix completion technique in the traditional low-rank matrix and proved that this number is essentially related to the coherence of the data in the matrix. Based on this minimum number of required collections, [42] also demonstrates that in the case of the Bernoulli distribution (or uniform distribution), the error of the recovered data is acceptable. However, in the Bernoulli distribution, too little data collected will rapid increase the probability of empty rows or columns in the matrix, resulting in a rapid increase in recovery error. Therefore, in [40], a sampling method of the cross-sampling model is proposed. By collecting data at intervals in each sensor node to ensure that there is data collected in each row and column, the amount of data that needs to be collected can be greatly reduced, and a low recovery error can be maintained. However, the recovery error can also be reduced by other means. Cheng *et al.* [41] proposed an STCDG data collection algorithm that recovers the columns of the collected data by using matrix completion and then inserts the columns that have not collected data into the recovery matrix. Since the data matrix has short-term stability, it can be directly obtained by the method of semidefinite programming, and the error of recovering data is also maintained.

Because the battery can be recharged, it is meaningless to simply reduce the energy consumption of the network, so the main goal in the EHWSN is to improve the energy utilization of nodes in the network [43]. However, as in traditional wireless sensor networks, EHWSN also has the problem of imbalancing energy consumption because the energy consumption of each node in the network is not uniform, resulting in low energy utilization of some nodes, thus reducing the energy utilization of the entire network. In traditional sensor networks, many schemes [44]–[47] have been proposed to balance the energy consumption of nodes in the network. In [44], a LEACH scheme was proposed. By specifying the selected node as the cluster head (CH), the CH collects the data in its own cluster for aggregation and then sends the aggregated data packet directly to the sink or the next CH node, thereby balancing the energy consumption of the network. However, this method requires the user to define the expected probability of each node becoming a CH. Therefore, there is a LEACH scheme based on a genetic algorithm (GA) [45], which directly uses the GA to select a better CH, but the CH is randomly selected, and the remaining energy of each node is also not considered in the process of selecting the CH. In [46], according to the distance of the node from the sink, the nodes are divided into different gradients. The nodes near the sink have a low gradient, and the nodes far from the sink have a high gradient. A cluster with a low gradient has a small cluster, and a node with a high gradient has a large cluster, so the number of clusters in the near-sink region will be larger than that in the far sink region. Therefore, there are more CHs in the near-sink area. Each CH collects the data of the nodes in the cluster. After aggregation, the data is transmitted to the sink according to the principle of gradient descent. This scheme has many CHs in the near-sink area and relatively few CHs in the far-sink area, so the energy balance of the CHs can be well-maintained, thereby alleviating the energy imbalance of the entire network and improving the energy utilization of the network.

Another issue that is closely related to this paper is delay [48]. The delay of network transmission has always been one of the research priorities in WSN, and many factors affect the transmission delay. First, the reliability of the transmission is one of the important factors affecting the delay. When the reliability of the transmission is high, the number of required retransmissions requires decreases. Generally, the method of improving the transmission reliability is to increase the transmission power of the sensor node, thereby increasing the signal-to-noise ratio of the transmission channel so that the received data signal of the receiver is enhanced. If the transmit power of the node is fixed, the reliability of the data transmission is fixed. At this time, it is necessary to adopt some mechanisms to ensure the reliability

of transmission. The commonly used guarantee mechanism is the retransmission mechanism. In the retransmission mechanism, after the sender sends the data, it waits for the receiver to return the ACK signal of the received data. When the sender receives the ACK signal, it starts the transmission of the next data. Otherwise, when the sender waits for more than a certain threshold, the packet is resent until the ACK signal is received or the number of resends exceeds a certain threshold. This process generates a large number of delays. This paper uses a method to improve the reliability of the transmission and reduce the number of retransmissions. Since there is a surplus of energy away from the sink area, this part of the residual energy can be used to improve reliability. Therefore, reducing the delay away from the sink area node, which is the maximum delay of the network, also improves the energy utilization of the network.

III. THE SYSTEM MODEL AND PROBLEM STATEMENT

A. THE NETWORK MODEL

The network model used in this paper is a typical data collection network model, similar to [7] and [13]; its characteristics are as follows:

There is a total of n sensor nodes of the same type in the network, which are randomly and uniformly deployed in a circular area with radius R, and the sink is located at the center of the circle. The communication radius of each sensor node is r. The routing algorithm in the network is the shortest routing algorithm, the number of hops that each node requires to reach the sink can be obtained. The nodes are layered according to the number of hops, and the node with hop count 1 belongs to the first layer.

B. THE ENERGY CONSUMPTION MODEL

The energy consumption model used in this paper is similar to [39]. The sensor nodes transmit a data packet in a transmission cycle. The transmission cycle consists of three parts, the working time T_{on}, the sleep time T_{stby}, and the transition time (work-sleep transition) T_{start}, so the transmission cycle can be expressed as:

$$T = T_{start} + T_{on} + T_{stby}$$

Correspondingly, its power can also be divided into three parts: working power P_{on}, sleep power P_{stby}, and transition power P_{start}. The power is small in the sleep state, so it can be considered as 0. The power of the transition state mainly comes from the frequency synthesizer, so $P_{start} = P_{syn}$. The working power P_{on} is transmitted power P_{Tx}, amplifier power P_{PA}, and circuit power $P_{circuit}$.

The amplifier power can be expressed as:

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

where η represents the drain efficiency of the amplifier, ξ is the peak-to-average ratio, and ξ is related to the modulation technique. The modulation technique used in this paper is M-QAM, and the constellation size is M, so the peak-to-average ratio is expressed as $\xi = 3\frac{\sqrt{M}-1}{\sqrt{M}+1}$.

Therefore, the energy consumption of transmitting a packet can be expressed as:

$$\begin{aligned} E_{total} &= P_{on}T_{on} + P_{start}T_{start} \\ &= (1+\beta)P_{Tx}T_{on} + P_{circuit}T_{on} + 2P_{syn}T_{start} \end{aligned} \quad (1)$$

The transmit power can be obtained from the receiver power, according to the equation of the k^{th} path loss model [39], and the transmit power can be expressed as:

$$P_{Tx} = P_{rx}G_d \quad (2)$$

where $G_d = G_1 d^k M_1$, G_1 is the gain factor of 1 m, M_1 is the link margin, d is the communication radius of the node, k is the exponent order, usually in the range 2-4, and in this paper, $k = 3$ is selected.

The receiver power P_{rx} is related to the signal-to-noise ratio (SNR).

$$P_{rx} = 2BN_f\sigma^2 \cdot SNR \quad (3)$$

The signal-to-noise ratio is related to the error probability. In the M-QAM modulation technique, the relationship between the signal-to-noise ratio and the error probability is as follows:

$$P_e = \frac{4}{b}(1 - \frac{1}{\sqrt{M}})e^{(-\frac{3}{M-1})\frac{SNR}{2}} \quad (4)$$

Therefore, it is necessary to determine the error probability to obtain the signal-to-noise ratio, obtain the signal-to-noise ratio to obtain the received power, and obtain the working power of each node to transmit a data packet. Finally, the energy consumption of a transmit packet can be obtained.

C. THE MATRIX COMPLETION MODEL

The matrix completion technique used in this paper is similar to Ref. [49]. Matrix completion technology is an emerging technology that can recover all the data in the low-rank matrix with less data. For a matrix X of size $N_1 \times N_2$, considering that the set of known data is Ω, the data collection process $P_\Omega(X)$ is defined as:

$$P_\Omega(X) = \begin{cases} X_{i,j} & (i,j) \in \Omega \\ 0 & otherwise \end{cases}$$

When the known dataset already contains sufficient information, there is a specific matrix with rank r that is consistent with the known data, so data recovery can be performed by minimizing the rank of the matrix.

$$\begin{aligned} &\min(rank(X)) \\ &\text{s.t. } P_\Omega(X) = P_\Omega(M) \end{aligned}$$

where rank(.) represents the rank of the matrix.

However, the problem of minimizing the rank of a matrix is an NP-hard problem. Therefore, the problem of minimizing

rank can be transformed into a nuclear norm optimization problem as in Ref. [49].

$$\begin{aligned} &\min \ \|X\|_* \\ &\text{s.t. } P_\Omega(X) = P_\Omega(M) \end{aligned} \tag{5}$$

However, when the matrix is large, the traditional method of solving the nuclear norm optimization problem is difficult, so we use a singular value thresholding algorithm (SVT) [49] to solve the problem of minimizing the nuclear norm optimization problem.

The SVT algorithm is one of the algorithms for solving matrix completion, which solves the equivalence problem of Eq. (5):

$$\begin{aligned} &\min \ \tau \|X\|_* + \frac{1}{2}\|X\|_F^2 \\ &\text{s.t. } P_\Omega(X) = P_\Omega(M) \end{aligned} \tag{6}$$

The SVT algorithm is essentially a Lagrange multiplier method, and the Lagrange function of Eq. (6) is:

$$\mathcal{L}(X, Y) = \tau \|X\|_* + \frac{1}{2}\|X\|_F^2 + <Y, P_\Omega(X) - P_\Omega(M)>$$

where $<X, Y>$ represents the standard inner product between two matrices.

Reference [49] proved that the original problem of this Lagrange function is the same as the optimal value of its Lagrange dual problem; the following iterative formula can be obtained by using the convex optimization theory and the subgradient method for the Lagrange dual problem.

$$\begin{cases} \mathcal{L}\left(X^k, Y^{k-1}\right) = \min\limits_X \mathcal{L}\left(X, Y^{k-1}\right) \\ Y^k = Y^{k-1} + \delta_k P_\Omega\left(M - X^k\right) \end{cases} \tag{7}$$

Finally, according to [46], formula (7) can be converted into:

$$\begin{cases} X^k = D_\tau(Y^{k-1}) \\ Y^k = Y^{k-1} + \delta_k P_\Omega\left(M - X^k\right) \end{cases}$$

where D_τ is a singular value contraction operator and τ is the singular value threshold.

D. PROBLEM STATEMENT

The main purpose of this paper is to design an efficient algorithm for adaptive data collection to improve the energy utilization of EHWSN. The scheme has the following characteristics:

(1) Improving the energy utilization of the network. In EHWSN, since the sensor node is rechargeable, there is no obvious significance in reducing the energy consumption of the data transmitted by the node. The energy of the nodes during the daytime is sufficient, so many data can be collected during the daytime, but the storage capacity of the battery is limited, and there is no solar radiation at night; thus, only the energy stored in the battery can be used, so less energy is available at. Therefore, to avoid wasting energy, the main goal is to improve the energy utilization in EHWSN. The energy utilization is the ratio of the available energy to total energy consumption, as follows:

$$\max(\eta) = \max\left(\sum_{i=1}^n E_i / \sum_{i=1}^n E_{ava}^i\right) \tag{8}$$

where η represents the energy utilization of the network, E_i represents the actual energy consumed of node i, and E_{ava}^i represents the available energy at node i.

(2) Reducing the network transmission delay. In a planar network, the energy consumption of the nodes in each layer in the network is uneven, and the energy consumption of the nodes far away from the sink is much lower than that of the nodes near the sink, which will seriously affect the energy utilization of the network. Therefore, the duty cycle of the nodes far away from the sink can be increased, thereby increasing the energy utilization of the network and reducing the delay of network transmission. Considering that the duty cycle of node i is τ_i, the second optimization goal is:

$$\min(\mathbb{C}) = \min \max \sum_{i \in S} \delta(\tau_i) \tag{9}$$

where S is the data transmission path and δ is the single-hop delay of the node.

In summary, our overall optimization goal is as follows:

$$\begin{cases} \max(\eta) = \max\left(\sum_{i=1}^n E_i / \sum_{i=1}^n E_{ava}^i\right) \\ \min(\mathbb{C}) = \min \max \sum_{i \in S} \delta(\tau_i) \end{cases} \tag{10}$$

TABLE 1. Network parameters.

Parameter	Value	Value
T	Transmission period	200 ms
T_{start}	The time of transient mode	$5\ 10^{-6}$s
η	The power of 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	$3.981\ 10^{-21}$
B	The channel bandwidth	10^4
L	The number of bits in a packet	10^3
N_f	The receiver noise figure	10
b	The modulation rate	5
P_e	The error probability	$2\ 10^{-4}$

IV. THE DESIGN OF ACMC SCHEME

A. RESEARCH MOTIVATION

Matrix completion techniques have evolved for a long time and can recover all the data from a low-rank matrix with missing data. In sensor networks, some high-coherence data

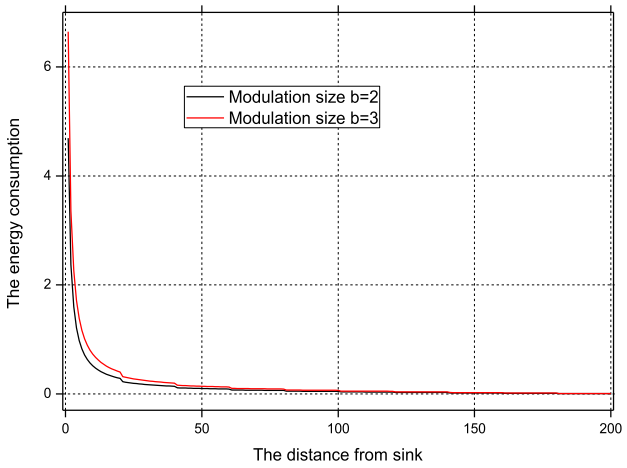


FIGURE 1. The energy consumption at different locations in the planer network.

are collected, so there have been some studies investigating the application of matrix completion techniques on sensor networks [40], [46]. The data collection scheme proposed in this paper differs from the previous scheme as follows:

(1) In the previous scheme, because the energy of the nodes in the network is limited, to extend the network lifetime, the purpose of the previous scheme was to reduce energy consumption by reducing the amount of data collected. In [46], due to the large coherence between some data, in each collection round, some sensor data is randomly collected according to a certain ratio. After the collection is completed, all data can be recovered through matrix completion technology. However, in this scheme, the sampling rate is fixed, that is, the amount of data collected per round is fixed, so the energy consumption will also fluctuate within a range (the position of the sensor node also affects the energy consumption). Therefore, in EHWSN, the sampling rate can only be selected according to the lowest available energy, so that a large amount of energy is wasted when there is considerable available energy.

(2) Second, the energy consumption of the planar network is unbalanced, there also lead to low energy utilization of the network. Fig. 1 shows the energy consumption of nodes at different distances from the sink. Obviously, the energy consumption of the nodes near the sink is much larger than that of the nodes far from the sink. When the energy of the nodes near the sink is exhausted, the nodes of the far sink will have a large of energy remaining.

Thus, designing a suitable scheme to ensure the energy utilization of the entire network is a challenging issue. In this paper, we propose an adaptive data collection scheme to solve this problem. It can effectively improve the energy utilization of EHWSN and appropriately alleviate the energy consumption imbalance in each node of the network. The following section will elaborate on this scheme.

B. OVERVIEW OF THE PROPOSED SCHEME

The main idea of this scheme is to use an adaptive sampling algorithm to adjust the sampling scheme of the current time slot in EHWSN to improve the performance of the network. When the energy is sufficient, more data is collected, and when the energy is insufficient, less data is collected. The scheme is divided into the following parts:

(1) **The available energy.** In EHWSN, since solar radiation is sufficient during the daytime, the energy converted from solar radiation can be directly utilized as available energy. At night, there is no solar radiation, so the node can only use the previously stored energy. Since it is not possible to accurately estimate how long the low solar radiation will last, a method of instantly allocating available energy is employed. In planar networks, the energy consumption of each node is different, and the energy consumption of the nodes closest to the sink is the largest. When the energy of these nodes is not exhausted, the energy of the other nodes in the network will not be exhausted. Therefore, we only need to consider the available energy of the node with the most energy consumption. Because only $\lfloor N/2 \rfloor$ data must be collected at most according to the cross-sampling principle, the energy consumption of collecting $\lfloor N/2 \rfloor$ data is taken as the threshold C_1. When the energy converted by solar radiation is higher than the threshold, the available energy is the energy converted by solar radiation. When it is lower than the threshold, since the energy of C_1 is consumed at most, the available energy is C_1. Second, since the energy consumption of the adaptive collection algorithm is unknown, to improve the network energy utilization, an overflow threshold α is also set to determine whether the energy stored in the battery is sufficient. If the result of the determination is that the energy is sufficient, we collect the data of all nodes; otherwise, the adaptive collection algorithm is adopted.

(2) **The adaptive collection algorithm**. In the above steps, we can obtain the available energy of each slot of EHWSN. Obviously, to improve the energy utilization of the network, when the energy is sufficient during the daytime, it is possible to collect more data. At night, there is less available energy, and only the energy stored in the battery can be used. It is impossible to accurately determine how long the night lasts, so as little energy as possible needs to be consumed at night. Therefore, there is less data collected at night. When there is no data in a row or a column of the matrix, the data error recovered by the matrix completion technique is very large. To avoid this phenomenon, the data is collected according to the cross-sampling principle [40].

The cross-sampling principle ensures that there is no data in adjacent rows and columns. As shown in Fig. 2, the column of the matrix can be divided into two areas. When the sensor node is in the white area, it is possible to collect data only when the slot is odd. When it is in the gray area, it is possible to collect data only when the slot is even so that collecting adjacent data can be avoided, and when the amount of collected data is small, the probability of the occurrence of empty lines or columns in the matrix is reduced.

Algorithm 1 The Scheme for Allocating Available Energy

Initialize: Considering that the battery capacity of each node is E_{cap}, a total of T_s slots are collected, the solar panel area is χ, and the remaining energy in the battery is E_{res}.
Input: the initial energy E_{init} for each node and the solar radiation F_i for each slot.
Output: the available energy E_{ai}
// C_1 is the maximum energy consumption for collecting
// $\lfloor N/2 \rfloor$ data according to the principle of cross-sampling

```
E_res = E_init
For i = 1 to T_s
    IF E_res ≥ α · E_cap or χF_i ≥ C_1
        IF χF_i ≥ C_1
            E_ai = χF_i
        Else
            E_ai = C_2
    // C_2 is the maximum energy consumption for
    // collecting all data, which can be calculated
    // according to Theorem 2
        End if
    Else
        E_ai = C_1
    End if
    E_res = E_res − E_u
    // E_u is the actual energy consumed and obtained by
    // Algorithm 2
    IF E_res > E_cap
        E_res = E_cap
    End if
End for
```

(a) Learning process of adaptive sampling algorithm: As increasingly more data is collected as the number of slots increases, the size of the matrix that needs to be recovered is also increasingly larger, and it will take longer for each recovery. To solve this problem, we use matrix completion technology for a small matrix. This smaller matrix is called a sliding window (as shown in the blue area in Fig. 2), and the size of the sliding window is fixed. When a new time slot is added, the oldest time slot is removed from the window. If the newly added slot is the current time slot, the sliding window is called the current sliding window.

The sliding window size used is $N \times T$. Since the learning process of the adaptive sampling algorithm occurs during the daytime, the energy of each node is sufficient, and as much data as possible can be collected. In the learning process, part of the collected data at the current slot is recovered using matrix completion technology in the sliding window. Comparing the recovered data with the collected data can cause a reconstruction error ε, then the actual reconstruction error ε is compared with the set error threshold ε_b. When $\varepsilon < \varepsilon_b$, the amount of some known data in the sliding window is reduced, and the sliding window is recovered until the known amount of data has been reduced to the minimum

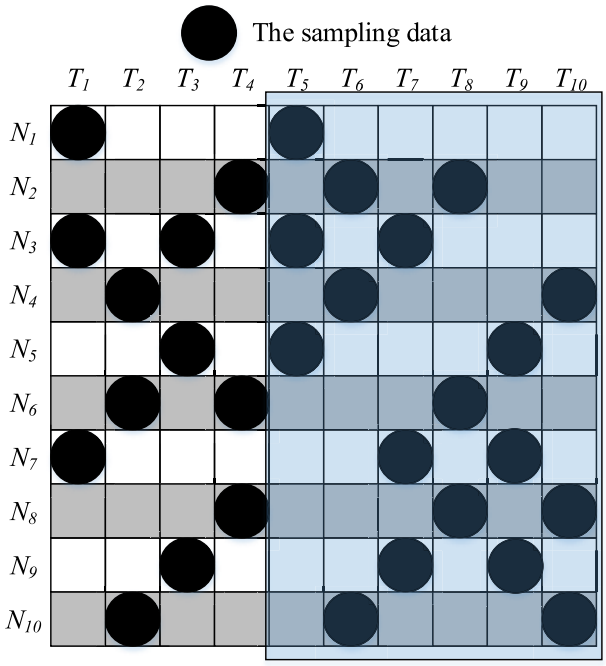


FIGURE 2. Cross-sampling model schematic diagram.

required, or the actual reconstruction error just meets $\varepsilon \geq \varepsilon_b$. Finally, the amount of data is recorded, and the initial amount of data collected by the adaptive sampling algorithm at night can be obtained.

(b) Collection process of the adaptive sampling algorithm: Since the rank differences of two adjacent matrices do not exceed 1 (see Theorem 1), the current initial collection of slots can be obtained from the previous window. Obviously, when the ranks of two adjacent matrices are the same, the amount of data collected by the current slot is the same as the first column of the previous window. When the ranks of two adjacent matrices are different, the amount of data collected by the current slot is the same as the last column of the previous window [38]. However, the rank of the matrix in the current sliding window is difficult to calculate, and in the data collection process, if the amount of data collected is insufficient, it continues to collect, so the minimum value of the first column and the last column of the previous window is used as the initial number of the current slot. Therefore, data is collected according to the initial amount, and the current sliding window is recovered by matrix completion to obtain the current reconstruction error ε. The current reconstruction error is compared with the error threshold ε_b.

When $\varepsilon \leq \varepsilon_b$, the data collection of the current slot ends. Otherwise, the amount of collected data is increased, and the comparison error is recovered again until the reconstruction error satisfies the requirement.

In Fig. 3(a), the entire process of the adaptive collection algorithm is shown. In the upper part of Fig. 3(a), since the collected data matrix has not reached the size of the sliding window at the beginning of data collection, the

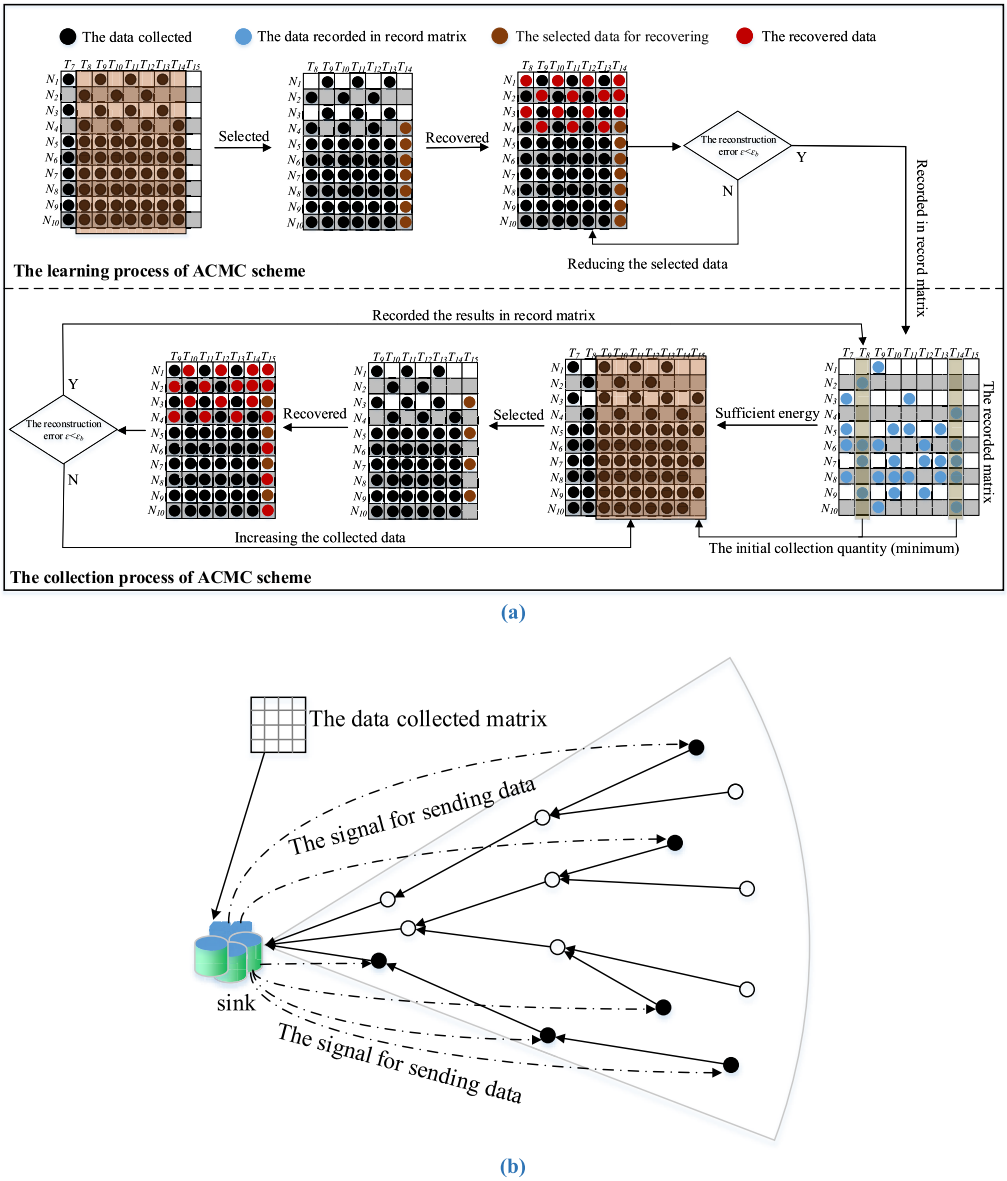


FIGURE 3. (a) The adaptive collection algorithm schematic diagram. (b) The data collection schematic diagram.

network cannot implement the collection process of the adaptive collection algorithm before the data collection matrix reaches the size of the sliding window. In Fig. 3(a), the size of the selected sliding window is 10×7, so the size of the sliding window does not meet the requirement and cannot be learned before T_7. When the current time slot is T_7, the sliding window is completely filled, so the learning process can be started. Reducing the amount of data at the current slot in the sliding window, the matrix completion technology is used to recover data, and the number is adjusted according to its reconstruction error. This process continues until the reconstruction error meets the requirements. Finally, the amount of data at the current slot in the window is stored in the record matrix. Thus, the entire learning process is completed, and a record matrix that records the minimum amount of data that needs to be collected in each slot can be obtained.

The learning process takes place when the available energy is sufficient, and the data must be collected before learning begins. When the available energy to each node is reduced, and it is unable to support the collection of large amounts of data, the adaptive collection process can begin. The adaptive collection process is shown in the lower half of Fig. 3(a). The initial number of adaptive sampled data is obtained from the first and last columns of the previous window. For example, when the current time slot is T_{15}, the previous window is T_8 to T_{14}, so the first column is T_8, the number to be collected is 4, the last column is T_{14}, and the number to be collected is 5. Because the data is not retrievable once collected, the initial number takes the smallest of the two columns, that is, 4,

and then broadcasts the corresponding node to send data to the sink, as shown in 3(b). Put the collected data into the current window for data recovery and continue to compare the reconstruction error; if the error is too large, increase the amount of data collected, looping to the reconstruction error to meet the requirements, and finally store the collection of this slot in the record matrix for later data collection. This method ensures that as little data is collected as possible with less available energy.

(3) **Adjusting the node duty cycle**. In the process of adaptive sampling, if all nodes adopt the same duty cycle, the data sent by the node of the far sink must pass through the node of the near sink, so the energy consumption of the nodes near the sink is still much higher than the nodes of the far sink. In EHWSN, the deployment positions of the nodes are not far apart, so that the solar radiation received by the nodes in the same network is considered the same, so the available energy is the same. Therefore, the energy consumption of the nodes in the far sink is small, which results in low energy utilization of the network. Therefore, to alleviate this phenomenon, different duty cycles can be used for nodes in different regions, thereby improving the energy utilization of the far sink node and simultaneously reducing the maximum delay of network transmission.

C. EMPIRICAL RESEARCH ON ACTUAL DATA

The data studied was from Intel Berkeley Research Lab [50], and they arranged 54 sensor nodes, each of which transmits the collected temperature, humidity, light intensity and other data to the sink at intervals of 30 seconds. Due to some data loss during the collection of this dataset, we extracted a round of data energy hours from April 28, 2004, and requested that these data be collected with at least one in the same slot. A total of 10 days of data was used. Since the No. 5 sensor and the No. 8 sensor had almost no data collected, we removed the data collected by these two sensors. Therefore, the final collection matrix size was 52×240, with a total of 52 sensor nodes, collecting 240 time slot data.

1) LOW-RANK FEATURE

The collected temperature and humidity data in different locations have strong coherence with different time slots. First, the singular value decomposition (SVD) method will be used to determine whether the matrix composed of temperature and humidity data has a low rank. Matrix $X_{N\times T}$ can be represented as follows:

$$X = U\Sigma V^T$$

where U, V, Σ are the three matrices after the matrix X is decomposed, U is a matrix of size $N \times N$, V is a matrix of size $T \times T$, and Σ is a diagonal matrix of size $N \times T$. The value on the diagonal of the diagonal matrix is the singular value, and the singular value decreases as the number increases ($\Sigma = diag(\sigma_1, \sigma_2, \cdots \sigma_r, 0, \cdots, 0)$). The rank of the matrix is r, and the number of nonzero singular values is equal to r. A matrix with low rank needs to be satisfied $r \ll \min(N, T)$.

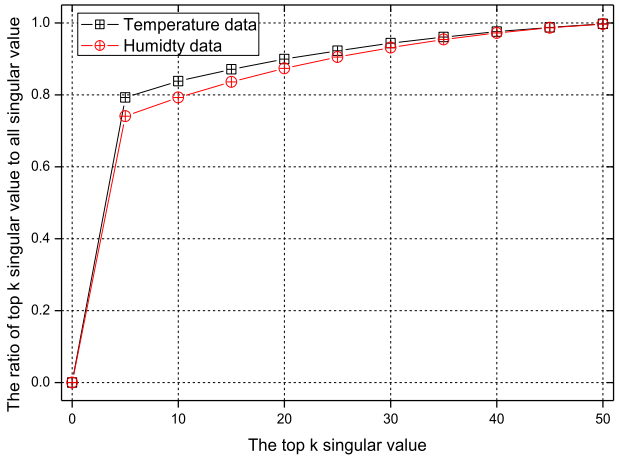


FIGURE 4. The low-rank feature of the matrix.

The value of the singular value σ_i can represent the proportion of the i^{th} part in the entire matrix. When the proportion of the first k singular values is close to the total weight of the matrix ($\sum_{i=1}^{k} \sigma_i^2 \approx \sum_{i=1}^{r} \sigma_i^2$), the matrix is considered to have low rank. Therefore, the proportion of the first k singular values in the total singular value can be obtained according to the following formula.

$$g(k) = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2} \tag{11}$$

Fig. 4 is the top k singular value ratio graph. The proportion of the first singular value is already close to 1. This result indicates that the collected data matrix has a good low-rank feature. The matrix of the property can also be recovered by matrix completion technology.

2) STABLE FEATURE

Temperature and humidity data usually change little over time. The data matrix can be determined to be stable according to the difference between the adjacent time slots. The difference between two consecutive time slots is expressed by:

$$gap(i, j) = \left| x_{i,j} - x_{i,j-1} \right|$$

where i represents the number of sensor nodes, j represents the time slot, and the smaller the $gap(i, j)$, the better the stability of the data around the time slot j.

We normalize the difference of consecutive time slots as follows:

$$\Delta gap(i, j) = \frac{gap(i, j)}{\max\limits_{1 \le i \le, 2 \le j < T} gap(i, j)} \tag{12}$$

Fig. 5 is the difference between adjacent slots according to Eq. (12), the x-axis represents the difference between two normalized time slots, and the y-axis represents the cumulative distribution function (CDF). It can be seen that more than 90% of the temperature data $\Delta gap(i, j)$ is less than 0.05, but the difference between the humidity data is larger when

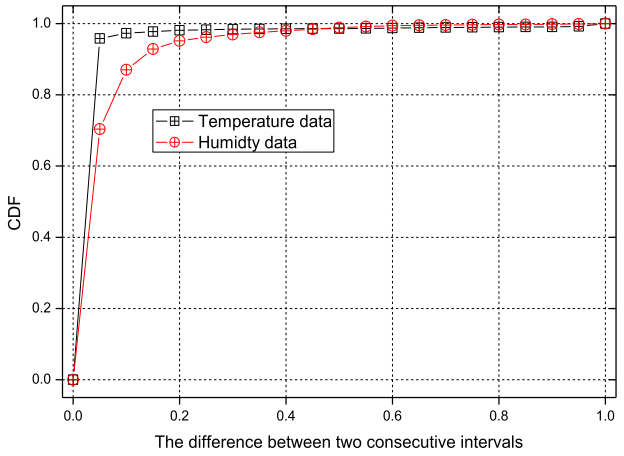


FIGURE 5. The stable feature of the matrix.

the difference after normalization is less than 0.8 and only covers 90% of the humidity data. Therefore, the stability of the temperature data is better than the stability of the humidity data.

3) RANK STABLE FEATURE

Theorem 1: Considering A is a matrix of size $m \times n$, B is a matrix of size $m \times k$. Then, the ranks of the matrices (A, B) where A and B are connected satisfy the following:

$$\begin{cases} \text{rank}\,(A, B) \geq \max\,(\text{rank}\,(A)\,, \text{rank}\,(\text{B})) \\ \text{rank}\,(A, B) \leq \text{rank}\,(\text{A}) + \text{rank}(\text{B}) \end{cases}$$

Proof: Obviously, after the matrices A and B are connected, the maximal linearly independent group of matrices A and B does not change. Considering the rank of matrix A is less than the rank of matrix B, the maximal linearly independent group of matrix A has rank (A) vectors, and that of matrix B is rank (B). Since rank (B) vectors in matrix B are linearly independent, there are rank (B) vectors in matrix A are required to represent these vectors. Therefore, the minimum rank of the matrix [A, B] is

$$\text{rank}([\text{A, B}]) \geq \max(\text{rank}\,(\text{A})\,, \text{rank}(\text{B}))$$

Considering the maximum value of the rank after the connection, the worst case after the connection is that the vector originally belonging to matrix A is obtained by the maximally linearly independent group of matrix A, and the vector originally belonging to matrix B is obtained by the maximally linearly independent group of matrix B, so the maximum value of rank is rank (A) + rank(B). ■

Theorem 2: Considering that the data collection matrix is X and the size of the sliding window is $N \times T$, the difference between the ranks of two adjacent sliding windows will not exceed 1.

$$|\text{rank}(X_{N\times T}\,(t-1)) - \text{rank}(X_{N\times T}\,(t))| \leq 1$$

where $X_{N\times T}\,(t)$ represents a sliding window in which the last column of the slot is t.

Proof: It is easy to see that the matrix $X_{N\times T}\,(t-1)$ and the matrix $X_{N\times T}\,(t)$ have a common part of size $N \times (T-1)$, and the matrix formed by this part is recorded as X_{pub}. The matrix $X_{N\times T}\,(t-1)$ has one more column than the matrix X_{pub}, which is equivalent to the matrix X_{pub} connected to a matrix with only one column to form $X_{N\times T}\,(t-1)$, so according to Theorem 1

$$\text{rank}\left(X_{pub}\right) \leq rank(X_{N\times T}\,(t-1)) \leq \text{rank}(X_{pub}+1) \quad (13)$$

Similarly, the relationship between the matrix $X_{N\times T}\,(t)$ and X_{pub} can also be obtained as follows:

$$\text{rank}\left(X_{pub}\right) \leq rank(X_{N\times T}\,(t)) \leq \text{rank}(X_{pub}+1) \quad (14)$$

Combined with (13) (14), can obtain

$$|\text{rank}(X_{N\times T}\,(t-1)) - \text{rank}(X_{N\times T}\,(t))| \leq 1$$

■

From Theorem 1 and Theorem 2, the rank difference of two adjacent sliding windows does not exceed 1, that is, the rank of adjacent matrices does not change too much, so using the data from the previous slot is meaningful.

D. THE CALCULATION OF ENERGY CONSUMPTION

In the previous section, the principle of adaptive sampling was briefly described, and it was also verified that the temperature data is low-rank and stable. In the ACMC scheme, the available energy in the current slot needs to be obtained first, and before the available energy is obtained, the two threshold values of C_1 and C_2 in Algorithm 1 need to be obtained.

In a planar network, a node deployed in a far-sink area must pass through a node near the sink area; that is, a node in the near-sink area must forward more data. Therefore, the energy consumption of the nodes in the near sink area is the fastest, and the available energy to each node in the same slot is equal; thus, it is only necessary to consider the amount of forwarding data of the near-sink node, and the amount of data forwarded by each node can be obtained by Theorem 3.

Theorem 3: In a planar network, the nodes are randomly and evenly distributed in a circular area of radius R, and the communication radius of each node is the same as r. In a slot, the amount of data forwarded by the node at a distance of l meters from the sink is:

$$D_l = \lambda_l + \left(1 + \frac{r}{l}\right)\lambda_{l+r} + \cdots + (1 + \frac{zr}{l})\lambda_{l+zr} \quad (15)$$

where z is a positive integer such that $l + zr$ is less than R, and λ represents the amount of data sent by each area of the network.

Proof: As shown in Fig. 6, considering that the node is located at area $\Theta_{l,k}$ that is a distance of l meters from the sink, the area $\Theta_{l,k}$ is an area surrounded by a circle of length d and two lines of angle θ. When the angle θ is small, the enclosed area $\Theta_{l,k}$ can be approximated as a rectangle, and the width

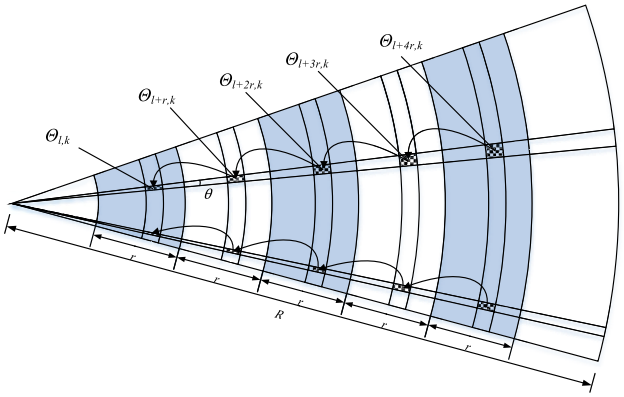


FIGURE 6. An illustration of node forwarding data quantity.

of the rectangle can be considered as $\alpha l\theta$, and α is a constant. Therefore, the area of $\Theta_{l,k}$ is

$$S_{l,k} = \alpha l\theta \cdot d$$

Therefore, the number of nodes in this area is as follows:

$$N_{l,k} = \alpha l\theta d \cdot \rho$$

where ρ is the node density.

Obviously, the nodes in the $\Theta_{l,k}$ region will forward the data generated by the $\Theta_{l+r,k}, \Theta_{l+2r,k}, \cdots, \Theta_{l+cr,k}$ region nodes, so that the total amount of data forwarded by the region $\Theta_{l,k}$ is

$$B_{l,k} = \alpha l\theta d\rho\lambda_l + \alpha\,(l+r)\,\theta d\rho\lambda_{l+r} + \cdots + \alpha(l+zr)\theta d\rho\lambda_{l+zr}$$

Thus, the amount of forwarding data of each node in the region $\Theta_{l,k}$ can be obtained as:

$$D_l = \frac{\alpha l\theta d\rho\lambda_l + \alpha(l+r)\theta d\rho\lambda_{l+r} + \cdots + \alpha(l+zr)\theta d\rho\lambda_{l+zr}}{\alpha l\theta d\rho} = \lambda_l + \left(1+\frac{r}{l}\right)\lambda_{l+r} + \cdots + (1+\frac{zr}{l})\lambda_{l+zr}$$

■

Theorem 3 gives the amount of data forwarded by each node in the network. According to Eqs. (1)(2)(3)(4), the energy consumption of a packet transmitted by the network can be obtained as:

$$E_b = \frac{4}{3}(1+\beta)N_f\sigma^2(M-1)\ln\left(\frac{4\left(1-\frac{1}{\sqrt{M}}\right)}{bP_e}\right)G_d BT_{on} + P_{circuit}T_{on} + 2P_{syn}T_{start} \quad (16)$$

Therefore, the energy consumption of the network can be obtained, so the maximum energy consumption C_1 of $\lfloor N/2 \rfloor$ data collected according to the cross-sampling principle in Algorithm 1 can be obtained. Thus, the available energy of each slot can be obtained. When the available energy of a node can collect more than $\lfloor N/2 \rfloor$ data, the available energy of the $\lfloor N/2 \rfloor$ data collected according to the cross-sampling principle still has remaining energy. Therefore, the data that has not been collected can continue to be collected. To balance the energy consumption of the network, the data of high-level nodes is preferentially collected until the available energy of the node is exhausted. When the available energy of the node cannot collect $\lfloor N/2 \rfloor$ data, the node has no energy supplement at this time, and the remaining energy is also determined to be insufficient, so the adaptive collection algorithm is used to collect as little data as possible to reduce the energy consumption. For the detailed process, see Algorithm 2.

Algorithm 2 The scheme for ACMC

Initialize: Considering the current i^{th} slot, the available energy for the current slot is E_{ai}, the maximum storage capacity of the battery is E_{cap}, and the energy consumption of transmitting a packet is E_b.
Input: the previous data collection matrix is $B = [\vec{B}_1^T, \vec{B}_2^T, \cdots, \vec{B}_{i-1}^T]$, and the recording matrix is $R = [\vec{R}_1^T, \vec{R}_2^T, \cdots, \vec{R}_{i-1}^T]$.
Output: the data collection vector $\vec{B}_i$, the record vector $\vec{R}_i$, and the actual consumed energy is E_u.

```
the nodes sorted by the distance from the sink, the farthest node number is 1.
IF E_ai > C_1
    The ⌊N/2⌋ data is collected according to the cross-sampling principle, and the collected result is assigned to the vector B_i
    x = C_1                    // x is the maximum energy consumption in network
    For j = 1 to N
        IF B_i(j) ≠ 0 // The j^th node has been collected
            continue;
        End if
        IF (d_j/d_N + 1)E_b + x ≤ E_ai // Eq. (15)
            x = x + (d_j/d_N + 1)E_b
            B_i(j) = 1  // Representing the data at node j can be collected
        End if
    End for
    Broadcasting according to B_i to collect data.
    R_i = learning(B)
    E_u = x
Else
    [E_u, B_i, R_i] = adapt(E_ai)
    Broadcasting according to B_i to collect data.
End if
```

The collected data of the current slot can be obtained by Algorithm 2, but in Algorithm 2, the adaptive sampling algorithm is not specifically given. Next, the specific form of the adaptive sampling algorithm will be given.

In the adaptive sampling algorithm, a sliding window must be defined; considering that the size of the sliding window is $N \times T$, this window is recovered by the SVT algorithm, and the number of data to be collected by the current slot is

adjusted according to the reconstruction error of the collected data so that as little data as possible is collected when the reconstruction error satisfies the requirement, and the reconstruction error can be calculated by:

$$\sigma = \frac{\sum_{i,j,W_{i,j}==1} (X_{i,j} - W_{i,j})^2}{\sum_{i,j,W_{i,j}==1} W_{i,j}^2} \tag{17}$$

where matrix X is the matrix after recovery, and matrix W is the collection matrix.

However, due to the large amount of available energy during the daytime, a large amount of data is collected, which causes the slot that just entered the night to collect only a small amount of data. When the slots in the sliding window are all at night, it may cause an empty line in the sliding window, which will seriously affect the accuracy of the recovery. Therefore, it is necessary to avoid this phenomenon.

Candès and Tao [42] prove that when the collected data obey the Bernoulli distribution (or uniform distribution), the probability that the matrix has an empty row or column is very low and can be ignored, so the minimum number of collections required is given as Theorem 4.

Theorem 4: In an $N \times T$ sliding window, when the probability of an empty row is less than that of a Bernoulli distribution, the minimum amount of data that needs to be collected per slot is

$$x \geq \frac{(N+2)(2\alpha - \alpha^2)}{2 + 4\alpha - 2\alpha^2} \tag{18}$$

where α is the sampling rate of the Bernoulli distribution, and N is the number of sensor nodes, that is, the number of matrix rows.

Proof: Considering event F is an event with an empty row in the matrix. When the sampling rate is α, the number of samples that can be obtained is $m = \alpha NT$. Under the Bernoulli distribution, the probability that each data in the matrix is collected is equal, so the probability that each data is collected is $\frac{m}{NT}$. Correspondingly, the probability of the data not being collected is $1 - \frac{m}{NT}$, so the probability of an empty row is

$$P_{Bernoulli}(F) = (1 - \frac{m}{NT})^T$$

Under the principle of cross-sampling, each slot will have maximum $\lceil N/2 \rceil$ data collected. Considering that x packets are collected, the total collection scheme has a total of $C^x_{\lceil N/2 \rceil}$ species (C represents the combination). When one of the datum must be collected, the total collection scheme has $C^{x-1}_{\lceil N/2 \rceil}$ species, so the probability that this data must be collected is $\frac{C^{x-1}_{\lceil N/2 \rceil}}{C^x_{\lceil N/2 \rceil}}$. Therefore, the probability that this data will not be collected is $1 - \frac{C^{x-1}_{\lceil N/2 \rceil}}{C^x_{\lceil N/2 \rceil}}$. In the cross-sampling principle, the slots that collect these sensors have $T/2$, so the probability of occurrence of event F is:

$$P_c(F) = (1 - \frac{C^{x-1}_{\lceil N/2 \rceil}}{C^x_{\lceil N/2 \rceil}})^{T/2}$$

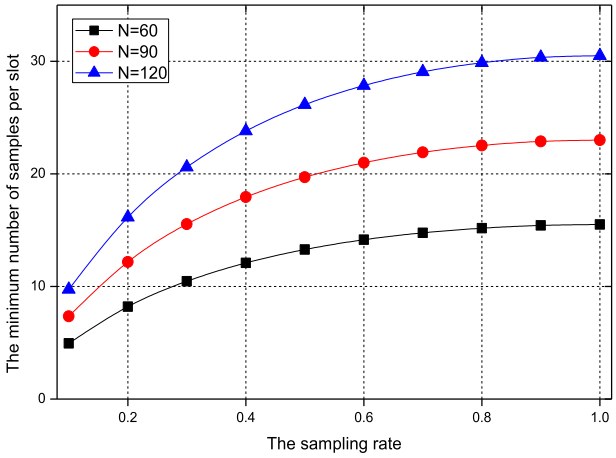


FIGURE 7. The minimum amount of data to be collected.

Therefore, it is necessary that the probability of an empty row is not greater than the Bernoulli distribution; that is, it needs to be satisfied.

$$P_c(F) \leq P_{Bernoulli}(F)$$

And that is:

$$(1 - \frac{C^{x-1}_{\lceil N/2 \rceil}}{C^x_{\lceil N/2 \rceil}})^{T/2} \leq (1 - \frac{m}{NT})^T$$

Reorganizing the above formula can obtain

$$x \geq \frac{(N+2)(2NTm - m^2)}{2(NT)^2 + 4NTm - 2m^2}$$

Substituting $m = \alpha NT$ into the above formula can obtain:

$$x \geq \frac{(N+2)(2\alpha - \alpha^2)}{2 + 4\alpha - 2\alpha^2}$$

■

From Theorem 4, the minimum amount of data that each slot needs to collect can be obtained. From (18), it is easy to see that when the probability of an empty row in the matrix is lower than that of the Bernoulli distribution, the minimum amount of data to be collected is only related to the sampling rate of the Bernoulli distribution and the number of sensor nodes (number of rows).

Fig. 7 shows the minimum amount of data to be collected when $T = 60$. It can be seen that the data to be collected increases with the increase of the Bernoulli distribution sampling rate, and as the number of nodes increases, the amount of data that needs to be collected will also increase. However, as shown in Fig. 8, the increase in the number of nodes does not affect the sampling rate under the principle of cross-sampling. The sampling rate of the cross-sampling principle is much lower than that of the Bernoulli distribution, where the probability of an empty row in the matrix is not greater than the Bernoulli distribution.

Therefore, the minimum amount of data that needs to be collected for adaptive sampling of each slot is obtained.

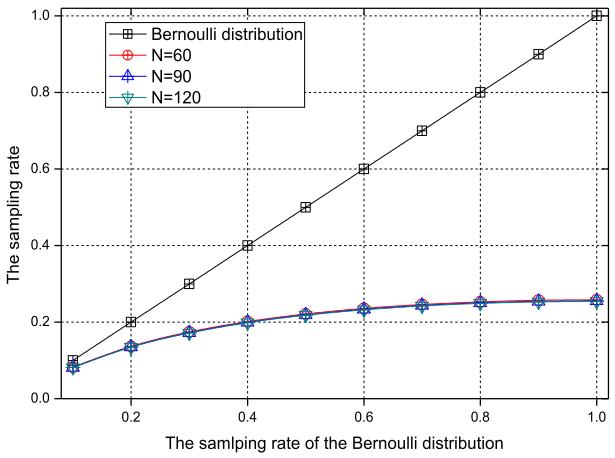


FIGURE 8. The sampling rate under the ACMC scheme.

Since the rank difference of the matrices in the adjacent two sliding windows does not exceed 1, the initial number of collections can be obtained from the previously collected slots. In the daytime, the energy is sufficient, there is no need to adopt the adaptive sampling algorithm, and the number of samples recorded will be large. Therefore, this number cannot be directly used for adaptive sampling, so after the current slot meets the size of the sliding window, an algorithm should be used to obtain the data that may be useful in the future. This process is called the learning process.

The learning process is the same as the idea of adaptive sampling. The amount of data that needs to be collected is adjusted based on the reconstruction error within the window. Obviously, increasing the amount of data collected will reduce the reconstruction error after recovery [36], and decreasing the amount of data collected will increase the reconstruction error after recovery. Therefore, the learning process of the ACMC scheme is Algorithm 3, and the collection process of the ACMC scheme is Algorithm 4.

Through the above adaptive collection algorithm, a complete data collection matrix is obtained, but the energy consumption of each node in the network is still unbalanced, resulting in considerable wasted energy. Therefore, the wasted energy is reduced by adjusting the duty cycle of each layer node. The duty cycle of each layer of the network is given Theorem 5.

Theorem 5: In planar networks, when the energy utilization of each node is the largest, the duty cycle of each node is as follows:

$$\begin{cases} max(\tau_i) \\ s.t\ E(\tau_i) \leq E(\tau_{opt}) \end{cases} \tag{19}$$

where $E(\tau_i)$ is obtained from Eq. (16), τ_i is the duty cycle of node i, and τ_{opt} is the duty cycle of the node that is closest to sink.

Proof: The node with the most energy consumption in the network must be located in the near sink area. In the adaptive collection algorithm, the node with the largest energy consumption in the network is not exhausted. Therefore, the nodes in other areas can increase the duty cycle. It is only necessary to ensure that the adjusted energy consumption does not exceed the near sink area to ensure that no energy nodes are present in the network. ■

By Theorem 5, the duty cycle of each layer of the network can be obtained so that the total energy consumption of the network can be obtained, and the total energy utilization of the network can be obtained. The duty cycle of the node is increased, and the delay of the network is also reduced. The delay of each layer node in the network is given in Theorem 6.

Theorem 6: In a planar network with a k layer, the transmission delay of the node of the i^{th} layer to the next layer is

$$\delta(\tau_i) = \frac{1}{1-(1-(1-P_e)^L)^{\tau_i TBb/L}} \tag{20}$$

Proof: The bit error rate that needs to be guaranteed for transmission is known, and the reliability of transmitting

Algorithm 3 The ACMC Scheme for Learning

Initialize: Considering the current i^{th} slot, the sampling rate is α, the size of the sliding window is $N \times T$, B is the collected data matrix, and R is the record matrix.
Input: The data collection matrix B and the reconstruction error σ_b.
Output: The record vector $\vec{R}_i$.

```
num = (N+2)(2α−α²) / (2+4α−2α²)   // The minimum amount of data
                                  // to collect
W_{N×T} = [B⃗^T_{i−T+1}, B⃗^T_{i−T+2}, ··· , B⃗^T_i]
While c ≥ num
    X = SVT(W_{N×T})              //Recovery using the SVT
                                  //algorithm
    σ = Σ_{i,j,W_{i,j}==1} (X_{i,j} − W_{i,j})² / Σ_{i,j,W_{i,j}==1} W²_{i,j}
                                  //The reconstruction errors
    IF σ ≤ σ_b
        n = C(σ_b − σ) // Reduced amount of data
        c = c − n
        IF current slot i is odd
            Randomly assigned n known data in B⃗_i to 0,
            and priority node number is even.
        Else
            Randomly assigned n known data in B⃗_i to 0,
            and priority node number is odd.
        End if
    Else
        break;
    End if
End while
For j = 1:N
    IF B⃗_i (j) ≠ 0
        R⃗_i (j) = 1
    Else
        R⃗_i (j) = 0
End for
```

Algorithm 4 The ACMC Scheme for Adapted Collection

Initialize: Considering the current i^{th} slot, the sliding window is $N \times T$, B is the data collection matrix.
Input: The data collection matrix B and the record matrix R.
Output: The data collection vector $\vec{B}_i$, the record vector $\vec{R}_i$.

IF the current slot is odd
  Randomly arrange the odd numbers in $1-N$ and store in S.
Else
  Randomly arrange the even numbers in $1-N$ and store in S.
End if
num $= \min(sum\left(\vec{R}_{i-1}\right), \text{sum}(\vec{R}_{i-T+1}))$
  // sum is the summation function, and the vector $\vec{R}$ is
  // summed with the amount of data that needs to be
  // collected
$W_{N\times T} = [\vec{B}^T_{i-T+1}, \vec{B}^T_{i-T+2}, \cdots, \vec{B}^T_i]$.
k=1
While 1
  While $k \leq num$
    Inform the $S(k)$ node to transfer data to the sink and assign the value to $\vec{B}_i(S(k))$
  End while
  $X = \text{SVT}(W_{N\times T})$
  $\sigma = \frac{\sum_{i,j,W_{i,j}=1}|X_{i,j}-W_{i,j}|}{\sum_{i,j,W_{i,j}=1} W_{i,j}}$ //The reconstruction error
  IF $\sigma > \sigma_b$
    num $=$ num $+ C(\sigma - \sigma_b)$
  Else
    break;
  End if
End while
For $j = 1{:}N$
  IF $\vec{B}_i(j) \neq 0$
    $\vec{R}_i(j) = 1$
  Else
    $\vec{R}_i(j) = 0$
End for
Calculate the energy consumption E_u according to Theorem 3 and $\vec{B}_i$

a data packet can be obtained as $(1-P_e)^L$, so the packet loss rate can be obtained is $1-(1-P_e)^L$.

The time required to transfer a packet is

$$T_p = \frac{L}{bB}$$

where b is the modulation size and B is the channel bandwidth.

In a transmission cycle, the time of the node runs is $\tau_i T$, so the data packet can be sent, at most, $\frac{\tau TbB}{L}$ times in one cycle. Therefore, when the node duty cycle is τ_i, the probability of unsuccessful transmission is $(1-(1-P_e)^L)^{\tau_i TBb/L}$, and the reliability of the transmission can be obtained; thus,

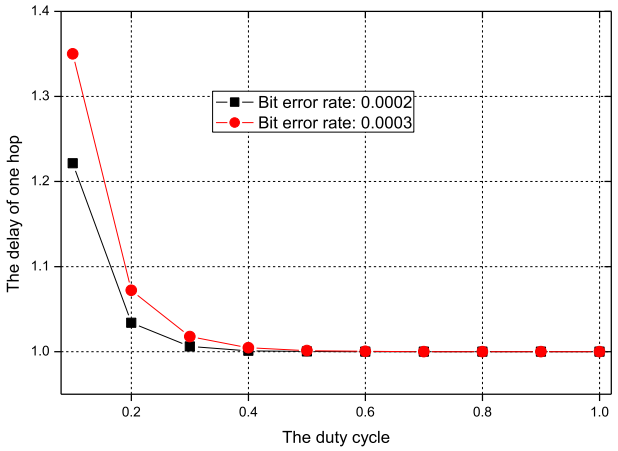


FIGURE 9. The delay under different duty cycles.

the delay of a single hop can be obtained as

$$\delta(\tau_i) = \frac{1}{1-(1-(1-P_e)^L)^{\tau_i TBb/L}}$$

■

Theorem 6 gives the single-hop delay of the node at different duty cycles. Fig. 9 shows the delay of different duty cycles. It can be seen that the delay of the node decreases as the duty cycle increases, which indicates that it is effective to reduce the delay of node transmission and increase the energy utilization by increasing the duty cycle.

V. THE EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL RESULTS

In the simulation experiment, the illumination data of the National Solar Radiation Data Base at DENVER/CENTENNIAL, CO in 2010 was used [51], considering that the solar panel size of each node is 20 cm × 10 cm, the initial energy of each node is 70 Wh, and the maximum stored energy of the battery is 110 Wh. The network collects data every hour and collects 10 days of data, that is, a total of 240 time slots, and the sliding window size is 52×60. In the ACMC scheme, the sampling rate of 0.5 represents that the probability of a row in the sliding window without any data is no more than the uniform distribution scheme (UDS) with a sampling rate of 0.5.

Fig. 10 shows the available energy and the actual maximum energy consumed of each node when the sampling rate of the network is 0.9. The actual maximum energy consumed does not exceed the available energy, and the network collects all the data directly when the available energy is much higher than the actual energy consumed; thus, during this time, there is no change in the actual maximum energy consumption. When the solar radiation begins to decrease, the energy converted by solar radiation is preferentially used in Algorithm 1, so the available energy during this time will gradually decrease. When the energy converted by solar radiation is no longer able to collect half of the data,

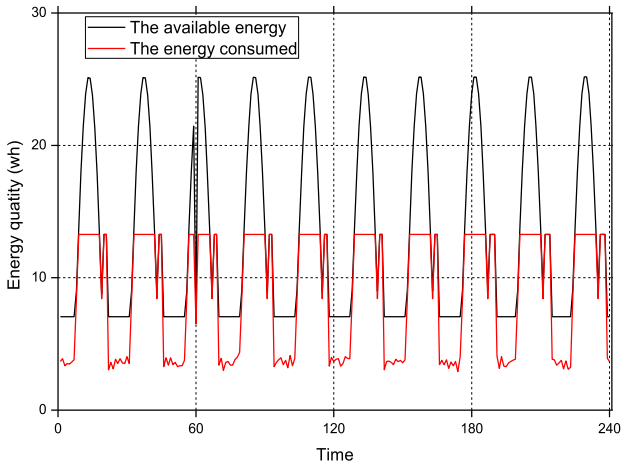


FIGURE 10. The available energy and actual energy consumed.

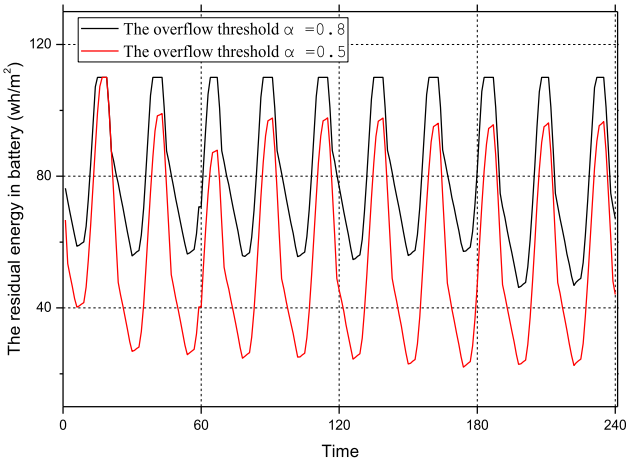


FIGURE 12. The residual energy at different overflow thresholds.

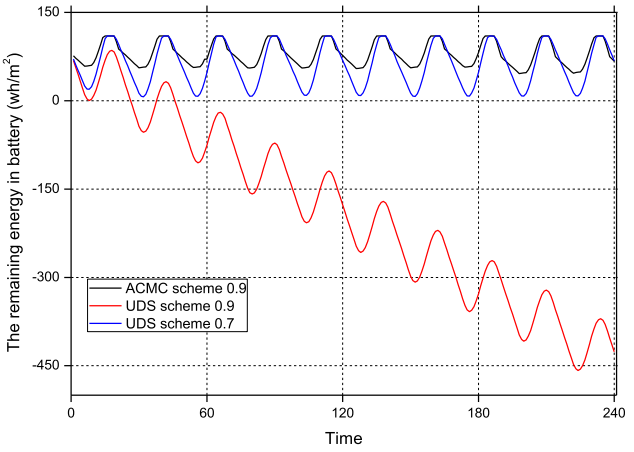


FIGURE 11. The remaining energy in the battery under different schemes.

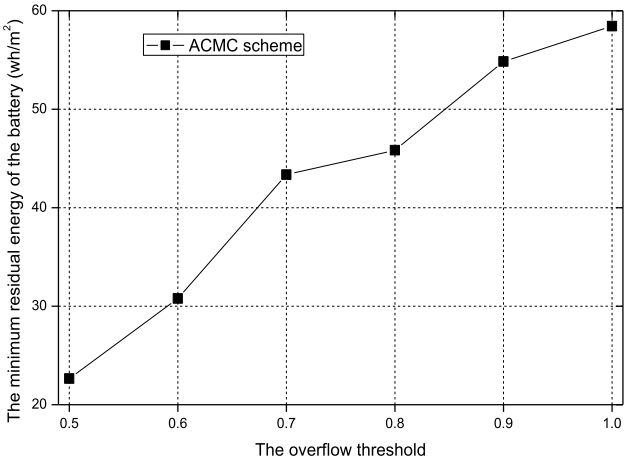


FIGURE 13. The minimum residual energy at different overflow threshold.

the ACMC scheme will continue to collect all the data if the residual energy in the battery is still sufficient (the residual energy in the battery is considered to be sufficient if the residual energy in the battery is greater than the overflow threshold), so that the actual energy consumption has reached the maximum and there is still high available energy after the solar radiation is reduced.

Fig. 11 shows the residual energy in the battery under the ACMC scheme and UDS scheme. It can be seen that in the UDS scheme, when the sampling rate is 0.9, the energy consumed by the node at night has exceeded the energy stored in the battery, so the UDS scheme with a sampling rate of 0.9 does not meet the requirements. In the ACMC scheme, when the sampling rate is 0.9, the residual energy in the battery is still sufficient, so the ACMC scheme can collect data. Therefore, the main comparison scheme is the UDS scheme with a sampling rate of 0.7.

Fig. 12 is the residual energy of the ACMC scheme after adjusting the overflow threshold. Since the overflow threshold is 0.8, there will be a large amount of energy remaining in the battery. Therefore, the overflow threshold can be appropriately lowered so that the number of slots for collecting all the sensor node data is increased, so the residual energy in the battery decreases. When the overflow threshold is at 0.5, the residual energy in the battery decreases compared with the overflow threshold at 0.8.

Fig. 13 is the minimum residual energy in the battery at different overflow thresholds; the lower the overflow threshold, the lower the minimum residual energy in the battery. This is because the lower the overflow threshold, the more slots that are determined to have sufficient energy. Therefore, the overflow threshold affects the residual energy in the battery. To improve energy utilization, in the next experiment, the overflow threshold is 0.5.

Fig. 14 shows the reconstruction error of the ACMC scheme and the UDS scheme. The sampling rate of the ACMC scheme is 0.9, and the sampling rate of the UDS scheme is 0.7. Since the sliding window size is 52×60, the first 60 time slots cannot calculate the reconstruction error. Fig. 14 shows that the reconstruction error of the ACMC scheme is significantly lower than the reconstruction error of

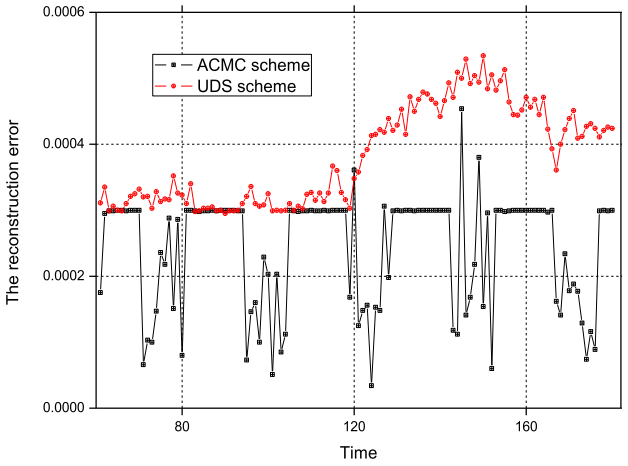


FIGURE 14. The reconstruction errors.

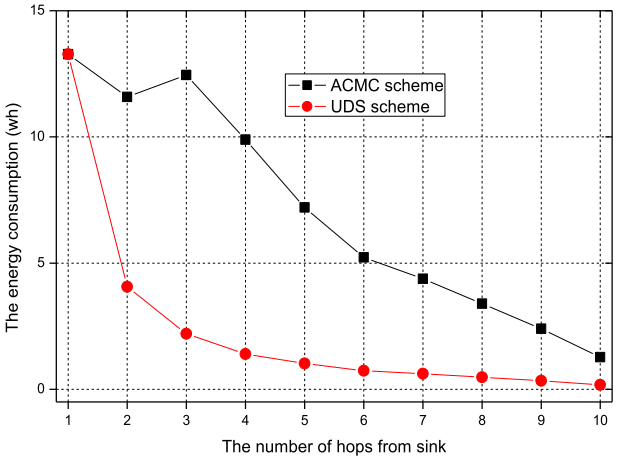


FIGURE 16. The energy consumption of each node under the ACMC and UDS schemes.

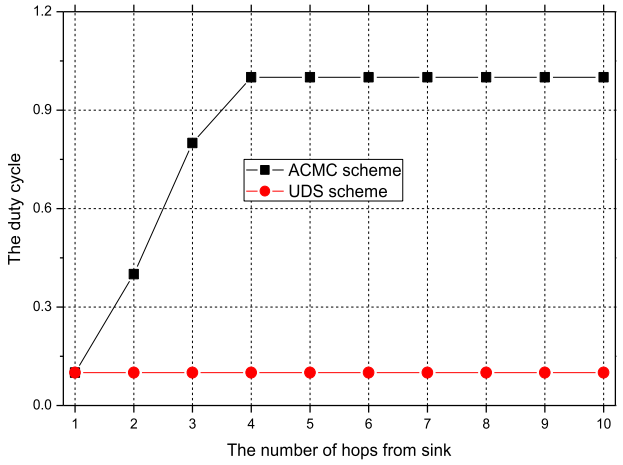


FIGURE 15. The duty cycle of each layer node.

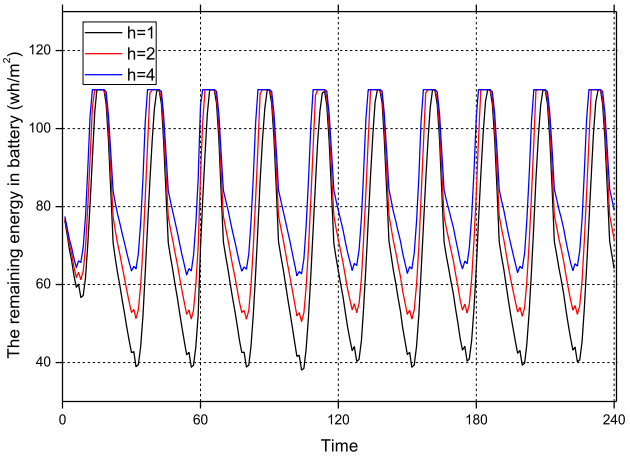


FIGURE 17. The remaining energy in the battery under different layers.

the UDS scheme. In the UDS scheme, the reconstruction error has significant growth because the rank of the adjacent matrix is always changing, and in the ACMC scheme, the amount of collected data will be adaptively adjusted to avoid this problem.

Fig. 15 shows the duty cycle of each node in the ACMC scheme; the higher the layer, the higher the duty cycle selected by the nodes in this layer. Fig. 16 shows the energy consumption of each layer of the network. After increasing the duty cycle of the node, the energy consumption of the nodes in the second and third layers is almost the same as the energy consumption of the nodes in the first layer. However, since the duty cycle can only be increased to 1, the energy consumption of the nodes in the later layers cannot be increased to the same level as the nodes of the first layer, so this scheme can alleviate the problem of imbalance in energy consumption.

Fig. 17 shows the remaining energy of the nodes in different layers. It can be seen that the battery of the first layer has a minimum energy of approximately 40 Wh, and the remaining energy of the nodes in the other layers is more than that of the first layer, thus making the entire scheme feasible.

B. PERFORMANCE COMPARISON IN ACMC and UDS

In this section, we will compare the performance differences between the ACMC scheme and the UDS scheme.

Fig. 18 is the cumulative energy consumption of the node with the most energy consumption in the network. It can be seen that the final energy consumption of the ACMC scheme is higher than the UDS scheme with a sampling rate of 0.7, and the UDS scheme has already exceeded the available energy when it starts collecting data. As shown in Fig. 19, when the UDS scheme starts collecting data, the energy utilization exceeds 1; that is, its energy consumption exceeds the available energy. However, the energy utilization rate of the UDS scheme after long-term operation is not as good as that of the ACMC scheme. The energy utilization rate of the UDS scheme is maintained at approximately 0.67, while the energy utilization rate of the ACMC scheme is maintained at approximately 0.7. The single node energy utilization of

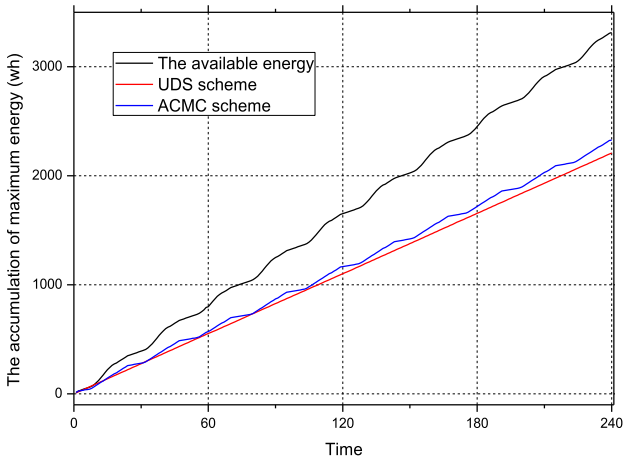


FIGURE 18. The cumulative maximum energy consumption.

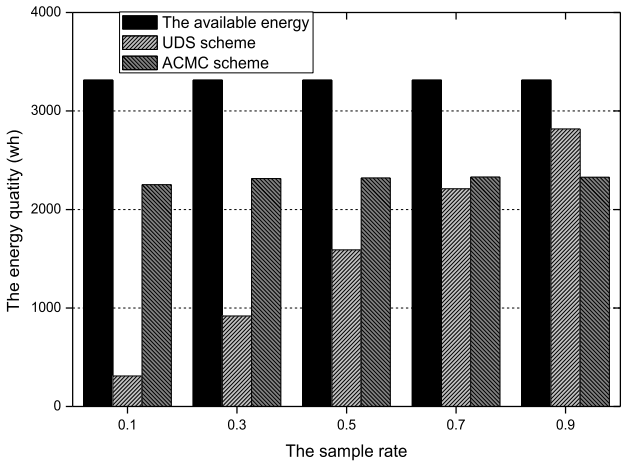


FIGURE 20. The maximum energy consumption of the network at different sampling rates.

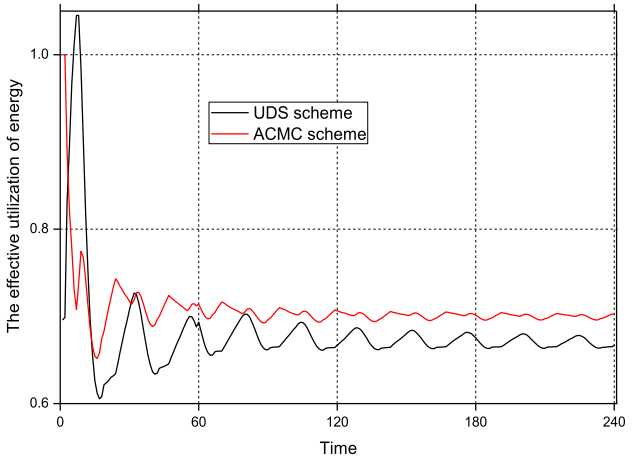


FIGURE 19. The maximum energy utilization of a single node.

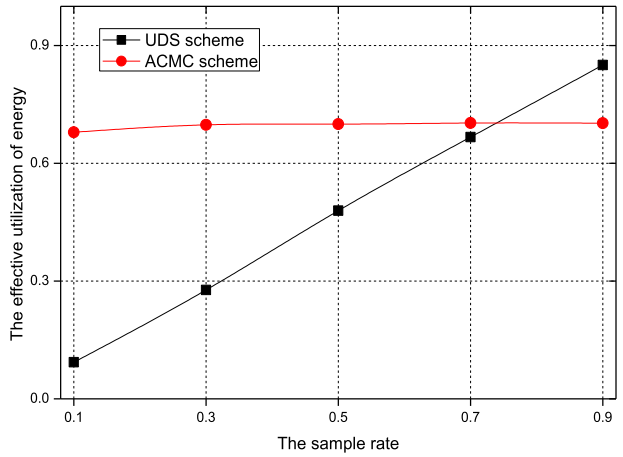


FIGURE 21. The energy utilization at different sampling rates.

the ACMC scheme is increased by 0.03 compared to the UDS scheme.

Fig. 20 shows the energy consumption of a single node in the network. It can be seen that if two schemes require the same sampling rate, the energy consumption of the ACMC scheme is much higher than that of the UDS scheme at low sampling rates, but at high sampling rates, the energy consumption of the ACMC scheme is lower than that of the UDS scheme. Fig. 21 shows the energy utilization of the network. The energy utilization rate of the ACMC scheme is relatively stable and does not change significantly. The energy utilization rate of the UDS scheme increases approximately linearly. The energy utilization of the ACMC scheme is between 0.68 and 0.7, and the energy utilization of the UDS scheme is between 0.1 and 0.85. In Fig. 20, it can also be seen that when the required sampling rate is increased from 0.7 to 0.9, the energy utilization rate of the ACMC scheme is reduced. This is because the required sampling rate is high, and the minimum amount of data that needs to be collected increases so that the energy consumption correspondingly increases, resulting in insufficient energy supplementation during the daytime. Therefore, the slot exceeding the overflow threshold will be smaller, so the slot for collecting all the data will be reduced, which will result in final energy consumption lower than the required sampling rate of 0.7.

Fig. 22 shows the cumulative total energy consumption of the network. The total energy consumption of the ACMC scheme is larger than that of the UDS scheme. As shown in Fig. 23, the energy utilization of the UDS scheme is also much lower than the energy utilization of the ACMC scheme. However, because the energy consumption of the network is still unbalanced, the overall energy utilization is also low. The final energy utilization of the ACMC scheme is maintained at approximately 0.073, and the final energy utilization of the UDs scheme is maintained at approximately 0.023. The energy utilization of the ACMC scheme increases by approximately 5%.

Fig. 24 shows the energy consumption of the two schemes at the same sampling rate. Even at high sampling rates, the total energy consumption of the UDS scheme is much

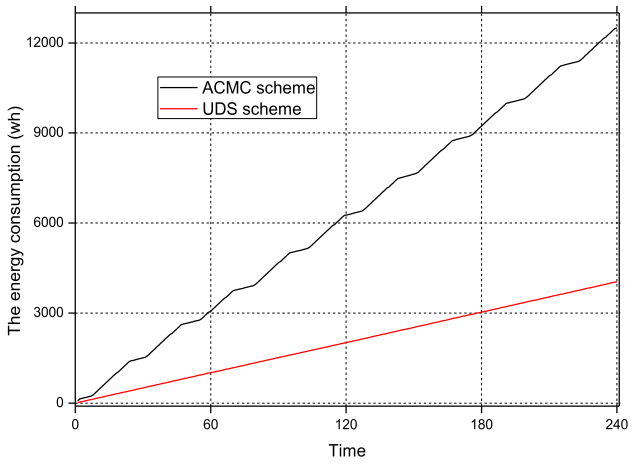


FIGURE 22. The cumulative total energy consumption.

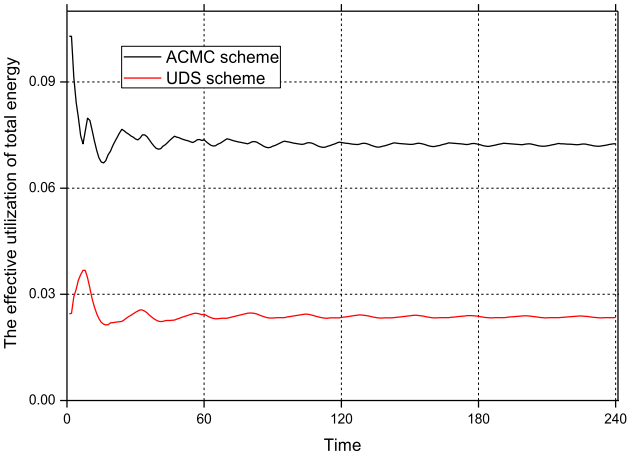


FIGURE 23. The energy utilization of the network.

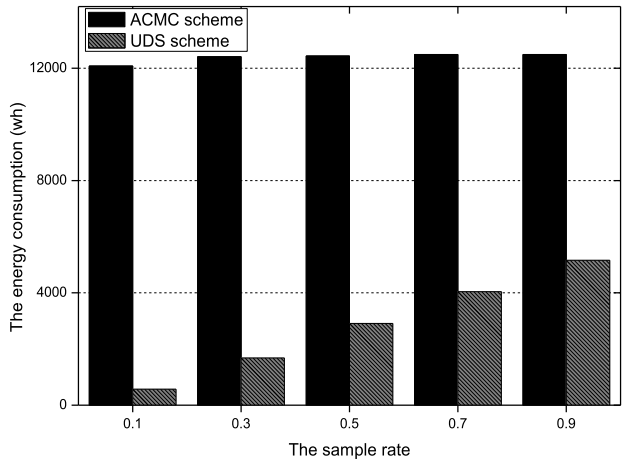


FIGURE 24. The total energy consumption at different sampling rates.

lower than that of the ACMC scheme. As shown in Fig. 25, the energy utilization of the ACMC scheme is maintained at approximately 0.08. Compared with the UDS scheme,

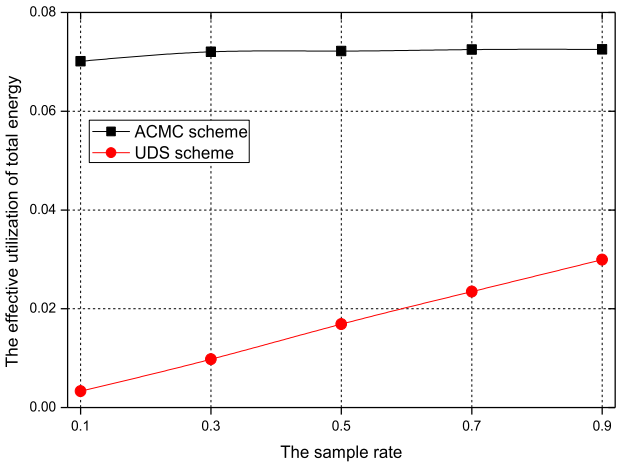


FIGURE 25. The energy utilization of the network at different sampling rates.

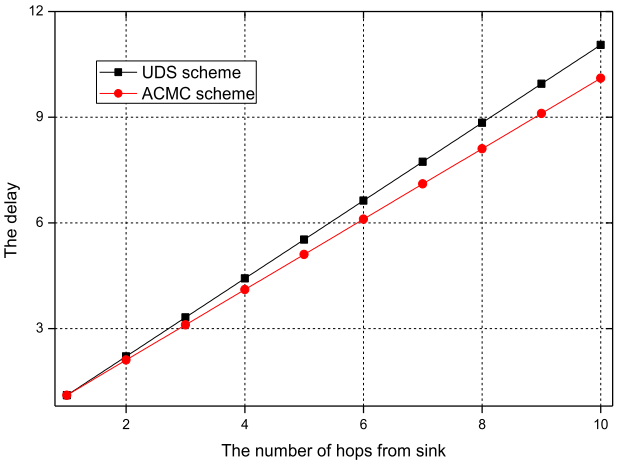


FIGURE 26. The delay of each layer node.

the energy utilization rate of the ACMC scheme is increased by 4.26%-6.68%.

Fig. 26 is the delay of each layer in the network. Because the ACMC scheme adjusts the duty cycle, the maximum delay of the ACMC scheme is better than that of the UDS scheme. The maximum delay of the ACMC scheme is reduced by approximately 9.4% compared with the UDS scheme.

VI. CONCLUSION

In this paper, we propose a data collection scheme based on matrix completion technology to optimize the performance of EHWSN. Different from the previous schemes, the scheme proposed in this paper can dynamically change the amount of data collected at each moment, collecting as little data as possible when the energy in the EHWSN is not sufficient. At the same time, it also guarantees the reconstruction error of recovered data. When the energy of the network is sufficient, more data is collected, and possibly all the data is collected. Therefore, this scheme can effectively improve the energy utilization of nodes. However, due to a large amount

of residual energy in the nodes far from the sink, reducing the amount of data collected cannot effectively improve the energy utilization of the network. Therefore, in the ACMC scheme, different duty cycles are also set according to the hop count of the node, which reduces the transmission delay and improves the energy utilization of the nodes away from the sink, thereby improving the energy utilization of the entire network. Therefore, the ACMC scheme is significant for research in the field of energy harvesting networks.

REFERENCES

[1] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the suitability of fog computing in the context of Internet of Things," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 46–59, Jan./Mar. 2018.

[2] Y. Liu, M. Ma, X. Liu, N. Xiong, A. Liu, and Y. Zhu, "Design and analysis of probing route to defense sink-hole attacks for Internet of Things security," *IEEE Trans. Netw. Sci. Eng.*, to be published, doi: 10.1109/TNSE.2018.2881152.

[3] T. Li, S. Tian, A. Liu, H. Liu, and T. Pei, "DDSV: Optimizing delay and delivery ratio for multimedia big data collection in mobile sensing vehicles," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3474–3486, Oct. 2018, doi: 10.1109/JIOT.2018.2847243.

[4] X. Wang *et al.*, "A privacy-preserving message forwarding framework for opportunistic cloud of things," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2018.2864782.

[5] J. Zhang *et al.*, "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2633–2645, Aug. 2018.

[6] W. Yang *et al.*, "Adding active slot joint larger broadcast radius for fast code dissemination in WSNs," *Sensors*, vol. 18, no. 11, p. 4055, 2018, doi: 10.3390/s18114055.

[7] W. Qi *et al.*, "Minimizing delay and transmission times with long lifetime in code dissemination scheme for high loss ratio and low duty cycle WSNs," *Sensors*, vol. 18, no. 10, p. 3516, 2018, doi: 10.3390/s18103516.

[8] X. Liu *et al.*, "Construction of large-scale low-cost delivery infrastructure using vehicular networks," *IEEE Access*, vol. 6, no. 1, pp. 21482–21497, 2018.

[9] M. Wu *et al.*, "An effective delay reduction approach through portion of nodes with larger duty cycle for industrial WSNs," *Sensors*, vol. 18, no. 5, p. 1535, 2018, doi: 10.3390/s18051535.

[10] X. Xiang, W. Liu, N. Xiong, H. Song, A. Liu, and T. Wang, "Duty cycle adaptive adjustment based device to device (D2D) communication scheme for WSNs," *IEEE Access*, vol. 6, no. 1, pp. 76339–76373, 2018.

[11] B. Huang, A. Liu, C. Zhang, N. Xiong, Z. Zeng, and Z. Cai, "Caching joint shortcut routing to improve quality of service for information-centric networking," *Sensors*, vol. 18, no. 6, p. 1750, 2018, doi: 10.3390/s18061750.

[12] M. Huang, A. Liu, M. Zhao, and T. Wang, "Multi working sets alternate covering scheme for continuous partial coverage in WSNs," *Peer-to-Peer Netw. Appl.*, to be published, doi: 10.1007/s12083-018-0647-z.

[13] H. Teng *et al.*, "Adaptive transmission range based topology control scheme for fast and reliable data collection," *Wireless Commun. Mobile Comput.*, vol. 2018, Jul. 2018, Art. no. 4172049, doi: 10.1155/2018/4172049.

[14] H. Teng *et al.*, "A novel code data dissemination scheme for Internet of Things through mobile vehicle of smart cities," *Future Gener. Comput. Syst.*, vol. 94, pp. 351–367, May 2019.

[15] T. Li, N. N. Xiong, J. Gao, H. Song, A. Liu, and C. Zhiping, "Reliable code disseminations through opportunistic communication in vehicular wireless networks," *IEEE Access*, vol. 6, no. 1, pp. 55509–55527, 2018.

[16] F. Xiao, X. Xie, Z. Li, Q. Deng, A. Liu, and L. Sun, "Wireless network optimization via physical layer information for smart cities," *IEEE Netw.*, vol. 32, no. 4, pp. 88–93, Jul./Aug. 2018.

[17] C. Zhang, R. Chen, L. Zhu, A. Liu, Y. Lin, and F. Huang, "Hierarchical information quadtree: Efficient spatial temporal image search for multimedia stream," *Multimedia Tools Appl.*, to be published, doi: 10.1007/s11042-018-6284-y.

[18] T. Li, Y. Liu, N. N. Xiong, A. Liu, Z. Cai, and H. Song, "Privacy-preserving protocol for sink node location in telemedicine networks," *IEEE Access*, vol. 6, no. 1, pp. 42886–42903, 2018.

[19] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, "PEA: Parallel electrocardiogram-based authentication for smart healthcare systems," *J. Netw. Comput. Appl.*, vol. 117, pp. 10–16, Sep. 2018.

[20] Z. Li, Y. Liu, A. Liu, S. Wang, and H. Liu, "Minimizing convergecast time and energy consumption in green Internet of Things," *IEEE Trans. Emerg. Topics Comput.*, to be published, doi: 10.1109/TETC.2018.2844282.

[21] X. Li *et al.*, "Differentiated data aggregation routing scheme for energy conserving and delay sensitive wireless sensor networks," *Sensors*, vol. 18, no. 7, p. 2349, 2018, doi: 10.3390/s18072349.

[22] Y. Ren, W. Liu, Y. Liu, N. N. Xiong, A. Liu, and X. Liu, "An effective crowdsourcing data reporting scheme to compose cloud-based services in mobile robotic systems," *IEEE Access*, vol. 6, no. 1, pp. 54683–54700, 2018.

[23] A. Liu and S. Zhao, "High-performance target tracking scheme with low prediction precision requirement in WSNs," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 29, no. 4, pp. 270–289, 2018.

[24] S. Fang *et al.*, "Feature selection method based on class discriminative degree for intelligent medical diagnosis," *Comput., Mater. Continua*, vol. 55, no. 3, pp. 419–433, 2018.

[25] Y. Liu, A. Liu, X. Liu, and X. Huang, "A statistical approach to participant selection in location-based social networks for offline event marketing," *Inf. Sci.*, vol. 480, pp. 90–108, Apr. 2019.

[26] C. Yang, Z. Shi, K. Han, J. J. Zhang, Y. Gu, and Z. Qin, "Optimization of particle CBMeMBer filters for hardware implementation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 9027–9031, Sep. 2018, doi: 10.1109/TVT.2018.2853120.

[27] X. Hu *et al.*, "Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 180–187, Apr. 2018.

[28] X. Liu, W. Liu, Y. Liu, H. Song, A. Liu, and X. Liu, "A trust and priority based code updated approach to guarantee security for vehicles network," *IEEE Access.*, vol. 6, pp. 55780–55796, 2018.

[29] Y. Zhang, Z. Tu, and Q. Wang, "TempoRec: Temporal-topic based recommender for social network services," *Mobile Netw. Appl.*, vol. 22, no. 6, pp. 1182–1191, 2017.

[30] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "EARS: Emotion-aware recommender system based on hybrid information fusion," *Inf. Fusion*, vol. 46, pp. 141–146, Mar. 2019.

[31] S. Xiao, H. Yu, Y. Wu, Z. Peng, and Y. Zhang, "Self-evolving trading strategy integrating Internet of Things and big data," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2518–2525, Aug. 2018.

[32] W. Hou *et al.*, "On-chip hardware accelerator for automated diagnosis through human-machine interactions in healthcare delivery," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 1, pp. 206–217, Jan. 2019, doi: 10.1109/TASE.2018.2832454.

[33] Y. Guo, X. Hu, B. Hu, J. Cheng, M. Zhou, and R. Y. K. Kwok, "Mobile cyber physical systems: Current challenges and future networking applications," *IEEE Access*, vol. 6, pp. 12360–12368, 2018.

[34] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4568–4578, Oct. 2018, doi: 10.1109/TII.2018.2816590.

[35] L. Guo, Z. Ning, W. Hou, B. Hu, and P. Guo, "Quick answer for big data in sharing economy: Innovative computer architecture design facilitating optimal service-demand matching," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1494–1506, Oct. 2018, doi: 10.1109/TASE.2018.2838340.

[36] X. Ju *et al.*, "An energy conserving and transmission radius adaptive scheme to optimize performance of energy harvesting sensor networks," *Sensors*, vol. 18, no. 9, p. 2885, 2018, doi: 10.3390/s18092885.

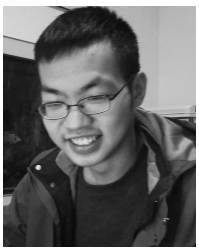
[37] Z. Li and Y. Yang, "RRect: A novel server-centric data center network with high power efficiency and availability," *IEEE Trans. Cloud Comput.*, to be published, doi: 10.1109/TCC.2018.2816650.

[38] T. Raptis, A. Passarella, and M. Conti, "Performance analysis of latency-aware data management in industrial IoT networks," *Sensors*, vol. 18, no. 8, p. 2611, 2018.

[39] R. Anane, K. Raoof, and R. Bouallegue, "Minimization of wireless sensor network energy consumption through optimal modulation scheme and channel coding strategy," *J. Signal Process. Syst.*, vol. 83, no. 1, pp. 65–81, 2016.

[40] K. Xie, L. Wang, X. Wang, G. Xie, and J. Wen, "Low cost and high accuracy data gathering in WSNs with matrix completion," *IEEE Trans. Mobile Comput.*, vol. 17, no. 7, pp. 1595–1608, Jul. 2018.

[41] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 850–861, Feb. 2013.
[42] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
[43] G. H. Badawy, A. A. Sayegh, and T. D. Todd, "Energy provisioning in solar-powered wireless mesh networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 8, pp. 3859–3871, Oct. 2010.
[44] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.
[45] J. L. Liu and C. V. Ravishankar, "LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks," *Int. J. Mach. Learn. Comput.*, vol. 1, no. 1, p. 79, 2011.
[46] D. Wu, J. He, H. Wang, C. Wang, and R. Wang, "A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 92–98, Aug. 2015.
[47] T. Amgoth and P. K. Jana, "Energy-aware routing algorithm for wireless sensor networks," *Comput., Electr. Eng.*, vol. 41, pp. 357–367, Jan. 2015.
[48] Y. Zhang, M. Chen, N. Guizani, D. Wu, and V. C. Leung, "SOVCAN: Safety-oriented vehicular controller area network," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 94–99, Aug. 2017.
[49] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010
[50] S. Madden. (2004). *Intel Lab Data*. [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html
[51] National Solar Radiation Data Base. *Denver/Centennial [Golden-NREL]*. [Online]. Available: https://rredc.nrel.gov/solar/old_data/nsrdb/1991-2010/hourly/list_by_USAFN.html



JIAWEI TAN is currently pursuing the master's degree with the School of Information Science and Engineering, Central South University, China. His research interests include data process, crowd sensing networks, and wireless sensor networks.

WEI LIU received the Ph.D. degree in computer application technology from Central South University, China, in 2014. He is currently an Associate Professor and a Senior Engineer with the School of Informatics, Hunan University of Chinese Medicine, China. His research interests include software engineering, data mining, and medical informatics. He has published over 20 papers in the related fields.

TIAN WANG received the B.Sc. and M.Sc. degrees in computer science from Central South University, in 2004 and 2007, respectively, and the Ph.D. degree from the City University of Hong Kong, in 2011. He is currently an Associate Professor with Huaqiao University, China. His research interests include wireless sensor networks, social networks, and mobile computing.

NEAL N. XIONG received the dual Ph.D. degree from Wuhan University (about sensor system engineering) and from the Japan Advanced Institute of Science and Technology, respectively. He was with Georgia State University, Wentworth Technology Institution, and Colorado Technical University about 10 years. He is currently an Associate Professor with the Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory.

He published over 200 international journal papers and over 100 international conference papers. Some of his works were published in the IEEE JSAC, IEEE or ACM TRANSACTIONS, ACM Sigcomm Workshop, IEEE INFOCOM, ICDCS, and IPDPS. He is a Senior Member of the IEEE Computer Society. He has received the Best Paper Award from the 10th IEEE International Conference on High Performance Computing and Communications (2008) and the Best student Paper Award from the 28th North American Fuzzy Information Processing Society Annual Conference (2009). He has been a General Chair, a Program Chair, a Publicity Chair, a PC Member, and an OC Member of over 100 international conferences. He has been a Reviewer of about 100 international journals, including the IEEE JSAC, IEEE SMC (Park: A/B/C), the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. He is serving as the Editor-in-Chief, an Associate Editor, or an Editor Member for over ten international journals (including an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, an Associate Editor for *Information Science*, the Editor-in-Chief for the *Journal of Internet Technology*, and the Editor-in-Chief for the Journal of *Parallel & Cloud Computing*), and a Guest Editor for over ten international journals, including the *Sensor* Journal, WINET, and MONET.

HOUBING SONG (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 2012. In 2017, he joined the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory. In 2007, he was an Engineering Research Associate with the Texas A&M Transportation Institute. He has served on the faculty of West Virginia University, from 2012 to 2017. He is an Editor of four books, including *Smart Cities: Foundations, Principles, and Applications* (Hoboken, NJ, USA: Wiley, 2017), *Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications* (Chichester, U.K.: Wiley–IEEE Press, 2017), *Cyber-Physical Systems: Foundations, Principles and Applications* (Boston, MA,USA: Academic Press, 2016), and *Industrial Internet of Things: Cybermanufacturing Systems* (Cham, Switzerland: Springer, 2016). He has authored over 100 articles. His research interests include cyber-physical systems, cybersecurity and privacy, the Internet of Things, edge computing, big data analytics, unmanned aircraft systems, connected vehicles, smart and connected health, and wireless communications and networking. He serves as an Associate Technical Editor for the *IEEE Communications Magazine*.

Dr. Song is a Senior Member of the ACM. He was the very first recipient of the Golden Bear Scholar Award, the highest campus-wide recognition for research excellence at West Virginia University Institute of Technology, in 2016.

ANFENG LIU received the M.Sc. and Ph.D. degrees in computer science from Central South University, China, in 2002 and 2005, respectively, where he is currently a Professor with the School of Information Science and Engineering. His major research interest includes wireless sensor networks. He is a member (E200012141M) of the China Computer Federation.

ZHIWEN ZENG is currently an Associate Professor with the School of Information Science and Engineering, Central South University, China. His major research interests include wireless sensor networks and distributed computing.