

Hierarchical Data Gathering Scheme for Energy Efficient Wireless Sensor Network

Jinping Chen and Zinan Chang*

*School of Computer Engineering, Jinling Institute of Technology, Nanjing 211169,
China*
*cjp@jit.edu.cn, *changzn@jit.edu.cn*

Abstract

Compressive Data Gathering (CDG) has recently emerged to provide energy efficient data aggregation in wireless sensor networks (WSNs). Existing CDG is done by forming a span-tree from all sensors to the data sink, which leads to a lot of energy consumption in maintain this tree. In this paper, we present a clustering based CDG scheme for large-scale WSNs. This scheme reduces traditional network communication cost by introducing in-network compression and recovers sensory data accurately by using compress sensing theory. With extensive simulation, we demonstrate that our scheme prolongs nearly 45% network's lifetime compared with the Hybrid compressive data gathering schemes.

Keywords: *Compressive data acquisition, WSN, Hierarchical, Energy efficient*

1. Introduction

Wireless sensor networks (WSNs) are composed of micro sensors which have capabilities of perceptual computing and communication. By means of cooperation between network nodes in Ad hoc way, WSNs can monitor and gather all kinds of information of environment or monitoring objects in network distribution fields in real time. Data collection and transmission are main tasks of WSNs [1, 2]. Information detected by nodes is transmitted to the sink node by wireless transmission way, and it is reconstructed in the sink node so as to restore the detection data of each node.

Multi-hop routing methods are mostly applied in traditional data transmission schemes. Each sensor node data is forwarded to the sink node through the upstream nodes on the path hop by hop along the routing tree. In this transmission mode, nodes closer to the sink also need to transmit information of other distant nodes apart from sending their detection data, so that loads of the closer nodes are higher. The transmission load imbalance brings about a consequence that the energy of sensor nodes near the sink node is the first to deplete, resulting in the network failure. In addition, the data transmission method does not make full use of the spatial correlation of detecting data to compress, resulting in the transmission efficiency of the node low and the energy consumption of wireless communication large.

The design of data acquisition and transmission scheme in WSNs needs to restrict wireless transmission bandwidth, energy constraints and the correlation of detecting data between nodes combining node capacity. However, the existing schemes still have many shortcomings. Literature [3] proposes compressive data gathering scheme (CDG) which uses the distributed code to compress the data correlation. Although it can achieve the load balance of network nodes, but because each node transmits the same amount of data, the transmission quantity cannot effectively reduce in the 2-D network. Literature [4] points out the shortcomings of CDG and proposes the Hybrid-CS solution to save the transmission quantity, which is similar to compressed sensing (CS) clustering scheme in literature [5]. But the two schemes cannot effectively reduce the overhead of time slots.

Literature [6] presents a data transmission scheme, which uses compressed sensing to compress data firstly, then applies and simulates network coding. But these two methods are independent with each other, which do not make full use of the characteristics of the two algorithms to combine them to further improve the efficiency of signal processing simultaneously.

In literature [7], Luo C. et al first apply compressed sensing to large-scale multi-hop wireless sensor network data collection. This method can effectively reduce the traffic in the network and maintain good load balancing that can prolong the network operation time. Literature [4] points out that the dense random projection cannot bring much performance improvement for networks, and the author also gives a data collection strategy of hybrid compressed sensing.

However, those existing CS based data gathering methods mostly adopt tree-type routing scheme which requires a large number of sensor nodes to participate in each CS measurement gathering. Too many sensor nodes participating in each CS measurement gathering leads to two problems [8]:

- (1) The communication cost of entire sensor networks is still high.
- (2) Each CS measurement is easy to be damaged due to sensor network is error-prone.

In fact, communication strategy also plays a role in network data gathering.

For sparse random projections data gathering can decouple routing from compression. In this paper, we have set up a larger-scale compression sampling scheme based on clustering. Meanwhile, we also design an efficient inter cluster communication time with slot switching optimization method.

The rest of this paper is organized as follows. In section 2, we supply the background knowledge of compress sensing. In section 3, we introduce existing compressive sensing and random projection data gathering firstly. The proposed cluster scheme and transmission optimization are given next. The simulation results are shown in section 4. Finally, we give our concluding remarks in section 5.

2. Background Knowledge of Compress Sensing

The basic idea of compressed sensing is conducting compressing and sampling at the same time. Firstly, collect the non-adaptive linear projection of the signal, namely the calculation of the signal measured value; then according to the reconstruction algorithm reconstruct the original data with the measured value [9]. If the value of the measured data is smaller than the original data, the amount of the compressed data can be obtained. Compressed sensing theory mainly includes signal sparse representation, the measurement of matrix designing and the reconstruction of signals. The prerequisite of compressed sensing is that the signal is sparse or in some conversion the signals can be represented sparsely. That is to say, a signal or in the case of its conversion, when k elements are non-zero, the signal is k -sparse, or in this conversion, it is k -sparse.

In general, the common signal in time domain is non-sparse. So, in order to apply compressed sensing, it is first to transform the signals, so as to find the sparse domain of signals. The general signal is sparse in the wavelet basis, Fourier basis and DCT basis. The mathematical description is: make the N -dimension real signal, $x \in R^{N \times 1}$, in an orthogonal basis $\{\psi_i\}_{i=1}^N$ (ψ_i is N -dimension column vector) expand, i.e.:

$$x = \sum_{i=1}^N \theta_i \psi_i \quad (1)$$

In formula (1), $S_i = \langle x, \psi_i \rangle = \psi_i^T x$. Coefficient can be written in matrix form, that we can get:

$$x = \Psi S \quad (2)$$

In formula (2), $\Psi = [\psi_1, \dots, \psi_N]$ is the orthogonal matrix (meeting $I = \Psi\Psi^T = \Psi^T\Psi$); expand coefficient vector $S = [S_1, \dots, S_N]^T$. It assumes that x is k -sparse, namely existing k ($k \ll N$) non-zero elements in S .

A measurement matrix is a way to get the signal; its design is the subject of compressed sensing, which can be interpreted as the signal acquisition in the measurement matrix physically. The signal is compressed during projection, and the measurement matrix is to find a measurement matrix $\Phi \in R^{M \times N}$ ($M \ll N$) satisfying the condition of RIP (restricted isometry property), executing a compression measurement for the signal x , i.e.:

$$y = \Phi x = \Phi \Psi S \tag{3}$$

That we can get M linear measurements $y \in R^M$, the small amount of linear projection includes enough messages to reconstruct signal x . The whole procedure is shown in Figure 1.

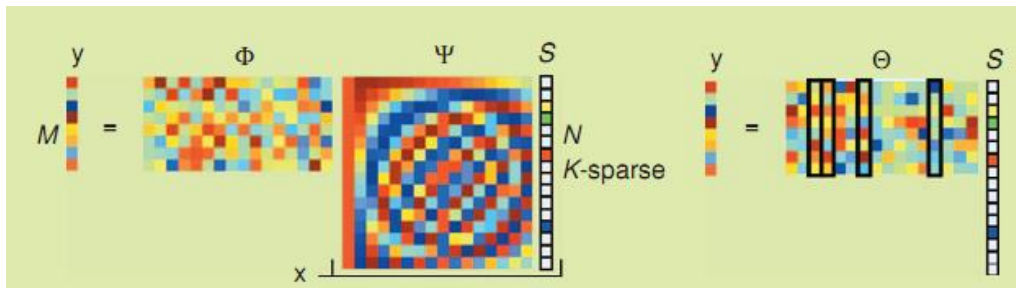


Figure 1. Components of Compressive Sampling

Signal reconstruction is to recover the original signal with a known measured value y and measurement matrix Φ . It is a problem of solving to recover x from y . From the point of view of equation (3), it seems to be impossible; because it is a ill conditioned equation whose unknown number is greater than the number of equation, existing infinitely many solutions. But take equation (2) into equation (3), making information operator as CS $A^{CS} = \Phi\Psi$, that we can get

$$y = \Phi\Psi S = A^{CS} S \tag{4}$$

Although the recovery of S from y is also an ill-conditioned problem, but because of the sparse coefficients, the number of unknowns is greatly reduced, making the signal reconstruction possible. The measurement matrix Φ satisfies the RIP condition, then it can be shown that x can be exactly reconstructed by solving the optimal l_0 norm problem, i.e.:

$$\tilde{x} = \arg \min \|x\|_0 \quad s.t. y = \Phi x \tag{5}$$

In formula (5), $\|\cdot\|_0$ represents l_0 norm, namely the number of non-zero elements in vector quantity x . While it is also an ill-conditioned problem to solve the minimization under l_0 norm. According to the convex optimization theory, it can approach the solution of l_0 norm with loosening the constraints by the optimization of l_1 norm. It can be proved that, in the case of the signal with sparsity, the optimal point of the objective function in constraint space between l_0 and l_1 is almost the same. So, use l_1 norm to reconstruct and optimize in the engineering application, as shown in formula (6):

$$\tilde{x} = \arg \min \|x\|_1 \quad s.t. y = \Phi x \quad (6)$$

3. Clustering Based Compressive Data Gathering (CDG)

3.1. Existing Typical CDG Scheme

Existing typical CDG scheme is based on the tree topology, mainly including the comparison between plain compressed sensing (Plain-CS) and hybrid compressed sensing data collecting strategy. What is shown in Figure 2 are the methods of two kinds of data collecting, in which the gray node represents compressing node with compressed sensing; white node expresses non-compressed node is only responsible for sending its information or transmitting received information; the link label denotes the number of the transmission packet in the process of a round of data collection. We can see from Figure 2 that the node energy consumption in Non-CS is unbalanced, and the transmission quantity of node near to the sink node is very large, while the leaf node is less. Literature [4] points out that, Plain-CS scheme (Figure 2b) cannot reduce total transmission quantity effectively. When the transmission quantity is less, we use traditional non-CS transmission scheme. When the transmission quantity goes over the threshold, it can reduce transmission quantity to use hybrid-CS scheme (Figure 2c) of CS transmission scheme effectively. This paper takes LEACH [10] as an additional comparative object, which is a most representative classic clustering algorithm, and other clustering route algorithms are almost the extensions of LEACH.

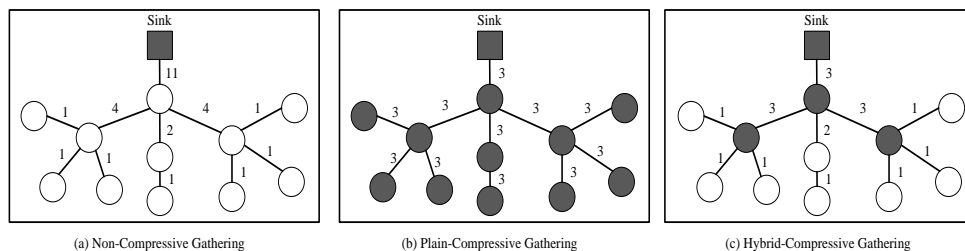


Figure 2. Typical CDG Based on Tree Topology

3.2. Clustering CDG Scheme

The system working manners are as follows: each sensor node periodically transmits signals with T sending period, and nodes are independent between each other without requiring time synchronization. The cluster head periodically collect signal and the collection cycle is also T . Accumulate the signal intensity value received in the period. After the cycle time slice, each sensor node sends the accumulated result to the cluster head, and the cluster head is sent to the sink node with data fusion processing. The sink node operates compressed sensing algorithm, calculating the collected signals selected by nodes. As shown in Figure 3, divide WSN with N nodes into H cell clusters, in which contain n nodes, and $H \times n = N$. In each intra-cell cluster, each node data is sent to Head-cell node through one hop directly; In inter-cell cluster, each Head-cell node transmits to the sink node by multi-hop scheme.

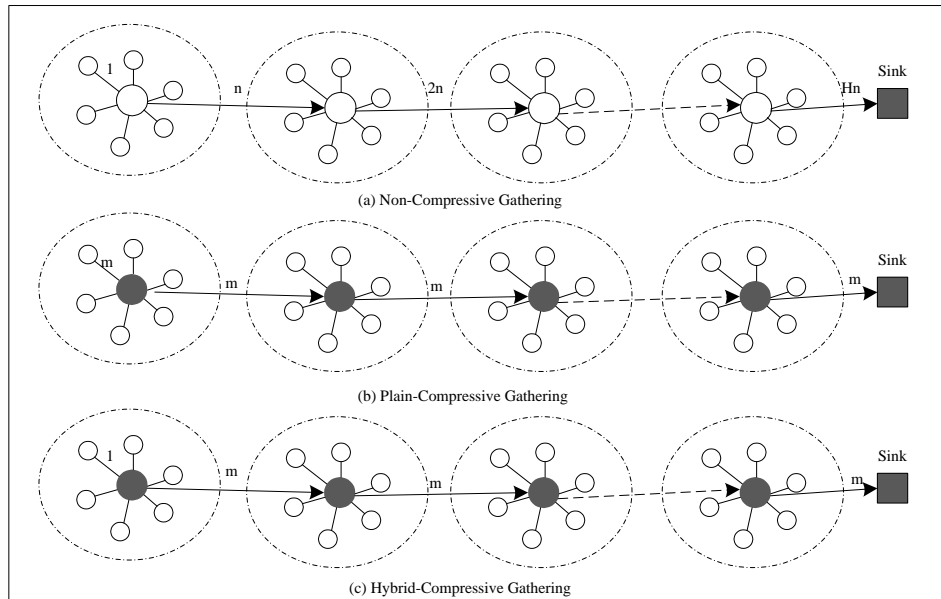


Figure 3. Cluster CDG Scheme

3.2.1. Clustering form Scheme

Because the energy heterogeneity is ubiquitous, different kinds of sensor node deploy different initial energy. In energy heterogeneous WSN, assume that the node energy value is distributed in an interval randomly, so that distinguish between the ordinary nodes and the high-energy nodes. The initial energy of ordinary node is E_0 ; that of the high-energy node is $E_0(1 + M_i)$; M_i is the energy multiple of the high-energy node over the ordinary node. Therefore the initial energy of the nodes can be described as a random distribution in the closed interval $[E_0, E_0(1 + M_{\max})]$. Suppose that the initial energy of the arbitrary node is $E_0(1 + M_i)$, then the total initial energy of the energy heterogeneous network is:

$$E_{total} = \sum_{i=1}^N E_0(1 + M_i) \quad (7)$$

Due to the different equipped initial energy of each node, the cluster head election must take into account the residual energy of nodes; the node with more residual energy has a higher probability of becoming a cluster head. So P_i , the weighted probability of each node being selected as the cluster head, is showed as the following formula (8); among them, P_0 is the percentage of the first round cluster head in all nodes, namely the probability that the first round node is selected as the cluster head.

$$P_i = \frac{P_0 N(1 + M_i)}{N + \sum_{i=1}^N M_i} \quad (8)$$

LEACH, Low Energy Adaptive Clustering Hierarchy, is the most classic energy efficient clustering routing protocol in WSN proposed by Heinzelman *et. al.* Its algorithm elects the cluster head node through circulation to divide the whole network energy load uniformly distributed to each node, so that improve the energy efficiency and prolong the network life cycle. In order to balance the energy consumption of each node, the cluster

head is randomly elected periodically in accordance with turns, each round of the election method is: each node generates a random number between 0 and 1. If the number is less than $T(n)$, the calculation formula of the node becoming the cluster head $T(n)$ is as follows:

$$T(n) = \begin{cases} \frac{P}{1 - p \times \left(r \bmod \frac{1}{p} \right)} & n \in G \\ 0 & other \end{cases} \quad (9)$$

In Formula (9), P is the percentage of the number of cluster heads in the total number of nodes; r is the current election round number, G is the node set of non cluster head nodes in recent $1/p$ round. The cluster head node broadcast the news in wireless channel. The other nodes choose the cluster head node of which the signal is the strongest to join and transfer data to the cluster head through one hop communication. The cluster head also transmits the aggregated data to the sink node through one hop communication. The protocol uses the way of random cluster head election to avoid the excessive consumption of energy of cluster head, that improves the survival time of the network; and the data fusion can effectively reduce the amount of traffic. But clustering protocol aims to data fusion, still using one hop communication. Despite of the small transmission delay, but it requires that the node has larger power communication ability, which is not suitable for large-scale network; even in small-scale networks, it can lead to shorter survival time for the ordinary nodes far from the sink node to use high power communication, and the communication traffic caused by the frequent cluster head election takes the energy.

In order to solve the problem of node load imbalance in case of energy heterogeneous WSN, considering the current energy of node in the process of clustering, so that the node with high energy at present has more chances to become the cluster head. The clustering system model of WSN is showed as Figure 3.

Firstly, promote calculation formula $T(n)$ that in isomorphism LEACH to Multilevel heterogeneous WSN, i.e., substitute formula (8) into formula (9), that formula (10) can be got.

$$T(n_i) = \frac{p_o N(1 + M_i)}{(N + \sum_{i=1}^N M_i) - p_o N(1 + M_i) \times r'} \quad (10)$$

In formula (10), $r' = r \bmod \frac{N + \sum_{i=1}^N M_i}{p_o N(1 + M_i)}$. Next, taking energy factor into account,

The product of energy mean value and threshold size can be considered as multi-attribute optimization method.

$$T(n_i) = \frac{p_o N(1 + M_i)}{(N + \sum_{i=1}^N M_i) - p_o N(1 + M_i) \times r'} \times (E_r + (r \bmod \frac{1}{p_i}) \Delta E_c) \quad (11)$$

Thereinto, E_r is the remaining energy percentage; ΔE_c is energy loss percentage; r represents the rounds of the node not being the first level cluster head continuously. Once elected as the first level cluster head, r is reset to 0. The improvement of formula (11) effectively solves the shortcomings of calculating algorithm $T(n)$ in LEACH protocol. So that the node whose energy consumption ratio is low is preferential to be the first level cluster head. And comprehensively consider the influence of node energy and the threshold size on the election of the first level cluster head, which makes the algorithm more fair and reasonable.

3.2.2. Clustering Transmission Scheme:

The transmitting steps are showed as Figure 4.

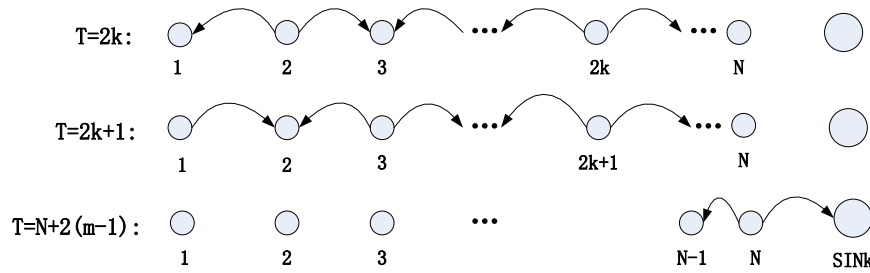


Figure 4. The Transmission Time Slot

1) Node N_1 is the first node. In the first time slot, send the product of its data and the elements corresponding to the measurement matrix, namely $S_1^1 = \psi_1^1 X_1$; in the second time slot, receive the data from node N_2 , and in next time slot, transfer the sum of its data and received data, namely $D(N_2) + \Psi_1^j X_1$. $D(N_2)$ is the data of node N_k in previous time slot. $\Psi_1^j X_1$ is the retransmission data of node N_i at j th time. N_1 empties data after sending.

2) Node N_D is the last node. When the pre-node sends data, node N_D receives data and takes away the data saved in the previous time slot. And store it into the data to be forwarded in the next time slot. Then transmit the data to the sink node in the next time slot.

3) Other intermediate nodes receive the information of their neighbor nodes, and send the sum of its data and the received data in the next time slot, i.e., when node $N_i + 1$ and node $N_i - 1$ send information, the received information of node N_i is $D(N_{i-1}) + D(N_{i+1}) - D(N_i) + \Psi_i^j X_i$, and transmit the data in the next time slot.

4) Each node stops forwarding data after completing retransmission of its data for m times and it can start to transmit the next data for first time after 2 time slots.

5) The odd or even node of each slot sends data, it receives the added value transmitted from its adjacent nodes.

6) In the process of information transmission of each node, the total transmission quantity of CS in ANC scheme is unchanged, still being $m * D$ in which m is the number of retransmission and D is the link depth, but the time slot overhead is changed from $m * D$ in the traditional CS scheme into $D + 2 * (M - 1)$. When $m \ll D$, the used time slot is about one-tenth of that of CS transmission scheme.

Table 1. Transmission and Time Slots Cost of 4 Schemes

Scheme	Transmission and TSC	
	Transmission	TSC
$N_0 - CS$		$[N - \beta] + \left[\frac{\beta + (1 + \beta)}{2} a \right]$
	$[a - 1] + \left[\frac{\beta(1 + \beta)}{2} a \right]$	
<i>Plain - CS</i>	$[m(N - \beta)] + [m\beta]$	$[m] + [m\beta]$

<i>Hybird – CS</i>	$[N - \beta] + [m\beta]$	$[a - 1] + [m\beta]$
<i>ANC</i>		$[N - \beta] + [m\beta]$
$[a - 1] + [\beta + 2(m - 1)]$		

4. Simulation Analysis

The network simulation software NS2 and MATLAB tools are used to simulate the energy heterogeneous WSN in this paper. Multilevel cluster head uses compressed sensing for data fusion; measurement matrix uses the Gauss stochastic matrix; and the reconstruction algorithm uses the OMP algorithm. The experiment simulates a square area of $L * L (L = 128)$. The sink is located in the center of the region, and in order to eliminate the influence of node deployment on the experiment, we deploy N , $64 * 64$, sensor nodes uniformly in the region. So each square cell, $4 * 4$ square region, has a sensor node. So that each $4 * 4$ cell contains one node exactly. For the transmission range R of the nodes we adopt a unit disk model, i.e., nodes can only communicate with all other nodes placed at a distance less than or equal to R . We use $R = \sqrt{5L^2 / N}$ as this guarantees that the structure is fully connected under any deployment of the nodes. For the energy of nodes, In addition to the base station energy that is not limited, the energy of the remaining sensor nodes is limited and the initial energy is not the same which is distributed randomly in the interval $[E_{init}, (1 + K \max) E_{init}]$. For the network measurement data, we use ocean surface temperature data of Florida coast [11], which is shown in figure 5, and we use the Signal-to- Noise- Ratio (SNR) to measure the reconstruction error. If x is the original measurement and \hat{x} is the reconstructed vector, then $SNR = 20 \log \frac{\|x\|}{\|\hat{x} - x\|}$. Select the simulation parameters in detail, as shown in Table 2.

Table 2. The Parameter Settings of Network Simulation Experiment

Parameter	Value	Parameter	Value
Network range	$400 * 200m^2$	E_{elec}	50nj/bit
Carrier frequency		2.412GHz	E_{start}
250nj/pack			
Signal sensitivity	-84dBm	Transmission range	10m
Max transmitted power	100mW	Signal amplifier multiple	2
Length of a packet	500bit	Path attenuation factor	2.0
Initial energy	[2,4]J	The MAC layer protocol	
CSMA/CA			

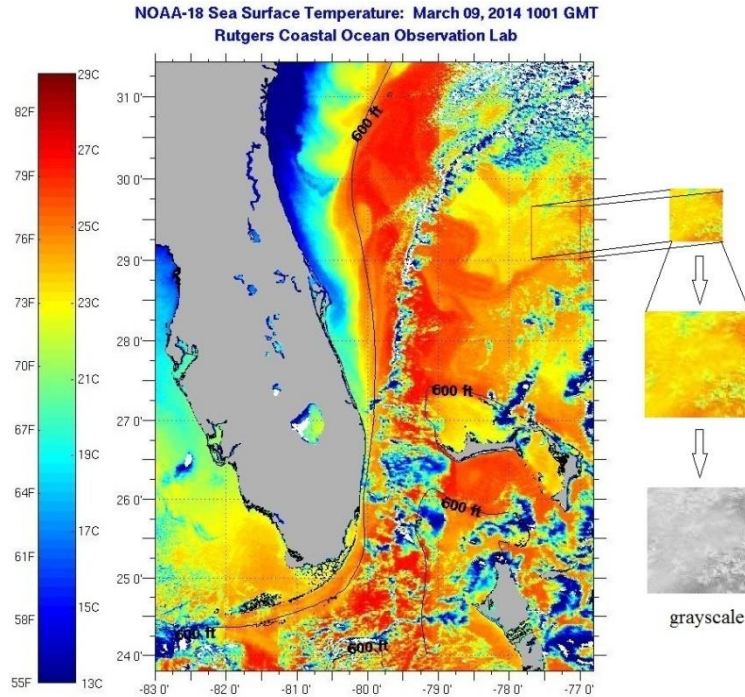


Figure 5. Ocean Surface Temperature Data of Florida Coast

The network life cycle is an important basis for evaluating the effectiveness of the energy of an algorithm. The simulation results in Figure 6 show the observed protocol performance by changing the proportion α of the number of high-energy nodes in that of the total nodes and the multiple M of the energy of high-energy nodes over that of the ordinary nodes using the proposed protocol. Figure 2 shows the life cycle from the beginning to the death of the first node, when the proportion α increases from 10% to 90% and multiple M increases from 1 to 9. It is observed that, LEACH protocol doesn't use the energy brought by the increase of α and M effectively. The lifecycles of the networks using the LEACH and Hybrid protocol remains almost unchanged, leading to the waste of high-energy node energy. So that the two protocols are not adaptive to the heterogeneous WSN. All nodes are considered as having the same energy, which doesn't consider the differences of node energy. While the proposed protocol makes full use of the high energy of the cluster head due to the mechanism based on the residual energy, so that the network life cycle greatly enhances.

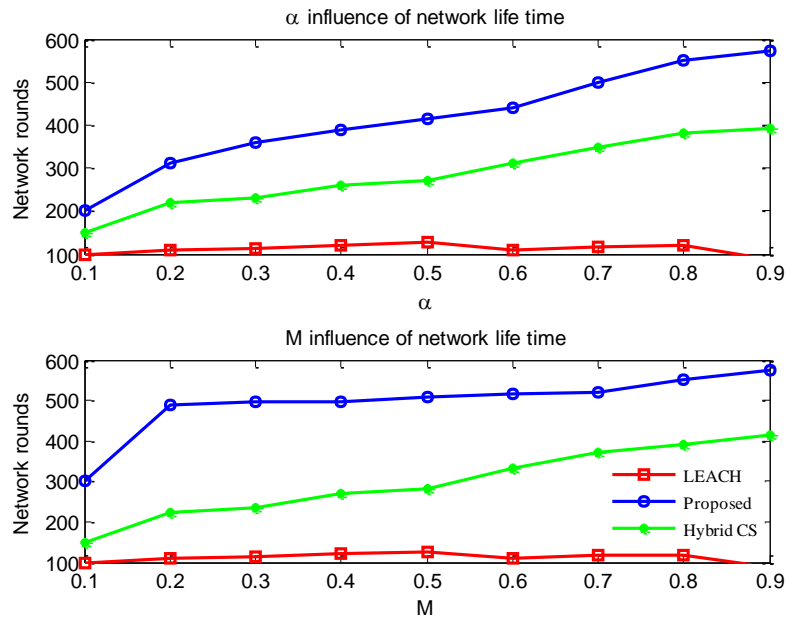


Figure 6. Ocean Surface Temperature Data of Florida Coast

The quality of the original data restored by the sink node is decided by the received packet number (the number of observations). Each data packet is regarded as a projection of the original data. But when reconstruct the original signal with compressed sensing recovery algorithm; the number of random sample will affect the quality of signal recovery. We simulate the signal recovery quality in the process of receiving all packets at the sink, as shown in Figure 7. As can be seen from the figure, when the number of the observation packets is 32, the signal to noise ratio (SNR) reconstructed suddenly increases greatly. But before the number of the observation packets is less than 32 (exactly 25), the recovery effect is very poor (SNR is not more than 15dB). That is because the number of the received observation data packets is less than the minimum threshold of restoring the original information. When the number of the projection data packets exceeds this threshold, the reconstruction effect does not obviously increase. When the observation number is 32, the SNR value reverts to 614.75dB, and when the observation number is 38, SNR is 619.7074.

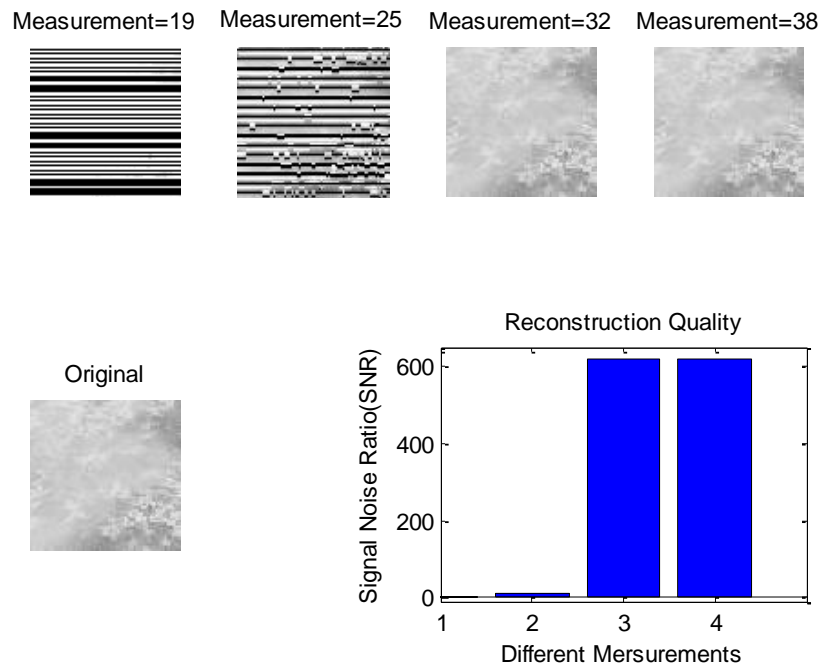


Figure 7. Reconstruction of Ocean Surface Temperature by Different Measurement Packets

5. Conclusion

Based on the thought of compressed sensing, this paper proposes an energy heterogeneous clustering routing protocol, and designs a weight coefficient based on normal distribution, so as to optimize the signal reconstruction performance brought by compressed sensing in case of the amount of data too little. The network life cycle and the remaining energy changing are compared in the classic LEACH protocol and the Hybrid method with NS2 network simulation software and MATLAB tool. The results show that the algorithm can not only make full use of the heterogeneous energy resources, balance the network energy consumption and prolong the network life cycle, but also can accurately restore the signal source.

Acknowledgements

This work was partially supported by the Natural Science Foundation of Jiangsu Province, China (Grant No. BK2012082), the Natural science fund for colleges and universities in Jiangsu Province, China (Grant No. 14KJB520012).

References

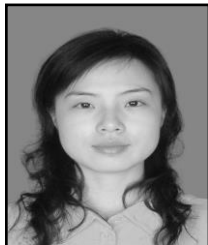
- [1] J. Vicente, R. Rocha and R. Neves, "GolfSense: A golf course WSN monitoring application", Computer Engineering and Systems (ICCES), 2010 International Conference on. IEEE, (2010), Cairo, Egypt
- [2] K.O.O. Bonhyun and S. Taeshik, "Implementation of a WSN-based structural health monitoring architecture using 3D and AR mode [J]", IEICE transactions on communications, vol. 93, no. 11, (2010), pp. 2963-2966.
- [3] C. Luo, F. Wu , J. Sun and C.W. Chen, "Compressive Data Gathering for Large-scale Wireless Sensor Networks", Proceeding of ACM Mobicom'09, (2009), New York, USA.
- [4] J. Luo, L. Xiang and C. Rosenberg, "Does Compressed Sensing Improve the Throughput of Wireless Sensor Networks", Communications (ICC), IEEE International Conference on, (2010), Cape Town, South Africa

- [5] H. Zheng, S. Xiao and X. Wang, "Energy and Latency Analysis for In-network Computation with Compressive Sensing in Wireless Sensor Networks", INFOCOM, 2012 Proceedings IEEE, (2012) March 25-30; Orlando, USA
- [6] S. Feizi and M. Médard, "A power efficient sensing/communication scheme: Joint source-channel-network coding by using compressive sensing", CoRR, (2011), Monticello Israel.
- [7] C. Luo, F. Wu, J. Sun and C. Chen, "Efficient measurement generation and pervasive sparsity for compressive data gathering", Wireless Communications, IEEE Transactions, vol. 9, no. 12, (2010), pp. 3728-38.
- [8] X.G. Wu, Y. Xiong, W.C. Huang, H. Shen and M.X. Li, "An efficient compressive data gathering routing scheme for large-scale wireless sensor networks", Computers and Electrical Engineering, vol. 39, (2013), pp. 1935-1946.
- [9] Candès E. "Compressive sampling", In: Proceedings of the international congress of mathematicians, (2006), Madrid, Spain.
- [10] W.R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", HICSS '00 Proceedings of the 33rd Hawaii International Conference on System Sciences, (2000), Massachusetts, USA
- [11] NOAA Sea Surface Temperature of Florida Coast, last access: http://marine.rutgers.edu/cool/sat_data/product=sst®ion=georgia¬humbs=0, March, 2014.

Authors



Chen Jinping, She is an experimentalist in the School of Computer Engineering at Jinling Institute of Technology. She received her Master degree in Computer Applications from the University of Nanjing Aeronautics and Astronautics in 2012, and her bachelor in Information Management and System Program from the Nanjing agricultural university in 2003. Her research interests include wireless sensor networks and compressed sensing technology, and large-scale simulation of complicated network systems.



Zinan Chang, She received the Bachelor Degree in Computational Mathematics from Nanjing Normal University in 2000, Master degree in Computational Mathematics from Nanjing Normal University in 2006. From 2006 to now, she is a lecture at the School of Computer Engineering, Jinling Institute of Technology and working for Jiangsu Information Analysis Engineering Laboratory. Her current research interests include information security and complex dynamical networks.