Multi-Attribute Compressive Data Gathering

Guangshuo Chen*, Xiao-Yang Liu*, Linghe Kong†*, Jia-Liang Lu*, Min-You Wu*
*Shanghai Jiao Tong University, China
†Singapore University of Technology and Design, Singapore
*{chengs, yanglet, linghe.kong, jlu, mwu}@sjtu.edu.cn, †linghe_kong@sutd.edu.sg

***Abstract*—The data gathering is a fundamental operation in wireless sensor networks. Among approaches of the data gathering, the compressive data gathering (CDG) is an effective solution, which exploits the spatiotemporal correlation of raw sensory data. However, in the multi-attribute scenario, the performance of CDG decreases in every attribute's capacity because more measurements are on demand. In this paper, under the general framework of CDG, we propose a multi-attribute compressive data gathering protocol, taking into account the observed inter-attribute correlation in the multi-attribute scenario. Firstly, we find that 1) the rapid growth of the demand on measurements may decline the network capacity, 2) according to the compressive sensing theory, correlations among attributes can be utilized to reduce the demand on measurements without the loss of accuracy, and 3) such correlations can be found on real data sets. Secondly, motivated by these observations, we propose our approach to decline measurements. Finally, the real-trace simulation shows that our approach outperforms the original CDG under multi-attribute scenario. Compared to the CDG, our approach can save 16% demand on measurements.**

I. Introduction

Wireless sensor networks (WSNs) [3] are widely used by scientist to monitor and interact with the physical environment. With hundreds or thousands deployed sensors, it is able to obtain a full-scale monitoring of the interesting area, *e.g.*, coal mines [11], forests [12] and indoor offices [1]. To get a real-time and precise overview of the target area, the data gathering is a key operation, where a large amount of sensor readings are packed or compressed, and then transmitted to the sink with the requirements of low delay and error rate.

In the network layer, effective global communication cost and economical energy consumption are two major challenges of the data gathering process[13]. The compressive data gathering (CDG) presented by C. Luo at el. [13] is the state-of-the-art approach, which deals with these two challenges both together. The CDG distributes computation and communication costs to all nodes, with the creative usage of the compressive sensing (CS) theory [5]. If the sensory data is sparse, the CDG can provide satisfied network capacity and accuracy. Theoretically, a sink receives compressed readings on sensor data, instead of directly measurements. Thus, the CDG treats the data gathering process as a recovery problem [10][7][9] and utilizes the sparsity of data to reduce communication cost.

In this paper, we are interested in adapting the CDG to deal with the multi-attribute scenario. A stringent restriction of the CS is that one attribute has to be sampled in a certain number of times in order to ensure the accuracy. The total amount of sampling measurements is proportional to the number of attributes, in the multi-attribute scenario. Such a large amount of sensor readings have noticeable impact on communication cost as the number of attributes increases. To the best of our knowledge, the existing discussions on the CDG does not take this into account. However, since multi-attribute scenarios are common in practice, we attempt to adapt the CDG to the multi-attribute scenario.

For the one attribute scenario, the number of measurements provided according to CS [5]. Fortunately, in the multi-attribute scenario, one WSN usually contains several attributes, *e.g.*, temperature, humidity, light illumination, and etc. These attributes are associated with the physical environment, thus they are physically related, such as humidity and temperature. The correlations among them provide the redundancy of information, so that one may obtain enough information for accurate reconstruction with fewer measurements. Therefore, communication cost may be saved by exploiting this new correlation one step further.

In this paper, by utilizing correlations among attributes, we propose an approach base on CDG for the multi-attribute data gathering scenario. Firstly, we analyse the CDG protocol and point out the key issues. Secondly, we propose our approach and analyse why it can reduce the number of measurements theoretically. Thirdly, experiments are made on real data sets. Our contributions are that 1) we propose an effective approach for the multi-attribute data gathering problem and 2) real data-driven simulations are performed, which shows that our approach can apparently reduce the demand of measurements without the loss of accuracy compared to the original CDG.

The rest of this paper is organized as following. Section II gives a review of of the compressive data gathering protocol. Section III describes the original CDG and proposes our approach. Section IV illustrates our observations on real data sets. Section V demonstrates the evaluation of performance. Section VI concludes the paper with discussions about future work.

II. Compressive Data Gathering

The problem of data gathering and collection has been widely studied in recent year. Compressive Data Gathering (CDG) is the state-of-the-art approach. C. Luo at el. proposed CDG [13] for efficient data gathering in large scale monitoring sensor networks. They also discussed efficient measurement generation and pervasive sparsity for CDG [14]. Since then, using compressive sensing theory on the data gathering becomes a popular topic. Xiang at el. discussed

978-1-4799-3083-8/14/$31.00 ©2014IEEE

about compressed data aggregation for energy efficient WSNs [15]. Caione at el. [4] also talks about using compressive sensing to extend lifetime of WSNs.

The CDG exploits compessive sensing to reduce global scale communication cost without introducing intensive computation or complicated transmission control overheads. With the benefit of sparse sensor readings, the CDG can achieve both capacity gain and load balancing.

The CDG is discussed popularly and proved to be a efficient data gathering method. However, it is common that one can like many types of sensor readings from one WSN, *e.g.*, light illumination, humidity temperature in GreenOrbs [12] or sea water salinity and temperature in CTD [2]. In the multi-attribute scenario, the increasing demand on measurements causes the decline of capacity, which weakens the performance of CDG.

Usually one can expect that there are correlations in these attributes[7]. We are interested in using the correlations among attributed to further improve the capacity and accuracy in CDG.

III. CDG under multi-attribute

The advantage of the original CDG approach [13] is two-fold, *i.e.*, 1) reducing global data traffic by compressing sensor readings and 2) prolonging network lifetime by distributing energy consumption. Our approach targets at data gathering problem under multi-attribute scenarios for large scale WSNs. Compared with the CDG, our approach further reduces global traffic by utilizing correlations among attributes and provides equal accuracy with less measurements.

A. Data Gathering Process

Without loss of generality, the same problem statements and assumptions as [13] are adopted as following. Considering a large scale WSN, sensors are deployed in a two-dimensional area. Each sensor are equipped with a battery with limited lifetime. There is only one sink in the network, which contains unlimited computing power and endurance.

Assume that sensors have limited transmission range so that multi-hop data transmission is required. In this case, sensors are connected by chain-type or tree-type topologies to the sink [13]. Since we focus on data gathering and reconstruction in this paper, we just make assumptions that all sensors 1) adopt the same routing strategy and 2) transmit through the corresponding routing tree by self-organization.

Suppose that N sensors are in a particular routing tree and are able to measure J different physical attributes in the same rate. Sensor readings generated by these sensors in one time slot is represented as $x_{i,j}$ where $i \in [1, N]$ and $j \in [1, J]$. Assume that each sensory vector $\mathbf{x}_j = (x_{1,j}, \cdots, x_{N,j})^T$ is a K_j-sparse signal.

The data gathering process in the original CDG approach is demonstrated as following. Each sensor generates M_j measurements of $x_{i,j}$ through an independently and identically distributed (i.i.d.) Gaussian random vector, and sends them to its next hop neighbour, which generates measurements itself and sums with all received data. Finally the weighted sum of all measurements $\mathbf{y}_j$ is transmitted to the sink. Mathematically $\mathbf{y_j}$ is represented by:

$$\mathbf{y}_j = \mathbf{\Phi}_j \mathbf{x}_j, \tag{1}$$

where $\mathbf{\Phi}_j$ is an $M_j \times N$ matrix.

The above equation is an ill-posed problem since $M_j < N$. However, according to the compressive sensing theory, an spare N signal can be recovered from less than N measurements if several certain conditions hold. Assume that all J attributes are sparse under the same basis $\mathbf{\Psi}$. Hence $\mathbf{x}_j$ can be represented as $\mathbf{x}_j = \mathbf{\Psi} \mathbf{z}_j$, where only K_j elements in $\mathbf{z}_j$ are nonzero. Further, Equation.(1) is reformulated as following:

$$\mathbf{y}_j = \mathbf{\Phi}_j \mathbf{\Psi} \mathbf{z}_j = \mathbf{A}_j \mathbf{z}_j, \tag{2}$$

where $\mathbf{A}_j = \mathbf{\Phi}_j \mathbf{\Psi}$.

According to the compressive sensing theory, if all nonzero entries of $\mathbf{z}_j$ are known, one can precisely recover $\mathbf{x}_j$ with only $M_j = K_j$ measurements by solving an l_0 minimization problem:

$$\begin{aligned} \min \quad & ||\mathbf{z}_j||_0 \\ s.t. \quad & \mathbf{y}_j = \mathbf{A}_j \mathbf{z}_j \\ & \mathbf{x}_j = \mathbf{\Psi} \mathbf{z}_j. \end{aligned} \tag{3}$$

Unfortunately, the nonzero entries are unknown, oversampling is required to guarantee the accuracy. In practise, the sink reconstruct sensor readings with $M_j = 3K_j \sim 4K_j$ measurements by solving an l_1 minimization problem:

$$\begin{aligned} \min \quad & ||\mathbf{z}_j||_1 \\ s.t. \quad & \mathbf{y}_j = \mathbf{A}_j \mathbf{z}_j \\ & \mathbf{x}_j = \mathbf{\Psi} \mathbf{z}_j. \end{aligned} \tag{4}$$

Truthfully, there are an amount of works aiming at solving minimization problems, which is beyond the scope of this paper. To ensure fairness, only the Lars algorithm [8] is used to solve all the l_1 minimization problems in this paper, which solves:

$$\min \quad \lambda ||\mathbf{z}_j||_1 + \frac{1}{2} ||\mathbf{y}_j - \mathbf{A}_j \mathbf{z}_j||_2^2 \tag{5}$$

instead of Equation.(4), where $\lambda > 0$ is a predefined threshold.

B. Capacity Analysis

C. Luo at el. [13] have proved that the ideal capacity of CDG under certain assumptions, as that 1) there is only one sink, deployed in the center of the area in the network, 2) all nodes share one radio channel and follow time-division multiple access control (TDMA), and 3) all nodes are in the same data rate. C. Luo proves that in a fixed date rate w, the capacity of the CDG, λ is inversely proportional to the number of measurement m, *i.e.*, $\lambda \propto w/m$.

Here we consider using CDG under multi-attribute scenario. Suppose that J attributes are required to send at a data rate w and each one costs m measurements. In the CDG, the total capacity remains, yet one attribute's capacity is reduced because λ need to divide by J. Consequently, the CDG performs quite well when measuring one attribute. Under multi-attribute scenario, the performance of CDG declines.

However, when keeping the data rate Jw, if measurements can be reduced, the performance becomes better. Hence, under multi-attribute scenarios, the major point is to reduce the demand on measurements without the loss of accuracy.

C. MCDG

Since all $\mathbf{z}_j$ are sparse signals, the joint vector of them is also a sparse signal. For instance, in the following equation,

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_J \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi} & & \\ & \ddots & \\ & & \mathbf{\Psi} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_J \end{bmatrix}, \quad (6)$$

since $\mathbf{\Psi}$ is an orthogonal matrix, $diag(\mathbf{\Psi}, \cdots, \mathbf{\Psi})$ is also a orthogonal matrix and is able to be treated as a basis to convert $(\mathbf{x}_1; \cdots; \mathbf{x}_J)$. Hence, it is convenient to regard all sensor readings as one sparse signal under the basis $diag(\mathbf{\Psi}, \cdots, \mathbf{\Psi})$. By adopting Equation.(2), when $J = 2$, the following representation is written:

$$\begin{bmatrix} \mathbf{y}_\triangle \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_\delta & \mathbf{\Phi}_\delta \\ \mathbf{\Phi}_1 & \\ & \mathbf{\Phi}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Psi} & \\ & \mathbf{\Psi} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}. \quad (7)$$

In Equation.(7), one part of measurements are shared by both $\mathbf{x}_1$ and $\mathbf{x}_2$, which is called joint measurements and the other part is individual measurements similarly. With joint and individual measurements, one can still formulated Equation.(7) as a l_1-norm minimization problem and obtain the sensor readings. Hereby, to joint measurements, each node is only required to sum its attributes with the same measurement coefficients and transmit $size(\mathbf{y}_\triangle)$ to the sink. In other words, the demand on measurements are saved.

In MCDG, the numbers of measurements on $\mathbf{x}_j$, are treated as:

$$M_j = m_\triangle + m_j, \quad (8)$$

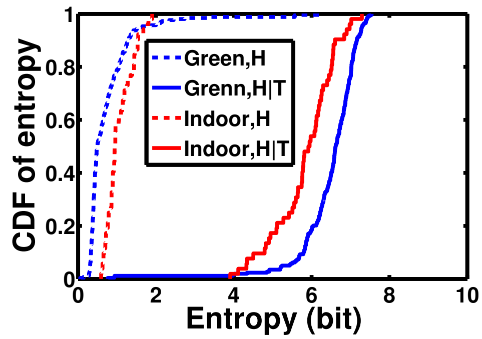
where $m_{(\cdot)} = size(\mathbf{\Phi}_{(\cdot)})$. Suppose the requirement of exact reconstruction with high possibility is $M_j > c_j K_j$, where c_j is a constant value. From the individual attribute sight, other attributes' measuring values on $\mathbf{y}_\triangle$ are regarded as noise. Hence, in other to exactly recover the signal, M_j is required to hold $M_j > (c_j + c_\delta) K_j$.

Hereby, it is required that the sparse coefficients of $\mathbf{z}_1$ and $\mathbf{z}_2$ are in the similar magnitude to avoid one attribute overshadowing another. Due to this reason, the choice on the basis is significant. In practise, wavelet basis performs better than discrete cosine transform (DCT) basis.

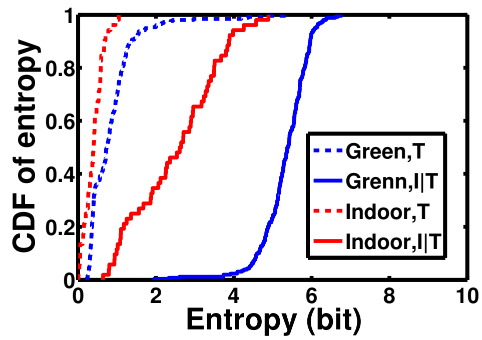
In MCDG, it is important that all sensor readings are in similar levels of magnitude. Otherwise, one attribute may overshadow another, which leads to the decline of accuracy. The solution of this problem is using weighted l_1-norm normalization, which is as following

$$\min \quad \gamma_1 ||\mathbf{z}_1||_1 + \gamma_2 ||\mathbf{z}_2||_1. \quad (9)$$

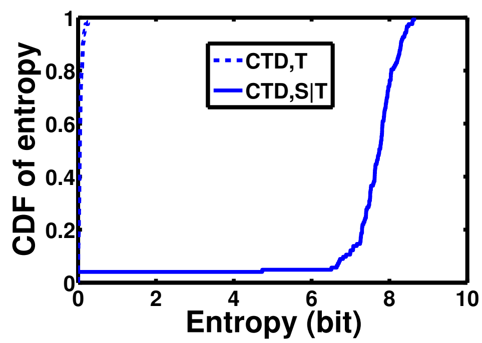
where both $\gamma_{(\cdot)}$ are trade-off coefficient.



(a) Compressibility between humidity and temperature.



(b) Compressibility between light illumination and temperature.



(c) Compressibility between salinty and temperature.

Fig. 1. The compressibility between attributes. H,T and I stand for humidity, temperature and light illumination. For example, Green,H,T represents the inter-attribute conditional entropy space between H and I (humidity and temperature).

IV. Observations on Real Data Sets

The correlations among attributes are of utmost importance in WSNs, which can be utilized to reduce the demand on measurements. In this section, by using the information theory, we illustrate that one attribute contains other one's information. Because of this hidden information, MCDG is able to achieve the required accuracy with less measurements.

A. Data Sets

We mine three real data sets, *i.e.*, CTD [2], GreenOrbs [12] and Intel Indoor [1].

1) CTD: The conductivity, temperature and depth (CTD) [2] data is collected by a ship using a Seabird 9/11 plus CTD

TABLE I
DATA SETS

Data Name	Matrix Size	Interval
Intel Indoor	49 nodes × 149 intervals	5 minutes
GreenOrbs	258 nodes × 500 intervals	30 seconds
CTD data	123 positions × 1000 intervals	1 dbar

underwater unit and water sampler. The project is carried by National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center (NDBC). The CTD data contains salinity and temperature of sea water in an area of Pacific Ocean.

2) GreenOrbs: The GreenOrbs [12] project is carried pointing at all-year round ecological surveillance in the forest, collecting sensory data such as temperature, humidity and light illumination, and content of carbon dioxide. In GreenOrbs, over 1000 nodes are deployed on an experiment region in Wuxi, China. In this paper, we choose sensor readings of relative humidity and temperature.

3) Intel Indoor: The Intel Indoor project [1] us carried by the Intel Berkeley Research Lab from Feb 28th to Apr 5th in the year 2004, where 54 nodes are used to measure humidity, temperature, light illumination and voltage. These attributes are obtained every 31 seconds.

The scale of these data sets is listed in Table.I. We compare inner-attribute entropy and inter-attribute conditional entropy to expose correlations across attributes. To GreenOrbs and Intel Indoor, every node's temporal sequence is adopted. To CTD, since sensor readings are obtained by one instrument in a specific position, its original sequence is adopt, where attributes are measured as the growth of the depth.

B. Entropy and Conditional Entropy

1) Discretization: sensor readings need to be discretized since they are real values. To an attribute sequence $\mathbf{x}_j$, the discretization carries on as following.

- Divide the range $[min(), max()]$ into Q equal sections, where Q is self-defined.
- By converting elements of $\mathbf{x}_j$ into sections, construct a state vector $\mathbf{u}_j = (u_{j,1}, \cdots, u_{j,N})^T$.

2) Inner-Attribute Entropy: The inner-attribute entropy is the entropy of state vector $\mathbf{u}_j$, which is mathematically represented as following:

$$H(\mathbf{u}_j) = -\sum_{k=0}^{Q-1} P(s_k) \cdot log_2 P(s_k) \tag{10}$$

where $P(s_k)$ is the probability of state s_k, $k = 1, \cdots, Q$. According to the information theory, the inner-attribute entropy is the amount of the information which would be need to specify $\mathbf{u}_j$.

C. Inter-Attribute Conditional Entropy

Intuitively, one can expect that correlations exist between two physically related attributes. In other words, one attribute may contain the other one's information. We use the inter-attribute conditional entropy to expose this kind of information, which is defined as the entropy of one attribute when the other attribute's reading in the same node is known, formulated as:

$$H(u_{i,t}|u_{j,t}) = H(u_{i,t}, u_{j,t}) - H(u_{j,t}), \tag{11}$$

where $H(u_{i,t}, u_{j,t})$ is the joint entropy of two attributes in one node. In this paper, we compute every node's inner-attribute and inter-attribute conditional entropy.

D. Potential Compressibility

In this section, the correlation between temperature and other two attributes, light illumination and humidity, is discussed. Temperature is chosen because it is included by all three data sets. Fig.IV shows the cumulative distribution functions (CDFs) of inner-attribute entropy and inter-attribute conditional entropy.

Here it is observed that inter-attribute conditional entropy is always far lower than inner-attribute entropy. For example, as shown in Fig.2(a), the entropy of humidity is between 4 to 8 bits, but the conditional entropy of humidity, when temperature is known, is less than 2 bits. This observation shows that the humidity and temperature can be stored together with less cost than do the same thing separately. In other words, jointing attributes is able to provide stronger compressibility, which leads to less demand on measurements according to the compressive sensing theory.

Without loss of generality, similar results can be observed from Fig.1(b) and Fig.1(c). We say that the distance of inner-attribute entropy and inter-attribute conditional entropy exposes the potential compressibility, *i.e.*, the potentiality of how many measurements can be possibly saved when using MCDG. However, sequences of sensor readings are not taken into consideration when computing the entropy. Hence, in practise, we still need more measurements in order to achieve satisfied accuracy of reconstruction.

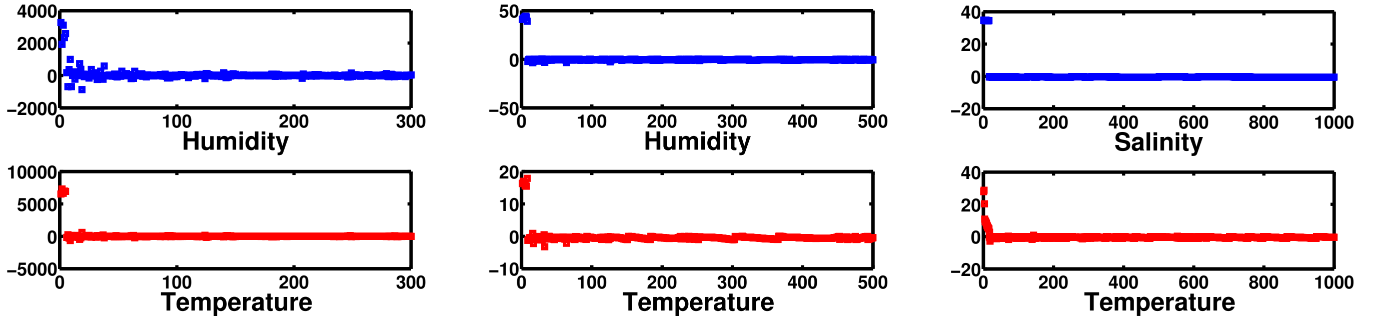
V. PERFORMANCE EVALUATION

In this section, the performance of MCDG is evaluated by comparing the accuracy and the demand of measurements with the original CDG.

The data sets of GreenOrbs [12], Intel Indoor [1] and CTD [2] are all adopted for the simulation. The detail of them has been presented in Section.IV-A.

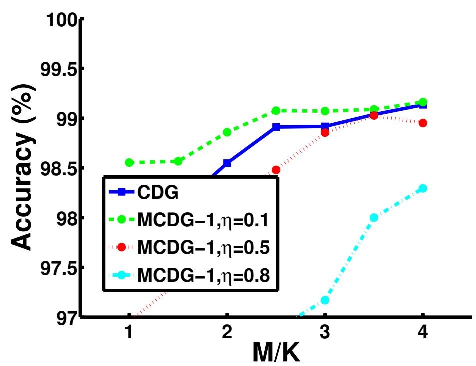
To CTD data, we regard it as if its readings are collected by multiple nodes. Specifically, sensor readings of sea water's salinity and temperature at the position (7.0N, 180W) on Mar 29th, 2008 are selected, which is also adopted by the CDG[13] in the simulation. Each attribute has 1000 readings sorted by the sea depth. Both of them are sparse in DCT and wavelet basis, as shown in Fig.2(c). Here, we set $K_j = 50$.

To GreenOrbs and Intel Indoor data, although they are obtained from multiple nodes, we still adopt temporal sequence readings of their node as if they are collected by one single routing tree. By this way, all the data can be treated as being collected from large scale WSNs. Here, to each data set, sensor

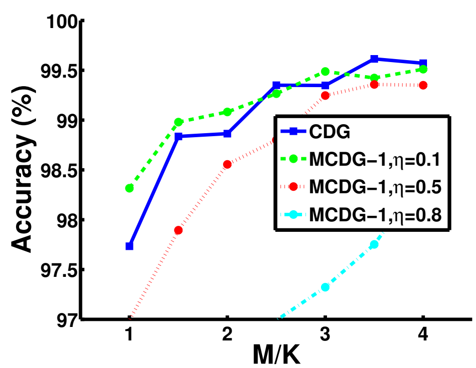


(a) GreenOrbs (b) Intel Indoor (c) CTD

Fig. 2. The sparse coefficients of attributes through Le Gall 5/3 (Spline 2.2) wavelet transform.



(a) Humidity



(b) Temperature

Fig. 3. The accuracy of MCDG on Intel Indoor.

readings of a specific node are chosen, containing relative humidity and temperature. Here K_j is set to 30 according to the observations of sparsity shown in Fig.2(a) and Fig.2(b).

A. Metric

In the simulation, the following metric is used to evaluate the accruacy of data gathering, *i.e.*

$$accuracy(\mathbf{x}, \hat{\mathbf{x}}) = 1 - \frac{||\mathbf{x}||_2}{||\hat{\mathbf{x}} - \mathbf{x}||_2}, \tag{12}$$

where $|| \cdot ||_2$ represents l_2-norm, and $\mathbf{x}, \hat{\mathbf{x}}$ are original and received readings, receptively.

The performance of MCDG and CDG is evaluated by comparing their accuracy of reconstruction through measurements $M = K \sim 4K$, where $K = \sum K_j$ and K_j is every attribute's sparsity . In the simulation, each node generates a vector of measurements through i.i.d Gaussian random process complying to $\boldsymbol{N}(1, 1/M)$, which has a good restricted isometry property (RIP) [13][14][6]. The simulation process is carried on 20 times and the average of results is adopted.

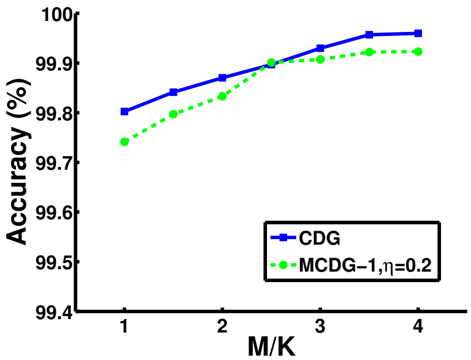
To MCDG, we use η representing the level of joint measurements, *e.g.*, when $J = 2$, suppose M measurements are used, where $\eta \times M$ are used as joint measurements and $0.5*(1-\eta)M$ are used individually by each attributes.

B. Simulation Results

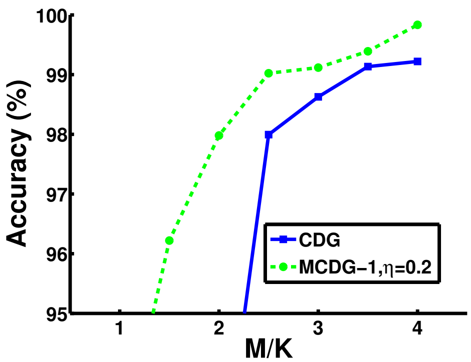
The performance of MCDG in different η is evaluated using Intel Indoor. As shown in Fig.V, MCDG outperforms CDG when $\eta = 1$, and when η is bigger, the performance of MCDG gets lower. With proper η, using MCDG can save measurements, *e.g.*, in order to achieve 99% reconstruction of humidity, MCDG with $\eta = 0.1$ takes $M = 250$ and CDG takes $M = 300$ so that 16% measurements are saved. When $M/K > 3$, both MCDG and CDG can achieve high accuracy, because the enough information is obtained by using measurements over demand. When achieving 99% accuracy, MCDG is able to save measurements with high possibility.

In the sea and forest environment, as shown in V-A, we find that MCDG can obtain the enhancement on two related attributes, as temperature and humidity, but loss its advantage without physical relationship.

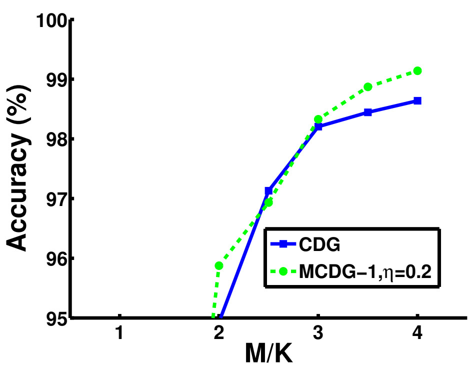
Consequently, to achieve 99% accuracy, MCDG can be used, which performs better than CDG. If higher accuracy is required, MCDG needs strong physical relationship among attributes. When the goal is to achieve over 99.5% accuracy, the number of measurements is the critical factor. All approaches performs almost on a par. Meanwhile the advantage of MCDG highly depends on the sensor readings. And η has to be calculated firstly when using MCDG. Hence the reasonable solution is firstly collecting the knowledge of readings by using CDG, and then utilizing MCDG to decline measurements.



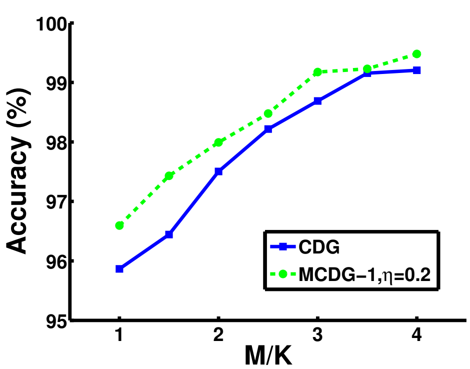
(a) Salinity of CTD



(b) Temperature of CTD



(c) Humidity of GreenOrbs



(d) Temperature of GreenOrbs

Fig. 4. The accuracy of MCDG in GreenOrbs(forest) and CTD(sea).

VI. Conclusion

In this paper, we presented an approach base on CDG, aiming at the multi-attribute data gathering problem. We theoretically analysed the disadvantage of the original CDG in the multi-attribute scenario. Then we observed sensor readings of two real data sets and exploited the correlations between attributes. Then by using distributed compressive sensing, our approach was proposed. Real data-driven simulation showed that our approach outperformed the original CDG in the multi-attribute scenario.

The future works are that 1) taking the effect of route strategy into consideration, and 2) using data prediction to enhance the performance.

Acknowledgement

This research was supported by NSF of China under grant No.61100210, 61373155, 61303202, 61311130112, STCSM Project No.12dz1507400, No. 13511507800.

References

[1] Intel Indoor Project. http://www.select.cs.cmu.edu/data/labapp3/index.html.

[2] NDBC CDT. http://tao.noaa.gov/refreshed/ctd_delivery.php.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *Communications magazine, IEEE*, 40(8):102–114, 2002.

[4] C. Caione, D. Brunelli, and L. Benini. Distributed compressive sampling for lifetime optimization in dense wireless sensor networks. *Industrial Informatics, IEEE Transactions on*, 8(1):30–40, 2012.

[5] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[6] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.

[7] G. Chen, X.-Y. Liu, L. Kong, J.-L. Lu, Y. Gu, W. Shu, and M.-Y. Wu. Multiple attributes-based data recovery in wireless sensor networks. In *Globecom 2013 - Ad Hoc and Sensor Networking Symposium (GC13 AHSN)*, Atlanta, USA, Dec. 2013.

[8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[9] L. Kong, M. Xia, X.-Y. Liu, G. Chen, Y. Gu, M.-Y. Wu, and X. Liu. Data loss and reconstruction in wireless sensor networks. *IEEE TPDS*, 2014.

[10] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu. Data loss and reconstruction in sensor networks. In *IEEE INFOCOM*, 2013.

[11] M. Li and Y. Liu. Underground structure monitoring with wireless sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*, pages 69–78. ACM, 2007.

[12] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, et al. Does wireless sensor network scale? a measurement study on greenorbs. In *INFOCOM, 2011 Proceedings IEEE*, pages 873–881. IEEE, 2011.

[13] C. Luo, F. Wu, J. Sun, and C. W. Chen. Compressive data gathering for large-scale wireless sensor networks. pages 145–156, 2009.

[14] C. Luo, F. Wu, J. Sun, and C. W. Chen. Efficient measurement generation and pervasive sparsity for compressive data gathering. *Wireless Communications, IEEE Transactions on*, 9(12):3728–3738, 2010.

[15] L. Xiang, J. Luo, and A. Vasilakos. Compressed data aggregation for energy efficient wireless sensor networks. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, pages 46–54. IEEE, 2011.