Data Gathering in Wireless Sensor Networks Through Intelligent Compressive Sensing

Jin Wang †, Shaojie Tang ‡, Baocai Yin †, Xiang-Yang Li ‡ §
† Beijing Key Laboratory of Multimedia and Intelligent Software Technology,
College of Computer Science and Technology, Beijing University of Technology
‡ Department of Computer Science, Illinois Institute of Technology
§ Tsinghua National Laboratory of Information Science and Technology, Tsinghua University

***Abstract*—The recently emerged compressive sensing (CS) theory provides a whole new avenue for data gathering in wireless sensor networks with benefits of universal sampling and decentralized encoding. However, existing compressive sensing based data gathering approaches assume the sensed data has a known constant sparsity, ignoring that the sparsity of natural signals vary in temporal and spatial domain. In this paper, we present an adaptive data gathering scheme by compressive sensing for wireless sensor networks. By introducing autoregressive (AR) model into the reconstruction of the sensed data, the local correlation in sensed data is exploited and thus local adaptive sparsity is achieved. The recovered data at the sink is evaluated by utilizing successive reconstructions, the relation between error and measurements. Then the number of measurements is adjusted according to the variation of the sensed data. Furthermore, a novel abnormal readings detection and identification mechanism based on combinational sparsity reconstruction is proposed. Internal error and external event are distinguished by their specific features. We perform extensive testing of our scheme on the real data sets and experimental results validate the efficiency and efficacy of the proposed scheme. Up to about 8dB SNR gain can be achieved over conventional CS based method with moderate increase of complexity.**

I. Introduction

One of the main application scenarios of wireless sensor networks (WSN) is continuous environmental monitoring of physical phenomenon such as temperature, humidity and light over a geographical area during certain time period [1]. Generally, the sensed physical data is transmitted from the sensor nodes to the data sink through multi-hop routing. Since the sensor nodes usually have limited computing ability and power supply, a primary goal of data gathering is to collect the sensed data at required accuracy with least energy consumption.

Conventional methods such as distributed source coding techniques [2]–[4], in-network collaborative wavelet transform [5] [6] and clustered data aggregation [7]–[10] have been proposed to reduce the global data communications. They utilize the spatial correlation in sensed data at sink or sensor nodes, but they are either not effective enough to cope with abnormal events or may bring extra computational and communication overheads. Recently the combination of the compressive sensing (CS) theory with WSNs [11]–[15] provides a new avenue for data gathering. According to the compressive sensing theory, a sparse signal can be accurately reconstructed with a relative small number of measurements, and the measurements are accomplished by random linear projections onto sensed data. Then it brings the benefits of simple compression at sensor nodes without introducing excessive computational and control overheads, which meets the limited resource constraint of sensor nodes.

Existing CS based data gathering solutions are established on an assumption that the sensed data has a constant sparsity. Therefore, a fixed transform basis according to the prior information of sensed data is adopted to achieve the sparsity in transform domain. The number of measurements is also fixed and determined by the relation between least required number of measurements and the sparsity. Unfortunately, such assumption is unlikely to hold for real sensed data. In fact, most natural signals are nonstationary and the sparsity varies in temporal or spatial domain. Note that sparsity is the guarantee of accurate reconstruction of measured data in CS theory, and it also has direct influence on the amount of required measurements. For example, if the sparsity of sensed data degrades and is not strictly K sparse in one snapshot, then more measurements are needed for accurate reconstruction, otherwise the reconstruction may fail. Thus existing CS based data gathering solutions will not perform well when the sensor readings may vary a lot.

To address the above challenges, in this paper we propose an adaptive data gathering scheme based on CS. The "adaptive" here has two fold meanings: the first one is that the CS reconstruction becomes adaptive to the sensed data, which is accomplished by the adjustment of autoregressive (AR) parameters in the objective function; the other is the number of measurement required to the sensed data is tuned adaptively according to the variation of data. Spatial correlation is common in the sensed physical data, thus we introduce the AR model to exploit such local correlation. We incorporate the AR model into the objective function of convex optimization, and the AR model fits the local structure of the sensed data adaptively by adjusting the parameters. Thus the CS reconstruction is capable of coping with the variation of sensed data adaptively. To further deal with the varied sensed data, each time when the reconstruction is accomplished at sink node, the result is approximately evaluated and forms a feedback to the sensor nodes if further measurements are needed. Specifically, the temporal correlation between historically reconstructed data

978-1-4673-0775-8/12/$31.00 ©2012 IEEE

is utilized to estimate current reconstruction result at sink node, then the error of the reconstruction is estimated. By approximating the reconstruction error as a function of the number of measurements, the number of extra measurements is determined. Thus the measurements can adjust adaptively according to the reconstruction result.

Abnormal event detection is an important task of sensor networks. The abnormal readings will degrade the sparsity of sensor readings and thus cause degradation in sensor data reconstruction. We incorporate the abnormal readings detection into the CS reconstruction novelly by utilizing combinational sparsity in different domains. Anomaly due to external error and internal error are then classified and identified by their specific features.

The proposed scheme is tested extensively on real data sets. Experimental results show that our adaptive reconstruction outperforms conventional CS based method in reconstruction quality. Up to 8dB SNR gain can be achieved with the cost of increased time complexity, which is constant times of conventional recovery. Experiments on testbed indicate that the MSE of reconstruction applying our anomaly processing scheme is less than 3.0, even with 10% abnormal readings due to both internal error and external events. The adaptive measurement is compared with the baseline scheme, less than half number of reconstructions and less than 20% re-query is required with our scheme at a cost of 20% more estimation error than the baseline.

The rest of the paper is organized as follows. Section II briefly introduces the fundamentals of CS and gives formulation of the problem. Section III elaborates our data gathering scheme. Experimental results are given in Section IV. In Section V we discuss related work. In Section VI we conclude the paper, with discussion of future directions.

II. Preliminary and Problem Formulation

A. Fundamentals of Compressive Sensing

According to CS theory [16]–[19], a sparse signal can be exactly recovered by solving a programming optimization problem from non-adaptive linear projections, which preserve the structure of sparse signals. Then sparse signal can be sampled much lower than Nyquist sampling rate.

Suppose $\mathbf{d}$ is an unknown vector in $\mathbb{R}^N$, and $\mathbf{d}$ has sparse representation in some domain(e.g. basis). If signal $\mathbf{d} \in \mathbb{R}^N$ can be decomposed as:

$$\mathbf{d} = \Psi^{-1}\mathbf{x} = \sum_{i=1}^{N} \varphi_i x_i \tag{1}$$

where $\Psi^{-1} = \{\varphi_1, \varphi_2, \cdots, \varphi_N\}$, φ_i is the i-th column vector of Ψ^{-1} and $\mathbf{x}$ is the coefficient sequence of $\mathbf{d}$ in Ψ domain. Usually Ψ is orthogonal transform basis such as DCT. We say $\mathbf{d}$ is sparse in the Ψ domain if the coefficient sequence is supported on a small set, i.e., most of the coefficients are zero. Furthermore, the signal $\mathbf{d}$ is K sparse if the number of non-zero coefficients in the coefficient sequence is K. If $K \ll N$, then $\mathbf{d}$ can be reconstructed by a small number of measurements from the acquisition system

$$\mathbf{y} = \Phi\mathbf{d} \tag{2}$$

where Φ is an $M \times N(M \ll N)$ measurement matrix and matrices Φ and Ψ are incoherent. When the number of measurements M satisfies:

$$M \geq cK \log \frac{N}{K} \tag{3}$$

where c is a positive constant, and the measurement matrix Φ satisfies the Restricted Isometry Property (RIP) [20], then it has been proved that the signal $\mathbf{d}$ can be exactly recovered by solving the following minimum l_1-norm optimization problem with very high probability [21], [22]

$$\mathbf{x} = \min ||\mathbf{x}||_1, \ s.t. \ \mathbf{y} = \Theta\mathbf{x} \tag{4}$$

where $\Theta = \Phi\Psi^{-1}$. The RIP quantifies how well the measurement matrix Φ preserves the norm of sparse vectors. In practice, a random matrix is usually chosen as the measurement matrix, e.g., the entries of Φ obey Gaussian or Bernoulli distributions.

In realistic applications the measurement $\mathbf{y}$ is usually polluted by noise, and the measurement is

$$\mathbf{y} = \Phi\mathbf{d} + \mathbf{e} \tag{5}$$

where $\mathbf{e}$ represents the additive noise during acquisition. Then $\mathbf{d}$ can be recovered by solving

$$\mathbf{d} = \min ||\Psi\mathbf{d}||_1, \ s.t. \ ||\Phi\mathbf{d} - \mathbf{y}||_2 \leq \epsilon \tag{6}$$

where ϵ is the error bound caused by $\mathbf{e}$.

Many approaches have been proposed to solve the above convex optimization problem, such as Interior Point (IP), Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP) and Projection onto Convex Sets (POCS) etc [23]–[26].

B. Problem Formulation

The CS theory provides a simple and universal data gathering solution for WSNs. The sensed data acquisition can be accomplished by making a small number of random projections of the data and transmitting the projection values. The burden of solving the optimization reconstruction has been shifted to the sink. Such asymmetric structure makes it quite suitable for resource constrained WSNs. Moreover, the property of far less measurements and democracy for equally measuring each sensed data will bring reduced communication cost and balanced network load. The two key features of compressive sensing make it a paradigm for data gathering in WSNs.

In a typical CS based data gathering scenario, the data readings for all the sensor nodes in a WSN can be represented as a vector $\mathbf{d} = [d_1, d_2, \cdots, d_N]^T$, where N is the number of nodes in the WSN. The samples to the raw data are obtained by a $M \times N(M \ll N)$ random matrix $\mathbf{b} = \Phi\mathbf{d} = d_1\phi_1 + d_2\phi_2 + \cdots + d_N\phi_N$, where M represents the number of measurement to $\mathbf{d}$ and ϕ_i is the i-th column vector of Φ. The sampling process is also the compression process in CS, which is performed individually by simple multiplications and addition at each sensor node. Then the M-dimensional vector rather than the N-dimensional raw data is transmitted to the sink. Thus the transmission is greatly reduced. The raw data is recovered from M samples by optimization reconstruction at sink. And the performance depends on the sparsity of the sensed data and reconstruction algorithm.

III. Data Gathering through Intelligent Compressive Sensing

In this section, we propose our adaptive data gathering scheme based on CS. The sensed data is adaptively reconstructed by incorporating AR model into the existing CS reconstruction. After that, the recovered data is evaluated to determine whether the recovery satisfies the accuracy requirement and whether further measurements are needed to improve the reconstruction quality. If so, a feedback is formed to the sensor nodes to take supplementary measurements. Since abnormal readings are common in sensor networks, we provide an abnormal readings detection and identification method to handle the anomaly in sensed data.

A. Adaptive Reconstruction

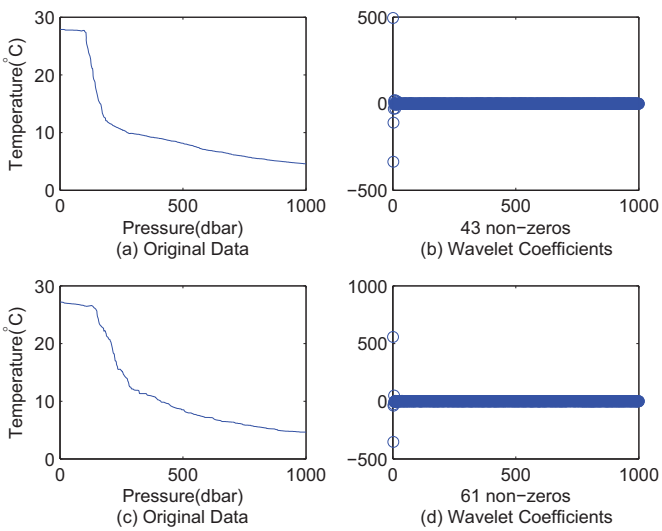


Fig. 1. The sparsity varies in most real natural signals

Prior effort has been made to introduce CS into data gathering, and existing CS based data gathering schemes assume that the sparsity of the sensed data is a constant and keep unchanged. Also a fixed set of bases such as DCT and wavelets is adopted to represent the sparsity of the entirety of sensed data. However, this does not hold true for natural physical data. Natural signals are usually nonstationary and their sparsity varies in spatial and temporal domains. As Fig. 1 shows, Fig. 1(a) plots the snapshot of 1000 temperature data readings from CTD data [27], Fig. 1(b) plots the corresponding coefficients after 6-level 5/3 wavelet transform. Most coefficients are near zero and about 43 coefficients are relatively large, thus the signal is approximately 43-sparse. Fig. 1(c) plots another snapshot of the temperature data and Fig. 1(d) shows its sparsity is about 61. Thus for existing CS based data gathering schemes, the reconstruction performance will degrade for varied data. Moreover, CS based compression methods need to find a space in which the signal is sufficiently sparse for optimal recovery, and it poses a challenge to the CS theory and its applications.

The sensor nodes are usually densely deployed to monitor the same physical phenomenon, thus high degree of spatial correlation is expected to exist in the sensed data. Various algorithms have exploited the spatial correlation to solve different problems such as data aggregation. Here we propose an adaptive CS reconstruction scheme that adopt the AR model to exploit the varying local spatial correlation in sensed data.

Suppose the readings from each sensor node is N dimensional vector $\mathbf{d}$, and we have measurements $\mathbf{b} = \Phi\mathbf{d} = \Theta\mathbf{x}$. Then the reconstruction can be formulated as the following minimization problem:

$$\mathbf{x} = \min \alpha \sum_{i}^{N} (d_i - \sum_{j \in S_i} a_{ij} d_{ij})^2 + \beta ||\mathbf{x}||_1 + \frac{1}{2} ||\Theta\mathbf{x} - \mathbf{b}||_2^2 \quad (7)$$

where S_i is the neighbor index set of the i-th node, which is also the support set of the AR model, a_{ij} is the j-th AR parameter for the i-th node, and d_{ij} is the data reading of the j-th neighbor for the i-th node. And (7) can be reexpressed as:

$$\min F(\mathbf{x}) = \min \alpha ||A\Psi^{-1}\mathbf{x}||_2^2 + \beta ||\mathbf{x}||_1 + \frac{1}{2} ||\Theta\mathbf{x} - \mathbf{b}||_2^2 \quad (8)$$

where A is a $N \times N$ matrix and each row of A contains the AR parameter for each node. The matrix A can be written as:

$$A = \begin{pmatrix} -1 & a_{11} & \cdots & a_{1j_1} & 0 & 0 \\ 0 & -1 & a_{21} & \cdots & a_{2j_2} & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & a_{N1} & \cdots & a_{Nj_N} & -1 \end{pmatrix} \quad (9)$$

where $j_i = |S_i|$, which denotes the cardinality of neighbor set of the i-th node.

We will solve the minimization in (8) in an alternative way. Specifically, we'll decompose the original problem (8) into two sub-problems of solving A and $\mathbf{x}$ separately. We'll initialize the AR parameter A_0 by smooth values at first, then $\mathbf{x}_0$ can be solved given A_0 in (8). And an improved estimation of AR parameter A_1 can be obtained given $\mathbf{x}_0$. The process iterates between revising $\mathbf{x}$ given A and revising A given $\mathbf{x}$.

First let's consider the sub-problem of estimating A given $\mathbf{x}$. Suppose the AR parameter for the i-th node is denoted as $\mathbf{a}_i$. Since many natural signals have slowly changing second order statistics, which means the AR parameter $\mathbf{a}_i$ is nearly constant or varies little piecewise, we can estimate $\mathbf{a}_i$ in a local area W_i. Then the sub-problem can be formulated as:

$$\mathbf{a}_i = \min G(\mathbf{x}) = \min \sum_{i \in W_i} (d_i - \sum_{j \in S_i} a_{ij} d_{ij})^2 \quad (10)$$

then the closed form solution of (10) is:

$$\mathbf{a}_i = (B^T B)^{-1} B^T \mathbf{n} \quad (11)$$

where the i-th row of matrix B consists of the neighbor readings of d_i, and column vector $\mathbf{n}$ is composed of all the d_i in W_i.

Then consider the sub-problem of estimating $\mathbf{x}$ given A, which is solving (8) given A. The optimality condition for (8) is:

$$\partial F(\mathbf{x}) = \mathbf{0} \quad (12)$$

Algorithm 1 Solution to the minimization of (8)

Input: Measurement matrix Φ, orthogonal transform matrix Ψ, measurements $\mathbf{b}$ and initial A_0
Output: $\mathbf{d}_k, A_k$.

$\mathbf{x}_0 = \mathbf{0}$; $\mathbf{y}_0 = \mathbf{0}$; $\tau_1 > 0$; $\tau_2 > 0$; $\epsilon > 0$; $\varepsilon > 0$; $k = 0$;
while $\frac{||\mathbf{d}_{k+1} - \mathbf{d}_k||_2}{||\mathbf{d}_k||_2} < \epsilon$ **do**
  while $\frac{||(\mathbf{x},\mathbf{y})_{k+1} - (\mathbf{x},\mathbf{y})_k||_2}{||(\mathbf{x},\mathbf{y})_k||_2} < \varepsilon$ **do**
    $\mathbf{s}_{k+1} = \mathbf{x}_k - \tau_1(\alpha\Psi A^T\mathbf{y}_k + \partial(\frac{1}{2}||\Theta\mathbf{x}_k - \mathbf{b}||_2^2))$;
    $\mathbf{t}_{k+1} = \mathbf{y}_k + \tau_2 A\Psi^{-1}\mathbf{x}_k$;
    $\mathbf{x}_{k+1} = \mathrm{sgn}(\mathbf{s}_k)\max\{0, |\mathbf{s}_k| - \tau_1\beta\}$;
    $\mathbf{y}_{k+1} = \min\{\frac{1}{\tau_2}, ||\mathbf{t}_k||_2\}\frac{\mathbf{t}_k}{||\mathbf{t}_k||_2}$;
  end while
  $(\mathbf{a}_i)_{k+1} = (B_k^T B_k)^{-1} B_k^T \mathbf{n}_k$;
  $k = k + 1$;
end while

We denote $f(\mathbf{x}) = ||\mathbf{x}||_2^2$, then for convex function f and its convex conjugate f^*, we have

$$y = \partial f(x) \Leftrightarrow x = \partial f^*(y) \tag{13}$$

Then by introducing a variable y, the optimality condition is

$$\alpha\Psi A^T\mathbf{y} + \beta\partial||\mathbf{x}||_1 + \partial(\frac{1}{2}||\Theta\mathbf{x} - \mathbf{b}||_2^2) = \mathbf{0} \tag{14}$$

$$\partial f^*(\mathbf{y}) = A\Psi^{-1}\mathbf{x} \tag{15}$$

We can apply operator splitting method to (14) and (15) with scalars $\tau_1, \tau_2 > 0$, then

$$\tau_1\beta\partial||\mathbf{x}||_1 + \mathbf{x} - \mathbf{s} = \mathbf{0} \tag{16}$$

$$\mathbf{s} = \mathbf{x} - \tau_1(\alpha\Psi A^T\mathbf{y} + \partial(\frac{1}{2}||\Theta\mathbf{x} - \mathbf{b}||_2^2)) \tag{17}$$

$$\tau_2\partial f^*(\mathbf{y}) + \mathbf{y} - \mathbf{t} = \mathbf{0} \tag{18}$$

$$\mathbf{t} = \mathbf{y} + \tau_2 A\Psi^{-1}\mathbf{x} \tag{19}$$

For (16) and (18), given $\mathbf{s}$ and $\mathbf{t}$, $\mathbf{x}$ and $\mathbf{y}$ can be obtained by

$$\mathbf{x} = \mathrm{sgn}(\mathbf{s})\max\{0, |\mathbf{s}| - \tau_1\beta\} \tag{20}$$

$$\mathbf{y} = \min\{\frac{1}{\tau_2}, ||\mathbf{t}||_2\}\frac{\mathbf{t}}{||\mathbf{t}||_2} \tag{21}$$

Finally the second sub-problem can be solved by the iteration of $\mathbf{s}$, $\mathbf{t}$, $\mathbf{x}$ and $\mathbf{y}$ from initial points $\mathbf{x}_0$ and $\mathbf{y}_0$. And the stopping criteria of iteration can be defined as the successive error of $\mathbf{x}$ and $\mathbf{y}$ is lower than a predefined threshold ε, that is:

$$\frac{||\mathbf{x}_{k+1} - \mathbf{x}_k||_2}{||\mathbf{x}_k||_2} < \varepsilon_x \quad and \quad \frac{||\mathbf{y}_{k+1} - \mathbf{y}_k||_2}{||\mathbf{y}_k||_2} < \varepsilon_y \tag{22}$$

The whole solution is depicted in Algorithm 1.

The initial choice of AR parameter $\mathbf{a}_i$ for a sensor node can be smooth values such as reciprocal of the number of neighbors of the node. The time complexity of proposed adaptive reconstruction is constant times of conventional recovery for the iterative process. According to our experimental results, the outer iteration of revising the AR parameters in A will converge fast, usually after about 5 outer iterations the AR parameters in A will be exact to make the algorithm converge to an accurate solution. Utilizing more neighbors of current node can provide more samples to estimate the AR parameters, however, the CS reconstruction performance may suffer from data overfitting if the order of the AR model is too high. Moreover, the spatial correlation decreases as the distance between two nodes increases. Considering that the sensed data readings from one-hop neighbors demonstrate direct correlation, we only use the one-hop neighbors of the sensor nodes in the AR model. The neighbors are determined by the physical location and topology of the sensor network, thus are irrelevant with the routings.

B. Abnormal Readings Processing

Anomalous data readings are prevalent in WSNs. Since the random measured data is information-dense, the perturbations of the measurements severely influence the successful reconstruction. We propose to address the problem by utilizing the combinational sparsity in different domains of sensed data. First we detect and recover the abnormal readings by combinational sparse of sensed data with anomalous readings. Then the recovered abnormal readings are identified and classified into two categories of internal error and external event by their specific patterns. Abnormal readings due to internal errors fail to represent the sensed physical data, thus they should be removed and replaced by its underlying normal readings, while anomaly due to external errors reflects the true scenarios in the environment and it can be preserved.

Inspired by the overcomplete representation basis in [13], we use the combinational sparsity to handle the anomaly. Sensed data with abnormal readings $\mathbf{d}$ can be decomposed into two parts:

$$\mathbf{d} = \mathbf{d}_n + \mathbf{d}_a \tag{23}$$

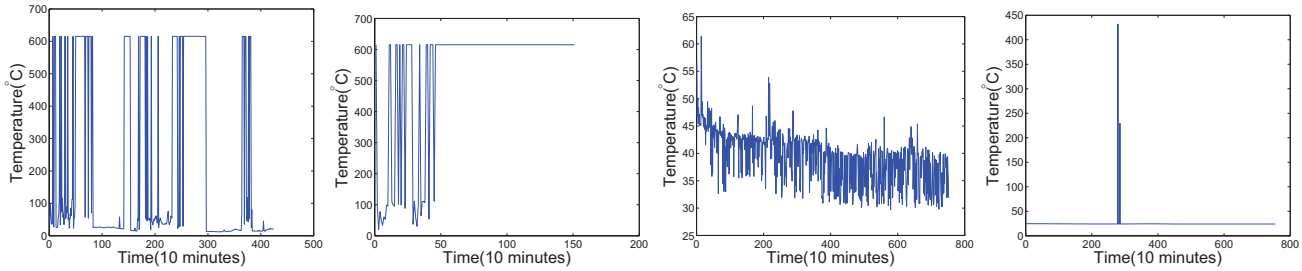
where $\mathbf{d}_n$ is the normal readings while $\mathbf{d}_a$ is the difference between abnormal readings and its underlying normal readings. $\mathbf{d}_n$ is sparse in some transform domain and as abnormal readings are usually sporadic in real sensed data, $\mathbf{d}_a$ can be viewed as being sparse in spatial domain. Thus the combinational sparsity in different domains is achieved in the following way:

$$\mathbf{d} = \Psi^{-1}\mathbf{x}_n + I\mathbf{d}_a = [\Psi^{-1} \quad I][\mathbf{x}_n^T \quad \mathbf{d}_a^T]^T \tag{24}$$

Here we introduce the auxiliary variable $\mathbf{d}'$ to incorporate the AR model into the abnormal detection:

$$\begin{aligned} \mathbf{d}' &= A\mathbf{d} + \mathbf{d} \\ &= [I \ I \ \Psi^{-1} \ I][(A\Psi^{-1}\mathbf{x}_n)^T \ (A\mathbf{d}_a)^T \ \mathbf{x}_n^T \ \mathbf{d}_a^T]^T \\ &= (\Psi')^{-1}\mathbf{x}' \end{aligned} \tag{25}$$

where $(\Psi')^{-1} = [I \ I \ \Psi^{-1} \ I]$. For both normal readings transform coefficients $\mathbf{x}_n$ and anomalous readings $\mathbf{d}_a$, it's obvious that $A\Psi^{-1}\mathbf{x}_n$ and $A\mathbf{d}_a$ are also sparse even with an initial A. Take the recovered CTD data in Fig. 1(a) with 160 measurements for example, for $A\Psi^{-1}\mathbf{x}_n$, only 4 out of 1000 coefficients have absolute value greater than 0.1. Thus $\mathbf{x}'$ is sparse in the combinational domain Ψ'.



(a) (b) (c) (d)

Fig. 2. The typical anomalous sensed temperature data from trace of GreenOrbs

Since $\mathbf{d}'$ doesn't really exist at sensor nodes as $\mathbf{d}$, we need to deduce the measurement of $\mathbf{d}'$ from $\mathbf{d}$. For the original anomaly contaminated sensed data $\mathbf{d}$, the random measurement is accomplished by $\Phi\mathbf{d}$, thus the measurement for $\mathbf{d}'$ is:

$$\Phi\mathbf{d}' = \Phi(A\mathbf{d} + \mathbf{d}) = \Phi(A + I)\mathbf{d} = \Phi'\mathbf{d} \tag{26}$$

where the new measure matrix $\Phi' = \Phi(A+I)$. It indicates that using a measure matrix $\Phi(A + I)$ to $\mathbf{d}$ is equivalent to using measure matrix Φ to $\mathbf{d}'$. An intuitionistic idea is to change the measure matrix into $\Phi(A + I)$ at sensor nodes. We shift such transform to the sink to keep sensor nodes unchanged, which means still a random measure matrix Φ is deployed at sensor nodes. Then when the original measurement $\Phi\mathbf{d}$ is obtained, the measure matrix Φ' for $\mathbf{d}'$ becomes

$$\Phi' = \Phi(A + I)^{-1} \tag{27}$$

and $\Phi(A + I)^{-1}$ is a matrix with random property, thus it can be used as measure matrix. Then (25) and (27) can be incorporated into CS reconstruction, and a large non-zero value in $\mathbf{d}_a$ indicates the location of the abnormal readings. Thus both the location and values of the abnormal readings can be solved.

We classify the abnormal readings into two categories to process them in different ways. Internal and external errors of anomalous readings demonstrate different features and we can use the recovered abnormal data to distinguish them from each other. Fig. 2 plots the prevalent anomalous sensed temperature readings from trace data of GreenOrbs [28] in several days with time interval of 10 minutes. Such anomaly is observed when the sensor node has low battery voltage, which is the main cause of abnormal readings in the system. Frequent dramatic fluctuation and exceeding of normal temperature range can be observed and thus distinguish themselves from normal readings. Usually abnormal readings due to external events have limited peak values and gradual change compared with that of internal errors. We can leverage such domain knowledge and prior information to develop heuristic rules to identify anomaly in sensed data. Specifically we utilize thresholding and variance analysis to distinguish the abnormal readings. A relative larger abnormal reading and its variance is more likely to indicate an internal error, while anomaly with relative smaller reading and variance is an external event with higher probability. The internal errors can be further classified by their unique features.

Take the faulty data in Fig. 2 as an example, which is the predominant anomaly we observed in the GreenOrbs trace. The detected abnormal readings are first compared with threshold h_1, then the variance of the readings within a time window containing n samples is computed. We set two thresholds h_2 and larger h_3 for the variation to indicate different levels of fluctuation. For abnormal data in right half of Fig. 2(b), the anomalous data will be identified as it's above h_1 and the variance of readings within a window will be far below h_2. For the abnormal data in Fig. 2(d), it exceeds both h_1 and h_3. The readings in Fig. 2(a) actually combines two cases in Fig. 2(b) and Fig. 2(d). The data in Fig. 2(c) will be below h_1 while its variation is between h_2 and h_3. By exploiting such specific features we can identify the typical abnormal readings caused by internal errors, and consider the other anomaly as external event occurs in sensor networks.

C. Adaptive Measurements

In most previous effort of CS based data gathering, the measurements and the reconstruction are separate. The sensed data is measured by random projection followed by optimization reconstruction. The data is assumed to be time-invariant and strictly K-sparse, thus the measurements are sufficient for accurate recovery. Nevertheless, due to the fluctuation in the time-varying sensed data, the sparsity of sensor readings may degrade as time varies. Thus the measurements may be insufficient for accuracy recovery. Although an adaptive measurement algorithm is proposed in [29] based on the adaptive CS theory, it jointly optimize the routing and compression to obtain optimal measurement, which greatly increases the complexity.

We propose to evaluate the reconstructed sensor readings and adjust measurements according to the evaluation. Each reconstruction is first checked whether the measurement is sufficient at sink by utilizing the temporal correlation between sensor readings. If not, a feedback is formed to inform the sensor nodes for small number of probing measurements to determine current reconstruction error. The reconstruction error can be formulated as a function of number of measurement, and it can be approximated by curve fitting with current reconstruction and probing measurement. The approximation is then utilized to specify the trend of decreasing error with increase of measurements. Thus the rough number of further

measurements required to satisfy the accuracy requirement is obtained.

According to [30], for a sparse signal $\mathbf{x} \in \mathbb{R}^N$ which has exactly k non-zero components, and the measurements are random Gaussian samples, the correct recovery can be declared if one of the following conditions is satisfied:

1) $\hat{\mathbf{x}}_{M+1} = \hat{\mathbf{x}}_M$, the reconstruction with $M+1$ measurements is the same with that of M measurements;
2) $||\hat{\mathbf{x}}_M||_0 < M$, $\hat{\mathbf{x}}_M$ has fewer than M non-zero entries.

However, the signal is approximately sparse in practice: most of the coefficients after orthogonal transformations are relatively small rather than exactly zero. For approximately sparse signal $\mathbf{x}$ under random Gaussian measurements, the distortion between original signal and recovery with M measurements can be bounded as:

$$d(\mathbf{x}, \hat{\mathbf{x}}_M) \leq \frac{d(\hat{\mathbf{x}}_{M+T}, \hat{\mathbf{x}}_M)}{\sin\theta} \tag{28}$$

where θ is the angle between the vector connecting $\mathbf{x}$ with $\hat{\mathbf{x}}_M$ and $N-(M+T)$ dimensional hyperplane [30]. And $E[\frac{1}{\sin\theta}] \geq \sqrt{\frac{N-M}{T}}$, from which the recovery error can be further approximated by

$$d(\mathbf{x}, \hat{\mathbf{x}}_M) \approx d(\hat{\mathbf{x}}_{M+T}, \hat{\mathbf{x}}_M)\sqrt{\frac{N-M}{T}} \tag{29}$$

(28) and (29) indicate that extra measurements are required to approximate the recovery error of current reconstruction.

Once the sink receives M measurements and recovers $\hat{\mathbf{x}}_M$, it is first compared with its last recovered normal reading and check if the difference D is within a threshold h. Such process is reasonable because many natural signals such as temperature will not change dramatically in a short time period. If the absolute difference is below h, then we can consider $\hat{\mathbf{x}}_M$ as a successful recovery. Otherwise the reconstruction is viewed as faulty one and our problem turns into how to determine the number of supplemental measurements efficiently. A baseline approach is to successively add measurements at a rate till the required accuracy is achieved. However, it brings much extra computational and communication burden. To reduce such complexity, we propose to fit the trend of decreasing error as a function of measurements and approximately gives the number of extra measurements T required. First extra $m(m \ll M)$ probing measurements are taken to determine the error of current recovery and approximate number of supplementary measurements. m can be defined as $m = \alpha M$, where α is a scale coefficient related to difference D. The larger D is, the higher probability the sparsity has increased dramatically, thus a larger α will take more measurement to capture the variation accurately. The m supplemental measurements is divided into n parts so that $m = nt$, and n is the number of probing reconstructions for error curve fitting. Then the sensed data with increasing number of measurements are recovered: $\hat{\mathbf{x}}_M, \hat{\mathbf{x}}_{M+t}, \cdots, \hat{\mathbf{x}}_{M+nt}$. Thus by (29) the recovery error $\hat{\mathbf{e}}_M, \hat{\mathbf{e}}_{M+t}, \cdots, \hat{\mathbf{e}}_{M+(n-1)t}$ can be estimated. The error function of number of measurement is then approximated by curve fitting to get the rough number of supplementary measurement required by the accuracy. We represent the error function and error bound as $\tilde{f}$ and ε respectively. Then the minimal T that satisfies $\tilde{f}(M+m+T) \leq \varepsilon$ approximates the extra number of measurements to the sensed data required by the error bound.

Note that when the sparsity of the sensed data decreases, the number of measurements can also be reduced accordingly to reduce communication cost. If the current reconstruction with M measurements already satisfies accuracy requirement, then $M - \alpha M$ out of M measurements at sink are used for a probing reconstruction and recovery error $\hat{\mathbf{e}}_{M-\alpha M}$ is estimated. If $\hat{\mathbf{e}}_{M-\alpha M}$ is below the error bound ε, then $M-\alpha M$ measurements are sufficient for an accurate recovery and sensor nodes are notified to send only $M-\alpha M$ measurement next time. Otherwise the number of measurement M will keep unchanged. Here we suggest not to reduce the number of measurement immediately after the sparsity of sensed data decreases to avoid fluctuations and frequent adjustments.

We use the real CTD trace data to illustrate our adaptive measurement scheme. Fig. 3 depicts the real recovery error

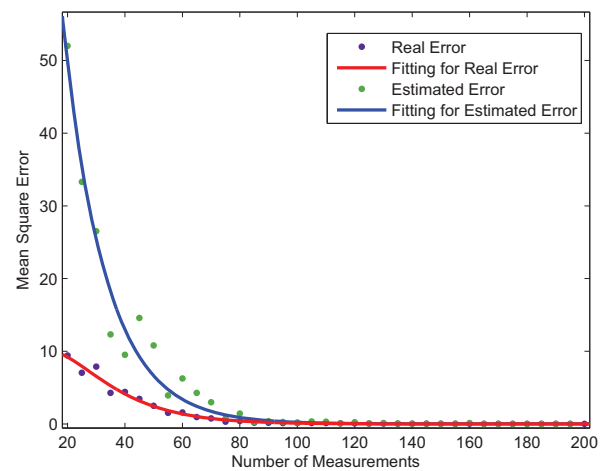
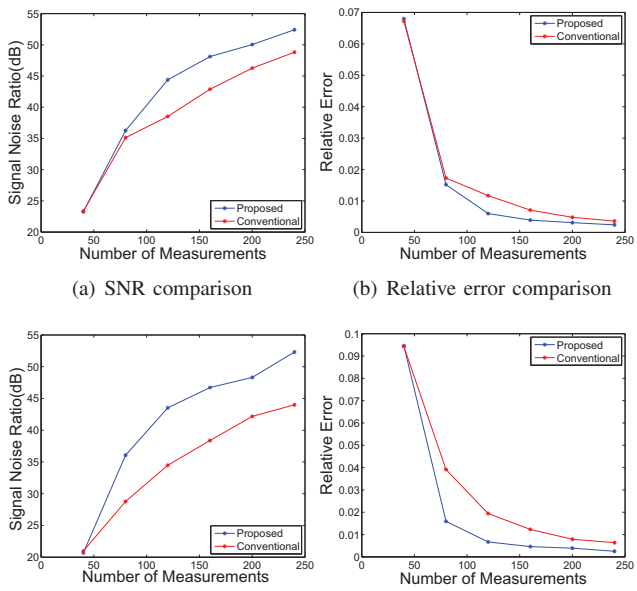


Fig. 3. Estimated recovery error and real error

and corresponding estimated error of 1000 reconstructed temperature data with different number of measurements. The original temperature data is approximately 43 sparse after 6-level 5/3 wavelet transform. The fitted curves are also plotted, from which the relation between measurements and recovery error can be formulated as an exponential function. Fig. 3 also clearly indicates that the difference between estimation and real error decreases as measurements increase, but non-monotonically. When there are enough measurements, around 80 in Fig. 3 for example, the estimation and real error begin to be coherent with each other very well. Thus we can use the error estimation to evaluate the recovery error when the real error is unavailable at sink. If the recovery is below accuracy requirement, extra supplemental measurements of sensed data is required, which can be approximately solved by the fitted error function of number of measurement.

Fig. 3 also indicates that when the measurement is severely insufficient, which corresponds to the scenario when the



(a) SNR comparison (b) Relative error comparison

(c) SNR comparison (d) Relative error comparison

Fig. 4. Reconstruction comparison of our scheme and conventional CS

sparsity of sensed data increases dramatically, the real reconstruction error and the estimated error is quite different and the reconstruction error can not be evaluated correctly. For such situation we suggest to increase the measurements in the baseline way, that is by successively increasing number of measurement by αM till the difference of reconstruction $\hat{\mathbf{x}}_M$ and $\hat{\mathbf{x}}_{M+\alpha M}$ is quite small.

IV. Experimental Results

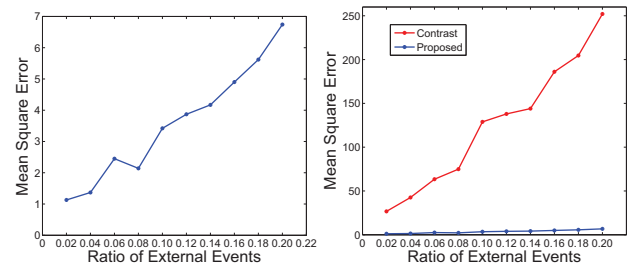
In this section, we will give the experimental results of our scheme on real sensor data sets.

Fig. 4 depicts the reconstruction quality comparison of recovered 1000 CTD temperature data between our scheme and conventional CS reconstruction. The data is from the CTD dataset on March 29, 2008 and March 26, 2008 respectively. The SNR and relative error E are reconstruction quality criteria as defined below:

$$SNR = 10\log_{10}\left(\sum_{i=1}^{N} d_i^2 / \sum_{i=1}^{N}(d_i - \hat{d}_i)^2\right) \tag{30}$$

$$E = ||\mathbf{d} - \hat{\mathbf{d}}||_2 / ||\mathbf{d}||_2 \tag{31}$$

where N is length of the signal and $\hat{d}_i$ is the reconstructed value of the i-th node. Gaussian random matrix is used as measurement matrix in both methods. The reconstructions are repeated 20 times to avoid fluctuation. We use the Fixed-Point Continuation (FPC) method [31] to solve the conventional CS reconstruction in (6). The figures indicate our scheme outperforms traditional reconstruction by up to about 8dB in SNR, and the relative error is about 0.30% with only 160 random measurements of the 1000 temperature data. The cost is that the time complexity of our adaptive scheme is constant times of that of FPC and the constant is about 5 according to our experiments. Also the topology of the sensor networks



(a) Reconstruction quality (b) Comparison with a contrast scheme

Fig. 5. Reconstruction quality of our scheme with different ratio of anomaly

is needed to specify the neighbors for each sensor node. Note that when the measurement is extremely insufficient for accurate reconstruction, which means the recovered data degrades severely and thus is useless, our method has a little lower SNR value. The reason is that when the recovered data greatly deviate from the original sensed data, the solution of AR model parameter will be inaccurate and thus affects the accurate reconstruction in turn.

Our proposed abnormal readings processing scheme is validated on a testbed with 5×10 sensors deployed in a $2m \times 1m$ rectangular area. The abnormal readings due to internal error is realized by supplying the sensor nodes with low power battery, and the patterns of anomaly is as depicted in Fig. 2. Anomaly due to external event is realized by changing the environmental temperature for some sensor nodes. The total anomaly ratio is set 10%, that is 10 out of 50 sensor nodes will produce anomaly of both internal and external error. And among these sensors the ratio of event caused anomaly is set from 2% to 20%. The ground truth of the abnormal readings are preserved by the duplicate sensors. The abnormal readings of sensed temperature data are first detected by the combinational sparse reconstruction. Then anomaly due to external event and internal error is identified by the domain knowledge based rules as defined in Section III-B. The length of time window to calculate the variation is set 10 successive readings. Anomaly due to external event will be kept while internal error caused abnormal readings will be replaced by its recovered underlying value. According to the domain knowledge and prior information, the threshold h_1,h_2 and h_3 are 10^2,10^2 and 10^4 respectively in our experiments. The reconstruction quality is as shown in Fig. 5. Fig. 5(a) indicates that recovery error increases as ratio of external events increases. The MSE of recovery error is less than 3.0 when the ratio is below 10%. In Fig. 5(b) we compare the result with a contrast scheme, in which internal and external error are not distinguished. Thus all the abnormal readings are considered as error and replaced by the estimated value.

We test our adaptive measurement scheme on the CTD trace data as shown in Fig. 1. We start from the current reconstruction with 90 measurement of the 1000 temperature data, and the accuracy requirement ε of the reconstruction MSE is 0.03. The scale coefficient α is set about 22% and thus there are 20 probing measurements at first. We set n as 4 in our experiment to reduce the complexity. Together

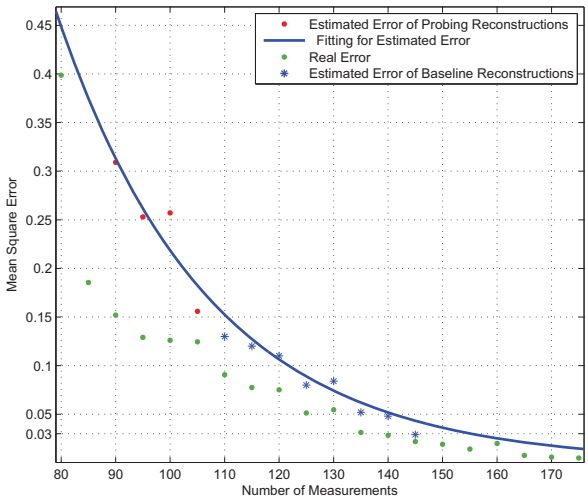


Fig. 6. Fitting result of estimated errors with 4 probing reconstructions

with initial construction $\hat{\mathbf{x}}_{90}$ we have 5 reconstructions with different number of measurements: $\hat{\mathbf{x}}_{90}$, $\hat{\mathbf{x}}_{95}$, $\hat{\mathbf{x}}_{100}$, $\hat{\mathbf{x}}_{105}$ and $\hat{\mathbf{x}}_{110}$. Then recovery error function $\tilde{f}$ can be approximately formulated by curve fitting of $\hat{\mathbf{e}}_{90}$, $\hat{\mathbf{e}}_{95}$, $\hat{\mathbf{e}}_{100}$ and $\hat{\mathbf{e}}_{105}$.

Fig. 6 shows the result of fitting result with 4 probing measurements. We compare it with the baseline scheme, in which the number of measurement is successively increased. The baseline scheme will take 11 more reconstructions and broadcasts of requiring measurements from sensor nodes, the error requirement can be satisfied with about 145 measurements. While the curve fitting formulates the trend of recovery error decrease and the fitted error function $\tilde{f}$ indicates about 155 measurements will be enough for the required accuracy. Then on the base of 20 probing measurements, 45 more measurements are required from sensor nodes. And the extra required number is 55 in baseline. Thus the proposed curve fitting scheme needs only about 5 extra reconstructions and 2 broadcast in our experiment, which is far less than that of baseline. We adopt curve fitting to estimate the extra number of measurements rather than find an accurate solution, thus error is inevitable. For example actually only about 140 measurement is enough as indicated in Fig. 6. And the estimated error is 20% more than the baseline scheme. However, considering the number of sensed temperature data is 1000, the measurements is still far less than that even with such error.

V. Related Work

In this section, we'll classify the data gathering techniques into two categories of conventional approaches and CS based. And we'll briefly introduce the prior work as follows.

Conventional non-CS approaches. Much effort has been made on non-CS data gathering in WSNs and prior effort mainly focuses on three directions: (1) Data aggregation techniques. Similar with the theory of vector quantization, clusters of sensed data are formed and a small subset of sensor nodes are selected, which are sufficient to reconstruct data of the whole sensor network. So only a subset of nodes are involved in the communication and the global communication cost incurred during data gathering is thus reduced. [7]–[10] share such similar idea. (2) Distributed source coding. These schemes are based on the Slepian-Wolf coding theory and spatial correlation is utilized at sink node [2]–[4]. However, they typically require exact knowledge of the correlation between attributes, otherwise they will fail. (3) Collaborative in-network compression. Collaborative in-network compression can be viewed as an extension of traditional transform in signal processing to sensor networks with irregular topology. Nodes can communicate with their neighbors and spatial correlation is exploited by collaborative transform such as distributed wavelet transform [5] and graph wavelet transform [6].

CS based approaches. Compressive sensing provides two features of universal sampling and decentralized simple encoding, which makes it a new paradigm for data gathering in sensor networks. [11] propose a universal compressive wireless sensing (CWS) scheme, in which sensed data is measured by synchronized amplitude-modulated analog transmissions to the fusion center in a single-hop network. [12] discusses methods to sparsify networked data and decoding algorithms when apply CS for data collection in WSN. [13] presents the first complete scheme to apply CS to data gathering in large scale sensor networks. The capacity gain of the scheme is validated by both analysis and simulation result. And different practical ways to make the sensed data sparse are discussed. [29] proposes an adaptive algorithm based on adaptive CS theory to collect data in sensor networks. By maximizing information gain per energy cost during each measurement, the scheme adaptively collects data in a energy efficient way. However, during each measurements the projection vector is obtained by solving an NP-hard optimization problem, which brings relative high computational and communication overhead. [14] compares the throughput of non-CS and plain-CS schemes and proposes a hybrid-CS scheme to improve performance, in which non-CS scheme is applied in the earlier stages of data collection and CS based collection is only applied at nodes whose traffic is larger than some extent. [15] discusses the problem from the aspect of energy efficiency. It investigates the energy efficiency of applying CS to data gathering in WSNs, aiming at minimizing energy consumption through joint routing and compressed aggregation. The optimization problem is proved NP-completeness and both optimal and near-optimal solutions are given.

VI. Conclusion

In this paper, we have illustrated a data gathering scheme for WSNs by intelligent compressive sensing. We suggest that the assumption of constant sparse of sensed data in conventional CS based data gathering is unrealistic for most natural signals. To address the problem of reconstruction quality degradation due to sensed data variation, we propose our adaptive data gathering scheme based on CS. By incorporating the AR model to exploit the local spatial correlation between sensed data of neighboring sensor nodes, the reconstruction is thus

adaptive to the variation of sensed data by adjusting the AR parameters. The measurement also adjusts with the sensed data adaptively by evaluating the recovery result and approximating the number of measurement required to satisfy the demand of accuracy. Since anomaly is prevalent in sensor networks, we provide an abnormal readings detection and identification scheme. The abnormal readings are novelly detected by combinational sparsity CS reconstruction. Specific features from real trace data are utilized to distinguish the external events and internal errors. Experimental results on real trace data validate the efficacy and efficiency of our scheme.

Since the property of information-dense for the random measurements in CS, the loss of measured sensor data will severely degrade the quality of data reconstruction. However, because of the unreliability of the sensor networks, packet loss is prevalent in a deployed sensor networks, which poses a great challenge to the application of the CS based data gathering in WSN. How to address this problem is of great significance, and we will investigate it in our future work.

Acknowledgment

This work was supported by the National Basic Research Program of China (973 Program) under Grant No. 2011CB302703, 2010CB328100, 2010CB334707, 2011CB302705, National Natural Science Foundation of China under Grant No. 61033004, 60825203, 60973056, 61170103, 61170216, and Beijing Natural Science Foundation under Grant No. 4102009. The research of authors is partially supported by NSF CNS-0832120, NSF CNS-1035894, Tsinghua National Laboratory for Information Science and Technology (TNList), program for Zhejiang Provincial Key Innovative Research Team, and program for Zhejiang Provincial Overseas High-Level Talents (One-hundred Talents Program).

References

[1] M. Li and Y. Liu, "Underground coal mine monitoring with wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, no. 2, p. 10, 2009.

[2] G. Hua and C. Chen, "Correlated data gathering in wireless sensor networks based on distributed source coding," *International Journal of Sensor Networks*, vol. 4, no. 1, pp. 13–22, 2008.

[3] J. Chou, D. Petrovic, and K. Ramachandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *INFOCOM 2003. IEEE Societies Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, vol. 2. IEEE, 2003, pp. 1054–1062.

[4] K. Yuen, B. Liang, and L. Baochun, "A distributed framework for correlated data gathering in sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 578–593, 2008.

[5] A. Ciancio, S. Pattem, A. Ortega, and B. Krishnamachari, "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 309–316.

[6] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *INFOCOM 2003. IEEE Societies Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, vol. 3. IEEE, 2003, pp. 1848–1857.

[7] S. Yoon and C. Shahabi, "The clustered aggregation (cag) technique leveraging spatial and temporal correlations in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 3, no. 1, pp. 3–es, 2007.

[8] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation," *IEEE Trans. on Parallel and Distributed Systems*, pp. 1010–1023, 2007.

[9] H. Gupta, V. Navda, S. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor networks," in *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2005, pp. 402–413.

[10] X. Xu, X. Li, P. Wan, and S. Tang, "Efficient scheduling for periodic aggregation queries in multihop sensor networks," *IEEE/ACM Transactions on Networking*, no. 99, pp. 1–1, 2011.

[11] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 134–142.

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

[13] C. Luo, F. Wu, J. Sun, and C. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of the 15th annual international conference on Mobile computing and networking*. ACM, 2009, pp. 145–156.

[14] J. Luo, L. Xiang, and C. Rosenberg, "Does compressed sensing improve the throughput of wireless sensor networks?" in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–6.

[15] L. Xiang, J. Luo, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," *Proc. of the 8th IEEE SECON (to appear)*, 2011.

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

[17] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

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

[19] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3. Citeseer, 2006, pp. 1433–1452.

[20] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[21] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[22] M. Iwen, "Simple deterministically constructible rip matrices with sublinear fourier sampling requirements," in *43rd Annual Conference on Information Sciences and Systems*. IEEE, 2008, pp. 870–875.

[23] S. Wright, *Primal-dual interior-point methods*. Society for Industrial Mathematics, 1997, vol. 54.

[24] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[25] D. Donoho, I. Drori, Y. Tsaig, and J. Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit*. Citeseer, 2006.

[26] E. Candès and J. Romberg, "Practical signal recovery from random projections," *IEEE Trans. on Signal Processing*, 2005.

[27] NBDC CTD data. [Online]. Available: http://tao.noaa.gov/refreshed/ctd_delivery.php

[28] 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 *Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 873–881.

[29] C. Chou, R. Rana, and W. Hu, "Energy efficient information collection in wireless sensor networks using adaptive compressive sensing," in *IEEE 34th Conference on Local Computer Networks, 2009*. IEEE, 2009, pp. 443–450.

[30] D. Malioutov, S. Sanghavi, and A. Willsky, "Compressed sensing with sequential observations," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008*. IEEE, 2008, pp. 3357–3360.

[31] E. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *Rice University Department of Computational and Applied Mathematics, Tech. Rep*, 2007.