Transmission Efficient Clustering Method for Wireless Sensor Networks using Compressive Sensing (Supplementary File)

Ruitao Xie, and Xiaohua Jia, *Fellow, IEEE*

Abstract—This supplementary file contains the supporting materials of the TPDS manuscript - "Transmission Efficient Clustering Method for Wireless Sensor Networks using Compressive Sensing". In this file, we first present the related works. Secondly, we present the analysis on the optimal value of the edge length D of cluster-square and discussions on determining the parameter H in our distributed implementations. Finally, we provide an additional performance evaluation. It includes the comparison between analytical results and simulation results, the simulations on the non-homogenous networks, and iteration times to converge of the iterative algorithm in our method.

1 Related Works

The sensor data in the sensor networks has spatial or temporal correlations. The correlated data is sparse in some transform domain, such as the wavelet domain and Fourier domain [1], [2]. The new technology of CS [3], [4], [5] motivates the investigations on data gathering with CS [1], [2], [6], [7], [8], [9], [10], [11] and target localization based on CS [12] in sensor networks. The data gathering with pure CS in the two dimensional area is conducted along the tree structure in Fig. 1 of the main paper. In the i^{th} round of projection, the node v_j generates a random measurement coefficient ϕ_{ij} (forms the measurement matrix Φ), and computes the data term $\phi_{ij}x_j$. Each leaf node transmits its term to the parent node. Once the parent node receives data from all its descendant nodes, it can add its own data term and all received data terms together and then send it to its upper parent node or the sink node. When the sink node receives data from all its descendant nodes, it adds them to form the i^{th} projection. In this way, the sink nodes collect M projections

$$y = \Phi x,$$

then it can use ℓ_1-norm minimization [13], [14] or other heuristic algorithms, such as orthogonal matching pursuit [15], to recover the original data of sensor nodes.

In real systems, the measurement coefficient ϕ_{ij} can be generated using a pseudorandom number generator seeded with the identifier of the node v_j [1]. Thus, given the identifiers of the nodes in the network, the sink can easily reconstruct the measurement matrix. The measurement matrix could be Gaussian random matrix or Bernoulli random matrix [1]. The number of projections M is determined by the total number of nodes N and the sparsity level of the original data [1]. If the sparsity of sensor data changes over time, the adaptive method in [11] can be used to adjust M. The transform basis is not required in generating the projections, but it is used in the recovery of original data. The discrete Fourier transform or discrete wavelet transform could be used as the transform basis [1].

Authors in [6] analyzed the network capacity when CS is utilized in data gathering and proved that the capacity gain is proportional to the sparsity level of sensor data. The load balance contributes to this capacity gain. Although researchers stated that the total transmission number can be reduced when the number of projections is low enough. However, if the required number of projections increases, for example if the sparsity level of signal is not sufficiently small, then the total number of transmissions may be larger than the case without using CS. Authors in [7] proposed to obtain measurements from spatially-localized sensors, that is to use sparse measurement matrix in CS, and showed that joint reconstruction is better than independent reconstruction with localized projection. Different from their works, we use general measurement matrix and analyze the optimal cluster size to minimize the number of transmissions.

Authors in [2] designed a data gathering scheme, where in each round of projection M furthest nodes away from the sink send their original data directly to one of the remaining nodes which apply the CS. They proved that this scheme can reduce the transmission cost. Authors in [8] researched how large throughput can be achieved with or without CS, or with the hybrid scheme, where compression is only applied on the nodes whose incoming traffic is reduced by

- *Ruitao Xie and Xiaohua Jia are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. E-mail: ruitaoxie2@student.cityu.edu.hk, csjia@cityu.edu.hk.*

the CS. The authors stated that application of CS in the simple scheme may bring no obvious throughput improvement, but application of CS in hybrid scheme can achieve significant throughput improvement. Authors in [9] applied hybrid CS in the data collection and proposed an aggregation tree with minimum energy consumption. The energy consumption of a link in [9] is a function of the link length. In the special case when the energy expense of sending one unit of data across link is 1, the energy consumption to collect one unit of data from each sensor in the network is equal to the number of transmissions.

Motivated by the CS theory, authors in [10] proposed a distributed energy-efficient sensor network scheme: In each frame, a randomly chosen subset of nodes participate in the sensing process and transmit data by randomly accessing the channel. The field reconstruction based on CS is performed at the fusion center with sufficient collision-free packets received. The sparsity of natural signals may vary in temporal and spatial domain. To address this challenge, authors in [11] presented an adaptive data gathering scheme by CS for wireless sensor networks. In their adaptive scheme, the autoregressive model is introduced into the CS reconstruction, and the number of measurements is adjusted according to the variation of the sensed data. Authors in [11] also proposed a novel abnormal readings detection and identification mechanism based on combinational sparsity reconstruction. Authors in [12] applied CS to count and localize targets in wireless sensor networks. In their works, a greedy matching pursuit algorithm is proposed to accurately recover sparse signals with high probability. Moreover, a framework for counting and positioning targets from multiple categories is also proposed.

The previous works use CS method on routing trees. However, clustering methods have many advantages over the tree methods [16], [17], [18], [19], [20], [21], such as fault tolerance and traffic load balancing. In addition, the previous works ignored the locations and node distribution of sensor nodes. While in sensor networks, the information of node distribution can help the design of data gathering method that uses fewer data transmissions [16], [17], [18], [19], [20], [21]. In this paper, we aim at using the information of node locations and distribution to propose a clustering method that uses the hybrid CS for sensor networks.

2 Optimal Value of D

The total number of transmissions of hybrid CS method in cluster structure is:

$$\begin{aligned} T &= \frac{N}{3}(1-\frac{3M}{2N})\cdot D+\frac{N}{3}(\frac{3M}{\lambda a^2}-1)\cdot\frac{1}{D} \\ &= c_1\cdot D+c_2\cdot\frac{1}{D}. \end{aligned} \tag{1}$$

Considering the above Eq. (1), T is a function of D, where D lies in the interval $[1, D_{\text{MAX}}]$.

In the analytical model, λa^2 is the number of nodes in a grid. To allow any nodes in a grid to be able to communicate with each other, a is set as $\frac{r}{\sqrt{2}}$. Thus, when the number of nodes in such small grid is sufficiently smaller than $3M$, $\frac{3M}{\lambda a^2}$ is much greater than 1, that is $c_2 > 0$, and the term $\frac{3M}{\lambda a^2}-1$ is approximate to $\frac{3M}{\lambda a^2}$.

With different M, the best value of D to minimize T is different. Firstly, if $M \geq \frac{2}{3}N$, that is $c_1 \leq 0$, T is a mono-decreasing function of D. T is lowest when D is largest.

Secondly, if $M < \frac{2}{3}N$, that is $c_1 > 0$, T changes in this way: when D is small, as the increase of D, T decreases; when D reaches a certain value, the further increase of D would lead to the increase of T. Thus, there is an optimal value of D such that the number of transmissions is lowest.

From

$$\frac{dT}{dD}=c_1-\frac{c_2}{D^2}=0,$$

we get

$$D|_{\frac{dT}{dD}=0}=\sqrt{\frac{c_2}{c_1}}=\sqrt{\frac{\frac{3M}{\lambda a^2}-1}{1-\frac{3M}{2N}}}. \tag{2}$$

It is approximate to

$$D|_{\frac{dT}{dD}=0}\approx\sqrt{\frac{\frac{3M}{\lambda a^2}}{1-\frac{3M}{2N}}}. \tag{3}$$

From $D|_{\frac{dT}{dD}=0}=D_{\text{MAX}}$, we get

$$M=\frac{2}{9}N. \tag{4}$$

When $M \leq \frac{2}{9}N$, $D|_{\frac{dT}{dD}=0} \leq D_{\text{MAX}}$, it is feasible, thus the optimal value D^* is $D|_{\frac{dT}{dD}=0}$; when $\frac{2}{9}N < M < \frac{2}{3}N$, $D|_{\frac{dT}{dD}=0} > D_{\text{MAX}}$, it is infeasible, thus D^* is D_{MAX}.

Following the above analysis, it is concluded that when M is less than $\frac{2}{9}N$, the optimal value D^* is calculated as in Eq. (5):

$$D^*=\begin{cases}\sqrt{\dfrac{\frac{3M}{\lambda a^2}-1}{1-\frac{3M}{2N}}}, & M<\frac{2}{9}N;\\ D_{\text{MAX}}, & \frac{2}{9}N\leq M\leq N.\end{cases} \tag{5}$$

When $M \geq \frac{2}{9}N$, and in the extreme case when $M = N$, the optimal value D^* is D_{MAX}.

3 Discussion on H

This section discusses how to determine the value of H. Let P_{th} denote the required probability that there is at least one node in an disk with area $\pi(Hr)^2$ (P_{th} is called threshold probability). Given the density λ of the Poisson distribution of sensor nodes and the threshold probability P_{th}, the value of H can be determined as follows. According to the Poisson point process, the probability that a field with area A contains n nodes is,

$$P(n) = \frac{e^{-\lambda A}(\lambda A)^n}{n!}, \text{ for } n = 0, 1, 2, \cdots.$$

Thus, an disk with area $\pi(Hr)^2$ contains at least one node with the following probability,

$$P(n \geq 1) = 1 - P(n = 0) = 1 - e^{-\lambda\pi(Hr)^2} \geq P_{\text{th}}.$$

Therefore, the threshold probability P_{th} can be met if the value of H is larger than

$$H \geq \sqrt{\frac{-\ln(1 - P_{\text{th}})}{\lambda\pi r^2}}.$$

That is, given P_{th}, we can determine the value of H, such that there is at least one node within H hops from the central point of a cluster. For example, when the transmission range r is $\sqrt{2}$ units, λ is 1 and P_{th} is 99.99%, H should be at least 1.21 unit.

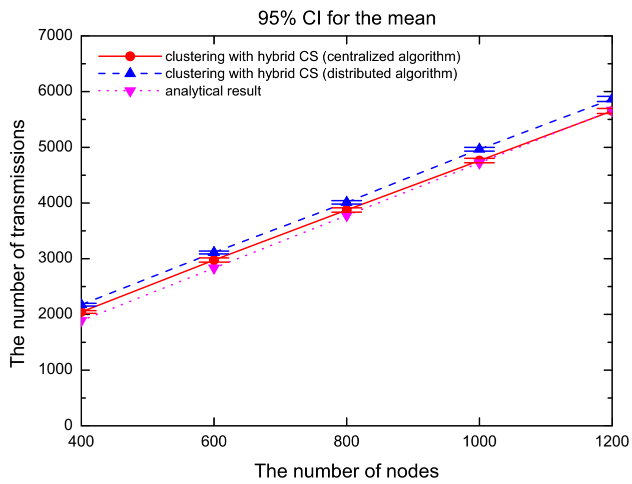

Fig. 1. The number of transmissions in analysis vs in simulation, where the compressive ratio is 10. The bars around the symbols on the lines represent the 95% confidence interval.

4 Additional Performance Evaluation

In Section 6.1 of the main paper, we have shown the simulation metrics and setup in detail.

4.1 Analytical Results versus Simulation Results

Fig. 1 shows the number of transmissions in analysis and in simulation, where the number of nodes varies from 400 to 1200 and the compressive ratio is 10. The number of transmissions $T(D^*)$ in analytical result is calculated from Eq. (7) with D^* from Eq. (8) in Section 3 of the main paper. It is shown that the gap between the simulation result and the analytical result is small. The number of transmissions in simulation conforms to the results obtained from our analytical model. It demonstrates that our analytical model is strong in analyzing the number of transmissions.

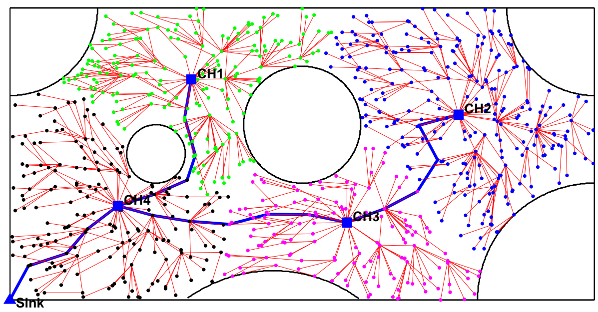

Fig. 2. An example of non-homogenous network in an irregular sensor field. The sensor nodes form 4 clusters.

4.2 Simulations on the Non-homogenous Networks

In this section, we evaluate the performance of our method on the non-homogenous networks in an irregular sensor field. As shown in Fig. 2, the irregular sensor field has 6 void areas in the rectangle field of 20 $\times$ 10 square units. In our simulations, sensor nodes are uniformly and independently distributed in the sensor field. The density of nodes varies from 2 to 6. The compressive ratio is set to 10. The other parameters are set as in Section 6.1 of the main paper.

Fig. 3 shows the number of transmissions of data collection methods in the irregular sensor field. Our method can reduce the number of transmissions significantly compared with other methods. The number of transmission of our method is slightly larger than that of the optimal tree with hybrid CS method.

Fig. 4 shows the reduction ratio of transmissions of our method compared with other methods in the regular and the irregular sensor field. The networks in the regular sensor field is set as presented in Section 6.1 of the main paper. As shown in Fig. 4, in the irregular sensor field, our method has the similar improvements to the case of the regular sensor field. The reduction ratio compared with the SPT with hybrid CS method decreases only 5% in the irregular sensor field, compared with the case of the regular sensor field.

Fig. 5 shows the change of the number of transmissions as the number of clusters increases. The density of nodes is set to 5. From the analysis in Section 3 of the main paper, for the networks in the regular sensor field, it is got that the optimal number of clusters is 3. As shown in Fig. 5, by using the centralized algorithm, the number of transmissions of 3 clusters is near to the minimum number of transmissions.

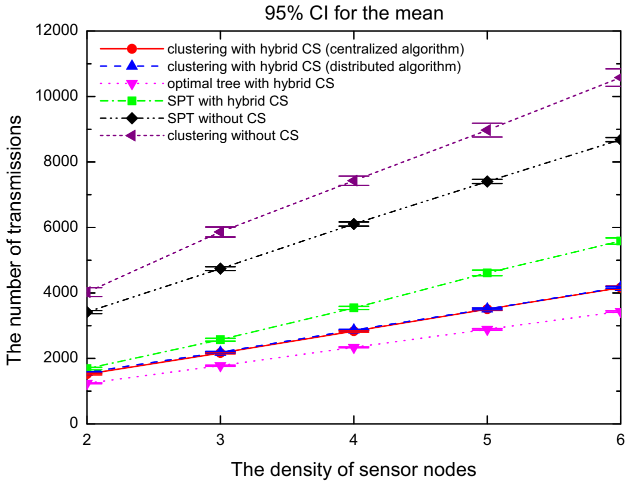

Fig. 3. The number of transmissions of data collection methods in the irregular sensor field.

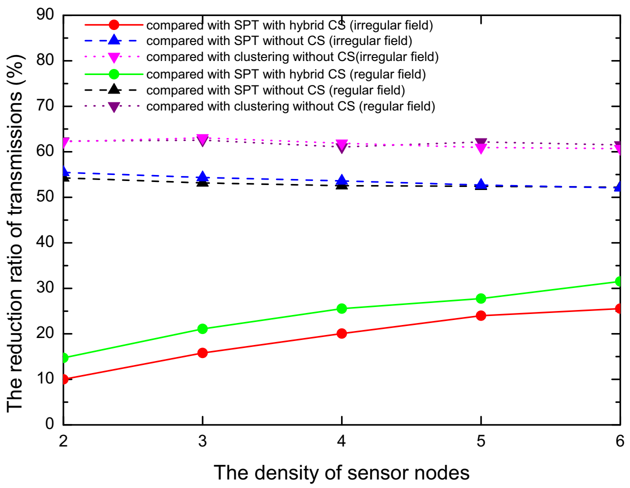

Fig. 4. The reduction ratio of transmissions of clustering with hybrid CS method compared with other methods in the regular and the irregular sensor field.

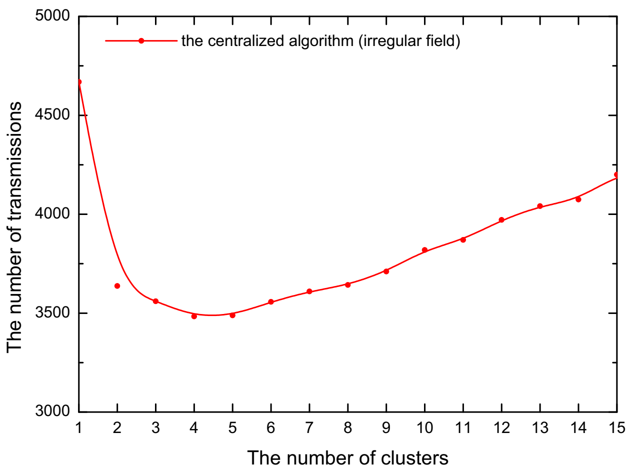

Fig. 5. The number of transmissions for different number of clusters, where the density of nodes is 5.

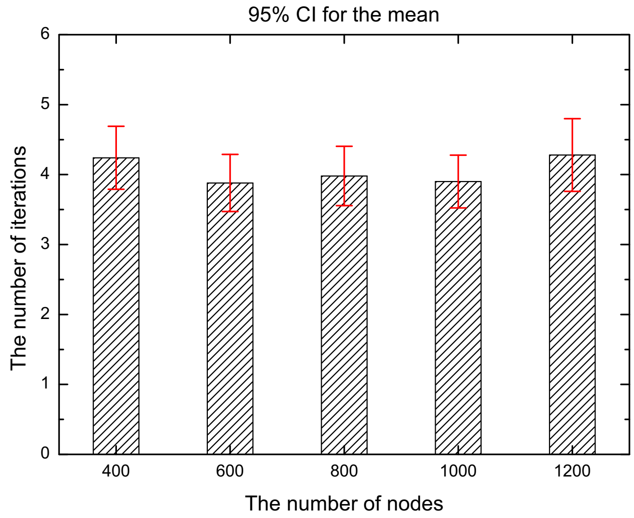

Fig. 6. Iteration times of our iterative algorithm to converge. The vertical lines around the top of bars represent the 95% confidence interval.

4.3 Iteration Times of the Iterative Algorithm in Our Method

In this section, we evaluate the iteration times of the iterative algorithm in Section 4.2 of main paper to converge. The compressive ratio is set to 10. The number of nodes varies from 400 to 1200. We do simulations on the regular sensor field. In our algorithm, the initial set of cluster heads is randomly selected by the sink. As shown in Fig. 6, our iterative algorithm takes 4 or 5 iterations on average to compute the CHs of clusters. In addition, our algorithm converges in the same rate as the number of nodes increases from 400 to 1200. It demonstrates that our algorithm is scalable in large scale networks.

References

[1] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, Mar. 2008.

[2] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Efficient measurement generation and pervasive sparsity for compressive data gathering," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3728–3738, Dec. 2010.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] R. Baraniuk, "Compressed sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[5] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[6] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. ACM MobiCom'09*, Sep. 2009, pp. 145–156.

[7] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Spatially-localized compressed sensing and routing in multi-hop sensor networks," in *GSN '09*, 2009, pp. 11–20.

[8] J. Luo, L. Xiang, and C. Rosenberg, "Does compressed sensing improve the throughput of wireless sensor networks?" in *ICC 2010*, May 2010, pp. 1–6.

[9] L. Xiang, J. Luo, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *Proc. IEEE Sensor, Mesh and Ad Hoc Communications and Networks (SECON'11)*, Jun. 2011, pp. 46–54.

[10] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1660–1670, Sep. 2011.

[11] J. Wang, S. Tang, B. Yin, and X.-Y. Li, "Data gathering in wireless sensor networks through intelligent compressive sensing," in *INFOCOM 2012*, Mar. 2012, pp. 603 –611.

[12] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," in *INFOCOM 2011*, Apr. 2011, pp. 2255–2263.

[13] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[14] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[15] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[16] M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 12, pp. 1844–1856, Dec. 2009.

[17] S. Soro and W. B. Heinzelman, "Cluster head election techniques for coverage preservation in wireless sensor networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 955–972, 2009.

[18] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *IEEE Netw.*, vol. 20, no. 3, pp. 20–25, May-Jun. 2006.

[19] W. Heinzelman, A. 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.

[20] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, Oct.-Dec. 2004.

[21] S. Bandyopadhyay and E. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *INFOCOM 2003*, vol. 3, Mar. 2003, pp. 1713–1723.