Data Compression Algorithm for Wireless Sensor Networks

Dawei Ye
College of Information Science and Technology. Hainan University
Haikou, PR China

Chong Shen*
College of Information Science and Technology. Hainan University
Haikou, PR China
Corresponding Author:*chongshen@hainu.edu.cn

Abstract—Energy consumption is a crucial problem affecting the lifetime of Wireless Sensor Networks (WSNs). In this research, we use data compression approach to reduce transmission data over wireless channels. A Distributed Wavelet Transform (DWT) is proposed for WSNs which is based on lifting scheme. The proposed scheme overcomes distributed wavelet algorithm issues such as time-consuming and calculation complexity. We use a Ring Topology based on Virtual Grid (VGRT) model that can effective eliminate spatial correlation in sensing data. Also in this paper, a novel data transmission mechanism named Data Selecting Mechanism (DSM) is designed to reduce the temporal correction in sensing data. It decreases the redundant data from its source, and cuts down the data needed for transmission. Simulation results are based on the Wavelet Lifting Transform (WLT) with DSM. Theoretically and experimentally, it is concluded that the proposed algorithm can reduce redundant data in WSNs efficiently and greatly decrease the energy consumption and prolongs the lifecycle.

Keywords—Data Compression; Wavelet Lifting Transform; Data Selecting Mechanism

I. INTRODUCTION

Nowadays, one of the key technologies to support ubiquitous computing is WSNs. With the advance of technology, sensors in WSNs are combined with sensing unit, process unit and communication unit. When we design a WSNs system, there are hundreds or even thousands of sensors distributed over a field [1-2]. Sensors are used to perform tasks such as target tracking (vehicles, chemical agents, or personnel), traffic controlling, environment monitoring and surveillance [3-4]. Each sensor node has its own processor, but due to the limited energy, reduced bandwidth and small storage resource, it is difficult to apply a large amount of data transmission in the networks. The physical characteristics of WSNs node bring a great challenge to information processing. If the networks only transmit the gathering data to the center console, it takes a big delay of the transmission, and also consumes the limited energy of nodes. Finally, it leads to the loss of power value of the entire networks [5].

Although existing data compression techniques for WSNs have been surveyed in [6]. Traditional wavelet algorithm without fully considering the issues mentioned above [7], it requires more calculation and large storage. Whether the gathering data is changing, every node process the data and transmit it, and it also takes too much energy to transmit this redundant data. To reach the objective of reducing data streams, the lifting scheme proposed to optimize the issues. Detail about WLT algorithm can be found in [8]. Furthermore, we propose DSM based on data differences in our work. By using the proposed algorithm, we can eliminate the temporal correction in sensing data.

The rest of this article is organized as follows. First, Section 2 presents the related work about data compression in WSNs and the issues we have faced. Then, in order to analyze, compare and solve the presented issues, the definitions of criteria based on the requirements of WSNs are described in Section 3. Section 4 provides an energy model to analyze the advantage of using our algorithm which is combining WLT with DSM. Experiment simulation system and numerical results are given in Section 5. Finally, Section 6 shows the conclusion of our work.

II. RELATED WORK

Over the past few years, the decreasing cost for WSNs has brought lots of attention to the development of a variety of distributed compression algorithm to enhance the performance of such networks. Data compression can be applied to cut down the amount of data in WSNs. The authors have discussed the using of static Huffman encoding to compress the sensing data in [9]. This proposed algorithm takes advantage of the characteristics of high correlation in sensing data. The biggest merits of this algorithm are simply in design and small amount of calculation.

In more recent work, Amar et al. [10] consider the data fusion problem in which each sensor node observes part of a data vector. Every data vector is compressed or encoded by means of KLT (Karhunen-Loeve Transform), those encoding matrices are sent to the fusion center. Then, we reconstruct the entire data vector from these compressed sensor observations with minimal MSE (Mean Square Error). The advantage of this algorithm is lower complexity, but they did not evaluate their energy consumption performance.

There are several works for wavelet transform approaches in the literature that propose a modified version of distributed
wavelets transform based on the lifting scheme. In Ciancio and Ortega’s work [7-8], they proposed the lifting scheme to generate 5/3 wavelet coefficients at each of the sensor nodes. Each sensor node needs data from its neighbors in this implementation, and computes the correspondent detail coefficient. A key concept of their work is to perform partial computations of the coefficients at each node, then reduce energy consumption in WSNs.

It is interesting to mention that data compression can be applied to reduce the amount of data sending by sensor nodes.

The algorithm proposed in [9-10] is a simple design and low calculation method, but it also doesn’t consider the spatial correlation in sensing data. For the physical properties of nodes, like the limited energy and low calculation, we find wavelet algorithm based on lifting can effectively eliminate the spatial correlation. There is a problem, each sensor node needs data from its neighbors in this implementation, and it is still waste much energy for nodes to communicate with other nodes. So, in this work, we bring DSM in every node, which will be detailed in Section 3. Processed by DSM, every node does not need to participate in WLT. It also decreases the redundant data from its source, cuts down the transmission data streams, which achieves the purpose of data compression.

III. SYSTEM ARCHITECTURE

A. Ring topology based on virtual grid (VGRT)

Here, we use a grid model that can effective eliminate spatial correlation in sensing data. The WSNs is divided into multi-cluster in order to processing the sensing data. Assuming that every cluster communicates with other clusters, larger cluster network is in the form of all sub-clusters. In such way, each cluster selects a node as the cluster head gathering the sensing data from each member nodes. Due to the random deployment in WSNs or other reasons, it makes uneven distribution within the cluster member nodes. Therefore, the additional energy consumption of some redundant nodes for monitoring, receiving and transmitting data is brought to the WSNs. So, we divide the cluster into a small virtual grid of M×N. Fig. 1 illustrates the scenario of VGRT.

![Fig. 1. Ring topology based on virtual grid (VGRT)](image)

Virtual grid A is closed to grid B, C, D, E, F, G, H and I. For any two adjacent grids X and Y, nodes in X can communicate with any nodes in Y, and the reverse is also true. In the case of assuming the coverage of detection area, we suppose that every virtual grid has one node working in it. Of course, other dormant nodes will be waked up at a right time to take place of the energy depleted nodes. Nodes will take turns as the first node in the ring and broadcast the data to virtual network. Based on this method, we don’t need to structure a new ring. It also can make sure the properly work of networks and prolongs the lifecycle of WSNs. The dispatch algorithm of VGRT has been discussed in [11].

B. Wavelet algorithm based on lifting

In our work, we use DWT based on the lifting scheme [4-5]. The lifting factorization provides a convenient representation of the transform as it assumes in place computation, and explicitly breaks down the transform into elementary operations that can be easy evaluated in terms of communication costs.

The lifting scheme is an alternative method to compute biorthogonal wavelets. It allows a faster implementation of wavelet transform, along with a full in-place calculation of coefficients.

We suppose the size of signal \( s^j = \{s_{j,l} | 0 \leq l \leq 2^j \} = (c_0, c_1, ..., c_{N-1}) \) is N. When N is even number, we define the sizes of the transformed low-frequency and high-frequency are N/2; but when N is odd number, the size of transformed low frequency is \((N+1)/2\), and high-frequency is \((N-1)/2\). So that, whatever the signal is, the size of transformed low-frequency is \(2^{j-1}\) and high-frequency is \(2^{j-2}\). As a result, the signal \( s^j \) after one wavelet decomposition will get a low-frequency signal \( s^{j-1}\) and a high-frequency signal \( d^{j-1}\). Here, \( s^{j-1}\) and \( d^{j-1}\) are the results of data after low-pass filter and high-pass filter, respectively. Generally, there are three steps in wavelet lifting algorithm: Split, Predict and Update. The whole process showed in Fig. 2 and following steps.

![Fig. 2. Process of combining DSM and WLT](image)

1) Split

The purpose of Split is to decompose \( s^j \) into two subsets \( \text{even}_{j-1} \) and \( \text{odd}_{j-1} \), which also have the correlation. If the correlation between the two subsets is better, we will get the best effect. By using Lazy Wavelet Transform, which means sampling with the data subsets, the signal can be split into two sequences of \( \text{odd}_{j-1} \) and \( \text{even}_{j-1} \). As following:

\[
\text{even}_{j-1} = \{s_{j,2l} | 0 \leq l \leq 2^{j-1} - 1 \} \quad (1)
\]

\[
\text{odd}_{j-1} = \{s_{j,2l+1} | 0 \leq l \leq 2^{j-1} - 1 \} \quad (2)
\]
2) Predict

During the processing of Split, some redundant data can be obtained in the dataset. Thus, our purpose of this step is using Predict again to eliminate the redundant data which has been left after first step. At the same time, it also gives a more compact representation of the data.

For a partial strong correlation signal, its subset is highly correlated. So, no matter what signal we know, odd subset or even subset, we can use one to predict the other within a reasonable accuracy. Normally we use even subset to predict odd subset. General, \( s_{j,2l} \) as the predictor of \( s_{j,2l+1} \)

3) Update

One pivotal characteristic of the low-frequency is that it has the same mean value with original signal. In other words,

\[
s = 2^{-j} \sum_{l=0}^{2^j-1} s_{j,l}
\]

There is no correlation between \( s \) and \( j \), so this ensures that the transformed coefficient \( s_{0,0} \) is the mean value of original signal. Then Update can guarantee this property.

The three steps mentioned above are equal to the level of signal wavelet transform, and then signal will decompose into low-frequency \( s^{l-1} \) and high-frequency \( d^{l-1} \).

\[
s_{j,2l} = \frac{s_{j-1} - d_{j-1}^{l-1}}{2}
\]

\[
s_{j,2l+1} = d_{j-1,l} + s_{j,2l}
\]

On the contrary, the inverse transformation algorithm of \( s^l \) can be approximated by the following description:

\[
s^l = \text{Merge}(\text{even}_{j-1}, \text{odd}_{j-1})
\]

The operations of (4)\(\text{(5)}\)\(\text{(6)}\) is calculated in-place, and the virtual codes can be described as following steps:

\[
\text{(even}_{j-1}, \text{odd}_{j-1}) : \text{Split} ( s^l )
\]

\[
\text{odd}_{j-1} = I ( \text{even}_{j-1} )
\]

\[
\text{even}_{j-1} = U ( d^{l-1} )
\]

Where U is step of Update, and P is step of Predict. After the steps, we can get the Forward Wavelet Transform. But for the inverse wavelet transform, only we need to do is changing the plus or minus sign. The operation of virtual codes can also be performed in three steps:

\[
\text{even}_{j-1} = U ( d^{l-1} )
\]

\[
\text{odd}_{j-1} = I ( \text{even}_{j-1} )
\]

\[
s^l = \text{Merge}(\text{even}_{j-1}, \text{odd}_{j-1})(\text{inverse wavelet transform})
\]

C. The data selecting mechanism

When we use VGRT to organize the WSNs, there are \( N \) nodes in a cluster: the ring is consisted of \( s_0, s_1, s_2, ..., s_{n-1} \). After assuming the dataset \( c_i \) is stored in the \( i \)th node ( \( s_i \) ) of ring, the collecting data of this cluster is represented as a vector \( d = (c_0, c_1, ..., c_{n-1}) \). \( d \) is different from other vector. Upon the VGRT model, the first element \( c_0 \) is next to the last element \( c_{n-1} \) in vector \( d \). Whether the sensing data changes or not, all nodes use wavelet transform to compress the sensing data and transmit it with the traditional transmission mechanism.

By using the VGRT module based on wavelet lifting algorithm, we can eliminate the spatial correction in sensing data. However, there still existed the temporal correction in sensing data.

In this paper, we design a new data transmission mechanism—DSM to reduce the temporal correction. In the initial state of the WSNs, every node will participate in the WLT in order to compress and transmit the sensing data to the centre node, and we can get the useful information by processing the compressed data at the control terminal. Using this method, it will decrease the redundant data from its source, cut down the data needed for transmission. Then, sensor nodes will select data between the new gathering data and the last data. If the differences between data are satisfied with particular threshold, this node will not participate in WLT, it only has the function of routing. At the same time, this node transmits the particular signal (i) to next node.

Sensors need to process data by DSM after gathering data, as seen in Fig. 2. If the sensing data of the node changes a little or never change, plays a role of transmitting data. And if data changes over the threshold (\( \Delta I \)), we need to use the WLT to process the data in this system when the internal networks also need transmitting data, then update the stored data. The designed process of DSM can be seen in Fig. 3.
IV. ENERGY ANALYZE MODEL

In order to fairly compare with the proposed algorithm, an energy model that takes into account both local processing and transmission costs has to be defined. Energy consumption for both process and transmission is highly dependent on the processor being use, we take an example that have been detailed in [12]. In this model, the energy analyzed with the transmission and reception of a k-bit packet over a distance D is

The energy of transmission \( E_{T_x} \) can be described as

\[
E_{T_x} = E_{elec} \cdot k + \epsilon_{amp} \cdot k \cdot D^2
\]  

(7)

and the received energy \( E_{R_x} \) can be defined by following equation

\[
E_{R_x} = E_{elec} \cdot k
\]  

(8)

where \( E_{elec} = 50nJ/b \) is the energy dissipated to run the transmit or receive electronics, and \( \epsilon_{amp} \approx 100pJ/b/m^2 \) is the energy dissipated by the transmission power amplifier.

Considering of the WLT that have been take into our work, we calculate the energy consumption of data processing and compression. Then \( E_p \) can be described as

\[
E_p = NCV_{dd}^2
\]  

(9)

where \( N \) is the number of clock cycles per transformation, \( C \) is the average capacitance switched per cycle, and \( V_{dd} \) is the provided voltage by sensor nodes.

The total energy \( E_{total} \) consumed at each sensor node will be split into three main components:

\[
E_{total} = E_{T_x} + E_{R_x} + E_p
\]  

(10)

Based on the VGR model, we analyze the energy dissipation in Fig. 4.

Fig. 4. The data streams in the WSNs

Additionally, we use \( a \) byte to represent the sensor data (such as temperature, moisture, pressure). The tag of node represented as \( b \) byte, also, \( b \ll a, b \approx i \), where \( i \) is the particular data mentioned above.

If the probability of \( |C| > \Delta t \) in X nodes is \( \alpha \) and the total byte of data is \( Y \), so the first receiving bytes \( Y_1 \) of sink node can be computed as

\[
Y_1 = K \cdot N \cdot \alpha
\]  

(11)

Where \( K \) is the wavelet transform coefficients, \( \Delta t \) is the threshold, and \( |C| \) is the differences between sensing data.

After all, the receiving bytes \( Y_2 \) will be changed as the following equation:

\[
Y_2 = K \times (X \times \alpha \times a + (N - X)(1 - \alpha) \times b)
\]  

(12)

When \( \alpha = 1 \), \( Y_2 = Y_1 \), and if \( \alpha \ll 1 \), we have \( Y_2 \ll Y_1 \). At this time, the amount of data transmitting and receiving will be greatly reduced. So, these are important theoretical foundations for our work.

Finally, the simulator computed compression ratio over all compressed data by using the following equation:

\[
\text{compression\_ratio} = \left(1 - \frac{\text{compressed\_size}}{\text{uncompressed\_size}}\right) \times 100\%
\]  

(13)

V. RESULTS

In this section, we evaluate the impact of the WLT with DSM to reduce the data streams. Our results are based on the MATLAB simulator. Based on the data from agriculture domain as this work is supported by “China N Agriculture T F”, and in order to simulating the WSNs, we use coordinate data to randomly generate 100 nodes for the ring, and there are 8 units of power in each node. Then, we suppose the distance between adjacent nodes in the ring is 5 meters. The temperature data used in the experiment is between 20 degree and 40 degree, and the threshold (\( \Delta t \)) is 2 centigrade. For the sake of simplicity, we set 200 seconds runtime and 100 times test to contrast with other algorithms. At last, take the mean value as the final experimental results.

The energy consumption contrast WLT with traditional Huffman encoding algorithm can be seen in the Fig. 5. We calculate the energy consumption of total nodes in Fig. 6, and then take the mean value as the average energy consumption. The figure shows the energy consumption of traditional Huffman encoding algorithm is 5.8 units, and WLT is 5.2 units. It occurs because the WLT is a faster implementation of the wavelet transform and allows a full in-place calculation of coefficients. Even if the calculation of energy consumed play a little part in total energy consumption, but it is still a great significance in the energy optimization.

![Performance Curve Contrast WLT with Huffman](image)

Fig. 5. Energy consumption contrast WLT with Huffman
Considering the WLT with DSM, the transmission of energy consumption will be decreased for the reduction of data streams. The fact can be illustrated in Fig. 7, in which analysis the energy consumption compares the combination of WLT and DSM with Huffman encoding. The energy consumption of total nodes shows in Fig. 8. In our implementation, energy consumption reduced to 4.7 units. Two reasons are concluded from the results: (i) the advantage of the WLT algorithm is that it allows a faster implementation and lower calculation than Huffman algorithm; (ii) each node do not need to send the whole processing data, just transmit the particular signal (i) to next node, which means the amount of data transmitting and receiving will be greatly reduced, as detailed in Section 3 and Section 4.

In the different way, we analyze the compression ratio of Huffman encoding and WLT with DSM algorithm. The proposed algorithm has a higher compression ratio than Huffman, as showed in Fig. 9. In such cases, Huffman encoding hasn’t considered the temporal and spatial correlation in sensing data, but WLT allows a faster implementation of wavelet transform and equals to half calculation of DWT. Additionally, compressing data with DSM, nodes will transmit less data streams. So our algorithm has improved the efficiency of compression ratio.

VI. CONCLUSION

WSNs are energy constrained, and the extension of their lifetime is one of the most important issues in the design of such networks. Usually, the WSNs collect a large amount of data from the environment, and it is inevitable that the temporal and spatial correlation in sensing data. So we need to reduce the redundant data in order to achieving the effect of saving energy.

In our work, we used the multi-cluster sensor networks based on VGRF, and a WLT with DSM algorithm is proposed and evaluated. This algorithm used lifting schemes to allow a full in-place calculation of coefficients. The main contribution of this algorithm is the capability of reducing energy consumption by decreasing transmission data streams. The results show that not only the data from a source can be reduced, but also energy for transmitting can be reduced. It prolongs the lifecycle of WSNs. However, we intent to use the information of the oldest data compression algorithm to decrease the energy consumed in WSNs.

ACKNOWLEDGMENT

The research is supported by the China International Cooperation Program (No. 2013DFR11020) and China
National Agriculture Technology Funding (No. 2013GB2E200394).

REFERENCES