Compressive Data Gathering for Large-Scale Wireless Sensor Networks

Chong Luo
Shanghai Jiao Tong University
No.800 Dongchuan Road
Shanghai, CHINA 200240
Microsoft Research Asia
No.49 Zhichun Road
Beijing, CHINA 100190
chong.luo@microsoft.com

Feng Wu
Microsoft Research Asia
No.49 Zhichun Road
Beijing, CHINA 100190
fengwu@microsoft.com

Jun Sun
Shanghai Jiao Tong University
No.800 Dongchuan Road
Shanghai, CHINA 200240
junsun@sjtu.edu.cn

Chang Wen Chen
SUNY at Buffalo
201 Bell Hall, Buffalo
NY 14260-2000, USA
chencw@buffalo.edu

ABSTRACT

This paper presents the first complete design to apply compressive sampling theory to sensor data gathering for large-scale wireless sensor networks. The successful scheme developed in this research is expected to offer fresh frame of mind for research in both compressive sampling applications and large-scale wireless sensor networks. We consider the scenario in which a large number of sensor nodes are densely deployed and sensor readings are spatially correlated. The proposed compressive data gathering is able to reduce global scale communication cost without introducing intensive computation or complicated transmission control. The load balancing characteristic is capable of extending the lifetime of the entire sensor network as well as individual sensors. Furthermore, the proposed scheme can cope with abnormal sensor readings gracefully. We also carry out the analysis of the network capacity of the proposed compressive data gathering and validate the analysis through ns-2 simulations. More importantly, this novel compressive data gathering has been tested on real sensor data and the results show the efficiency and robustness of the proposed scheme.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network communications, Wireless communication*

General Terms

Design, Performance

MobiCom'09, September 20–25, 2009, Beijing, China.

Keywords

Wireless Sensor Networks, Compressive Sampling

1. INTRODUCTION

This paper considers the data gathering problem in a large-scale wireless sensor network. Data gathering sensor network finds a variety of applications in infrastructure and habitat monitoring [8][23]. It is expected that the number of sensor nodes deployed could be on the order of hundreds or thousands. In general, data transmissions are accomplished through multi-hop routing from individual sensor nodes to the data sink. Successful deployment of such large scale sensor networks faces two major challenges in effective global communication cost reduction and in energy consumption load balancing.

The need for global communication cost reduction is obvious because such sensor networks typically are composed of hundreds to thousands of sensors, generating tremendous amount of sensor data to be delivered to data sink. It is very much desired to take full advantage of the correlations among the sensor data to reduce the cost of communication. Existing approaches adopt in-network data compression, such as entropy coding or transform coding, to reduce global traffic. However, these approaches introduce significant computation and control overheads that often not suitable for sensor networks applications.

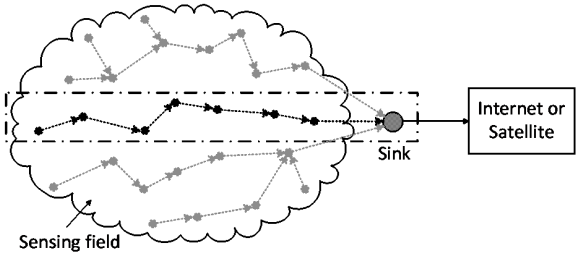
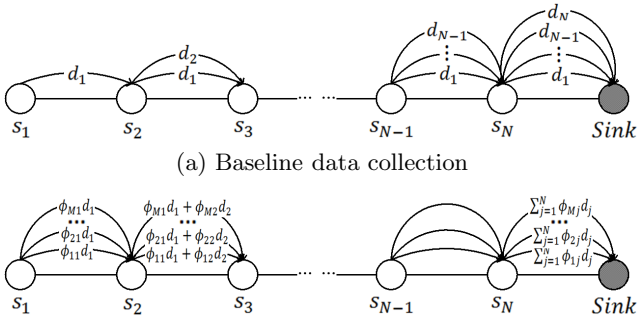


Figure 1: Data gathering sensor network



(a) Baseline data collection

(b) Compressive data gathering

Figure 2: Comparing baseline data collection and compressive data gathering in a multi-hop route

The need for energy consumption load balancing is also clear because of the required multi-hop data transmission for such large scale sensor networks. Fig. 1 shows such a network where sensors are densely deployed in the region of interest and monitor the environment on a regular basis. A simple but typical example is the highlighted route in Fig. 1. Suppose N sensors, denoted as s_1, s_2, ..., and s_N form a multi-hop route to the sink. Let d_j denote the readings obtained by node s_j. The intuitive way to transmit d_j, $j = 1, 2, ...N$ to the sink is through multi-hop relay as depicted in Fig. 2(a). Node s_1 transmits its reading d_1 to s_2, and s_2 transmits both its reading d_2 and the relayed reading d_1 to s_3. At the end of the route, s_N transmits all N readings to the sink. It can be observed that the closer a sensor is to the sink, the more energy is consumed. Clearly, the sensor nodes closer to the data sink will soon run out of energy and lifetime of sensor network will be significantly shortened.

This paper presents the first complete design to apply compressive sampling theory [13][4][7] to sensor data gathering for large-scale wireless sensor networks (WSNs), successfully addressing the two major challenges as outlined above. First, the proposed data gathering is able to achieve substantial sensor data compression without introducing excessive computation and control overheads. With elegant design, the proposed scheme is also able to disperse the communication costs to all sensor nodes along a given sensor data gathering route. This will result in a natural load balancing and extend the lifetime of the sensor network.

The basic idea of the proposed compressive data gathering (CDG) is depicted in Fig. 2(b). Instead of receiving individual sensor readings, the sink will be sent a few weighted sums of all the readings, from which to restore the original data. To transmit the i^{th} sum to the sink, s_1 multiplies its reading d_1 with a random coefficient ϕ_{i1} and sends the product to s_2. Upon receiving this message, s_2 multiplies its reading d_2 with a random coefficient ϕ_{i2} and then sends the sum $\phi_{i1}d_1 + \phi_{i2}d_2$ to s_3. Similarly, each node s_j contributes to the relayed message by adding its own product. Finally, the sink receives $\sum_{j=1}^{N} \phi_{ij}d_j$, a weighted sum of all the readings. This process is repeated using M sets of different weights so that the sink will receive M weighted sums.

With such design, all nodes transmit M messages and consume same amount of energy. Each node only performs one addition and one multiplication in order to compute one weighted sum. Comparing Fig. 2(a) and Fig. 2(b), careful readers will observe that, the first M nodes send more messages in CDG than in baseline transmission, while the rest of nodes send less messages in CDG. When N is large and M is much smaller than N, CDG can significantly reduce the total number of transmissions and save energy. The key problem now becomes whether the sink is able to restore N individual readings from M measurements when M is far smaller than N. Fortunately, the compressive sampling theory has a positive answer to this question.

This paper makes three main contributions. First, we extend the application of compressive sampling theory from one or a few sensors to large-scale multi-hop sensor networks. Beyond the basic idea, we propose a scheme which allows CDG to be practically applied to large sensor networks. Second, we carry out a theoretical analysis of the network capacity for CDG and validate the capacity gain of CDG through ns-2 simulations. Third and more importantly, we test CDG on two sets of real sensor data. The results show that CDG is practically applicable to various data gathering sensor networks. Even when sensor data exhibit little spatial correlations in which case conventional in-network compression approaches would fail, CDG is still able to reduce the traffic of bottleneck node by two to three times and significantly prolong the network lifetime.

The rest of this paper is organized as follows: Section II reviews related work on energy-efficient data gathering. Section III describes the proposed CDG scheme. Section IV presents the analysis of the network capacity for CDG and the results of ns-2 simulations. Section V demonstrates the test results on two sets of real sensor data. Section VI concludes this paper with some discussions.

2. RELATED WORK

The fundamental assumption of in-network data compression is that sensor nodes have spatial correlations in their readings. According to where the spatial correlation is utilized, we can classify existing in-network data compression techniques into two categories.

2.1 Conventional Compression

Conventional compression techniques utilize the correlation during the encoding process and require explicit data communication among sensors. Cristescu et al. [12] propose a joint entropy coding approach, where nodes use relayed data as side information to encode their readings. Again take the multi-hop route in Fig. 2 as an example. First, node s_1 encodes its reading d_1 into message p_1 using $H(d_1)$ bits, where $H(d_1)$ is the entropy of d_1. Then, when s_2 receives p_1, it encodes its reading d_2 into message p_2 using $H(d_2|d_1)$ bits, where $H(d_2|d_1)$ is the conditional entropy. Since d_1 and d_2 are correlated, $H(d_2|d_1)$ is smaller than $H(d_2)$. Therefore, jointly encoded messages cost less bits than independently encoded messages.

The above approach utilizes data correlation only unidirectionally. If data are allowed to be communicated back and forth during encoding, nodes may cooperatively perform transform to better utilize the correlation. Ciancio et al. [10] and Aćimović et al. [2] propose to compress piecewise smooth data through distributed wavelet transform. In doing so, even nodes first broadcast their readings. Upon receiving the readings from both sides, odd nodes compute the high pass coefficients $h(\cdot)$. Then, odd nodes transmit $h(\cdot)$ back and even nodes compute the low pass coefficients $l(\cdot)$.

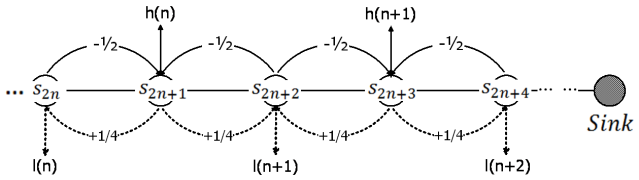


Figure 3: Cooperative wavelet compression

This process is illustrated in Fig. 3. Although wavelet decorrelation can be performed for multiple levels, it is not suggested to do so in distributed processing because of the communication overhead. After the transform, nodes transmit significant coefficients to the sink, usually in their raw form to avoid the complexity of entropy coding.

Quantization of a group of readings to one representative value is another form of conventional compression. The clustered aggregation (CAG) technique [26] forms clusters based on sensing values. By grouping sensors with similar readings, CAG only transmits one reading per group to achieve a predefined error threshold. Gupta et al. [15] exploit a similar idea. In each round of data gathering, it only involves a subset of nodes, which is sufficient to reconstruct data for the entire network.

There are two main problems with conventional compression techniques. First, the compression performance relies heavily on how the routes are organized. In order to achieve the highest compression ratio, compression and routing algorithms need to be jointly optimized. This has been proved to be an NP-hard problem [12]. Second, the efficiency of an in-network data compression scheme is not solely determined by the compression ratio, but also depends on the computational and communication overheads. However, joint entropy coding techniques perform complex computation in sensors, while transform based techniques require a large amount of data exchanges.

2.2 Distributed Source Coding

Distributed source coding techniques [9][11][18] intend to reduce complexity at sensor nodes and utilize correlation at the sink. They are based on the Slepian-Wolf coding theory [22], which claims that compression of correlated readings, when separately encoded, can achieve same efficiency as if they are jointly encoded, provided that messages are jointly decoded. This important conclusion not only eliminates data exchanges, but decouples routing from compression. After encoding sensor readings independently, each node simply sends the compressed message along the shortest path to the sink.

However, a prerequisite of Slepian-Wolf coding is that the global correlation structure needs to be known in order to allocate appropriate number of bits to be used by each node. This is hard to fulfill in a large-scale wireless sensor network. In view of this, Yuen et al. [27] adopts a localized Slepian-Wolf coding scheme. Based on the assumption that sensors outside immediate neighborhood have weak correlation in their readings, a node may only consider its data correlation with one-hop neighbors when determining the size of encoded message. We will show that, for a set of real sensor data which do not satisfy this assumption, the localized coding scheme will fail to compress such data.

Distributed source coding techniques perform well for static correlation patterns. However, when correlation pattern changes or abnormal events show up, the decoding accuracy will be greatly affected. Since detecting abnormal events is an important task of sensor network, when an abnormal event is captured by a side node, the originally assigned number of bits will be inadequate to encode the reading, and cause decoding error at the sink. More seriously, when the abnormal reading appears at a main node, it will cause errors within a large range of reconstructed sensor readings.

2.3 Compressive Sampling

With the emergence of compressive sampling theory [13] [4] [7], we have seen a new avenue of research in the field of in-network data compression. Compressive wireless sensing (CWS) [3] appears to be able to reduce the latency of data gathering in a single-hop network by delivering linear projections of sensor readings through synchronized amplitude-modulated analog transmissions. Due to the difficulties in analog synchronization, CWS is less practical for large-scale sensor networks. Rabbat et al. [21] leverages compressive sampling for data persistence, instead of data gathering, in a WSN. In an overview paper, Haupt et al. [17] also speculate the potential of using compressive sampling theory for data aggregation in a multi-hop WSN. However, no real scheme has been reported based on this initial idea.

When compressive sampling is applied to in-network data compression, it will bring a wealth of similar benefits as distributed source coding including simple encoding process, saving of inter-node data exchange, and decoupling of compression from routing. Furthermore, compressive sampling has two additional advantages. First, it can deal with abnormal sensor readings gracefully. This advantage will be detailed in the next section. Second, data reconstruction is not sensitive to packet losses. In compressive sampling, all messages received by the sink are equally important. However, in distributed source coding, received data are predefined as main or side information. Losing main information will cause fatal errors to the decoder. All these desired merits make compressive sampling a promising solution to the data gathering problem in large-scale wireless sensor networks.

3. COMPRESSIVE DATA GATHERING

The objective of compressive data gathering is two-fold: compress sensor readings to reduce global data traffic and distribute energy consumption evenly to prolong network lifetime. Similar to distributed source coding, the data correlation pattern shall be utilized on the decoder end. Besides, compression and routing are decoupled and therefore can be separately optimized.

3.1 Data gathering

The intuition behind CDG is that higher efficiency can be achieved if correlated sensor readings are transmitted jointly rather than separately. We have given a simple example in Section I, showing how sensor readings are combined while being relayed along a chain-type topology to the sink. In practice, sensors usually spreads in a two-dimensional area, and the ensemble of routing paths presents a tree structure. Fig. 4(a) shows a typical routing tree in which the sink has four children. Each of them leads a subtree delimited by the dotted lines. Data gathering and reconstruction of CDG are performed on the subtree basis.

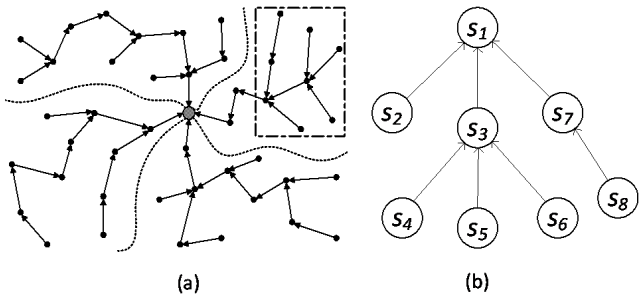


Figure 4: Data gathering in a typical routing tree

In order to combine sensor readings while relaying them, every node needs to know its local routing structure. That is, whether or not a given node is a leaf node in the routing tree or how many children the node has if it is an inner node. To facilitate efficient aggregation, we have made a small modification to standard ad-hoc routing protocol: when a node chooses a parent node, it sends a "subscribe notification" to that node; when a node changes parent, it sends an "unsubscribe notification" to the old parent.

The data gathering process of CDG is illustrated through an example shown in Fig. 4(b). It is the detailed view of a small fraction of the routing tree marked in Fig. 4(a). After all nodes acquire their readings, leaf nodes initiate the transmission. In this example, s_2 generates a random number ϕ_{i2}, computes $\phi_{i2}d_2$, and transmits the value to s_1. The index i denotes the i^{th} weighted sum ranging from 1 to M. Similarly, s_4, s_5 and s_6 transmits $\phi_{i4}d_4$, $\phi_{i5}d_5$, and $\phi_{i6}d_6$ to s_3. Once s_3 receives the three values, it computes $\phi_{i3}d_3$, adds it to the sum of relayed values and transmits $\sum_{j=3}^{6}\phi_{ij}d_j$ to s_1. Then s_1 computes $\phi_{i1}d_1$ and transmits $\sum_{j=1}^{8}\phi_{1j}d_j$. Finally, the message containing the weighted sum of all readings in a subtree is forwarded to the sink.

Assume that there are N nodes in a particular tree, and the sink intends to collect M measurements. Then all nodes send the same number of $O(M)$ messages regardless of their hop distance to the sink. The overall message complexity is $O(NM)$. When $M \ll N$, CDG transmits less messages than the baseline data collection (as shown in Fig. 2(a)) whose worst case message complexity is $O(N^2)$. More importantly, the transmission load is spread out uniformly so that the lifetime of bottleneck sensors and the entire network is greatly extended.

The i^{th} weighted sum can be represented by:

$$y_i = \sum_{j=1}^{N} \phi_{ij} d_j \tag{1}$$

The sink obtains M weighted sums $\{y_i\}$, $i = 1, 2, ...M$. Mathematically, we have:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & ... & \phi_{1N} \\ \phi_{21} & \phi_{22} & ... & \phi_{2N} \\ \vdots & & \vdots & \\ \phi_{M1} & \phi_{M2} & ... & \phi_{MN} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \\ \vdots \\ d_N \end{pmatrix} \tag{2}$$

In this equation, each column of $\{\phi_{ij}\}$ contains the series of random numbers generated at a corresponding node. In order to avoid transmitting this random matrix from sensors to the sink, we can adopt a simple strategy: before data transmission, the sink broadcasts a random seed to the entire network. Then each sensor generates its own seed using this global seed and its unique identification. With a pre-installed pseudo random number generator, each sensor is able to generate the corresponding series of coefficients. These coefficients can be reproduced at the sink given that the sink knows the identifications of all sensors.

In (2), d_i $(i = 1, 2, ...N)$ is a scalar value. In a practical sensor network, each node is possibly attached with a few sensors of different type, e.g. a temperature sensor and a humidity sensor. Then sensor readings from each node become a multi-dimensional vector. In this case, we may separate readings of each dimension and process them respectively. Alternatively, since the random coefficients ϕ_{ij} are irrelevant to sensor readings, we may treat d_i as a vector. The weighted sums y_i become vectors of the same dimension too.

When $M < N$, solving a set of M linear equations with N unknown variables is an ill-posed problem. However, sensor readings are not independent variables. In most cases, the sensor field follows a certain structure because of the spatial or temporal correlations. Hence, there exists a transform domain in which the signal is sparse. Under this assumption, we will explain in the following subsection whether the set of linear equations are solvable, what requirements M should meet to solve them, and how these equations can be solved.

3.2 Data recovery

3.2.1 *Recover spatially correlated data*

According to compressive sampling theory, a K-sparse signal can be reconstructed from a small number of measurements with a probability close to one. The weighted sums obtained in (2) are a typical type of measurements. Signal sparsity characterizes the correlations within a signal. An N-dimensional signal is considered as a K-sparse signal if there exists a domain in which this signal can be represented by K $(K \ll N)$ non-zero coefficients. Fig. 5(a) shows a 100-dimensional signal in its original time domain. Obviously, it is not sparse at all in this domain. Because of the signal correlation, it can be described more compactly in transform domains such as wavelet and DCT. Fig. 5(b) gives the representation of the same signal in DCT domain. We can see that there are only 5 non-zero DCT coefficients. Therefore, this signal is a 5-sparse signal in DCT domain.

In a densely deployed sensor networks, sensors have spatial correlations in their readings. Let N sensor readings form a vector $\mathbf{d} = [d_1\ d_2\ ...\ d_N]^T$, then $\mathbf{d}$ is a K-sparse signal in a particular domain Ψ. Denote $\Psi = [\psi_1\ \psi_2\ ...\psi_N]$ as the representation basis with vectors $\{\psi_i\}$ as columns, and $\mathbf{x} = [x_1, x_2, ...x_N]^T$ are the corresponding coefficients. Then, $\mathbf{d}$ can be represented in the Ψ domain as:

$$\mathbf{d} = \sum_{i=1}^{N} x_i\psi_i, \text{ or } \mathbf{d} = \Psi\mathbf{x} \tag{3}$$

Compressive sampling theory tells that a K-sparse signal can be reconstructed from M measurements if M satisfies the following conditions [6]:

$$M \geq c \cdot \mu^2(\Phi, \Psi) \cdot K \cdot \log N \tag{4}$$

where c is a positive constant, Φ is the sampling matrix as defined in (2), and $\mu(\Phi, \Psi)$ is the coherence between sampling basis Φ and representation basis Ψ. The coherence

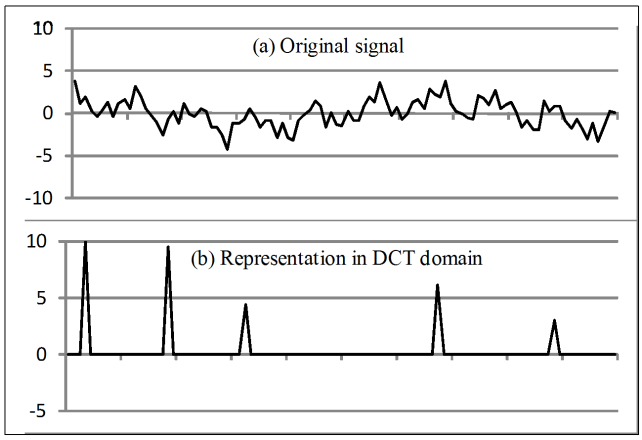


Figure 5: A 5-sparse signal in DCT domain

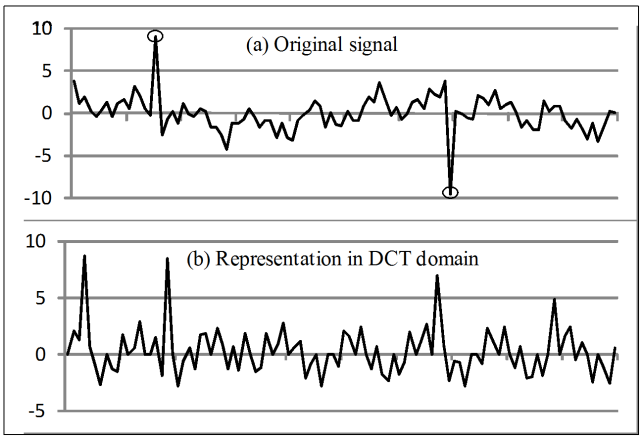


Figure 6: A signal with two abnormal readings

metric measures the largest correlation between any two elements of Φ and Ψ, and is defined as:

$$\mu(\Phi, \Psi) = \sqrt{N} \cdot \max_{1 \le i,j \le N} |\langle \phi_i, \psi_j \rangle| \quad (5)$$

From (5), we can see that the smaller the coherence between Φ and Ψ is, the less measurements are needed to reconstruct the signal. In practice, using random measurement matrix is a convenient choice, since a random basis has been shown to be largely incoherent with any fixed basis, and $M = 3K \sim 4K$ is usually sufficient to satisfy (4).

With sufficient number of measurements, the sink is able to reconstruct sensor readings through solving an l_1-minimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_{l_1} \quad s.t. \quad \mathbf{y} = \Phi\mathbf{d},\ \mathbf{d} = \Psi\mathbf{x} \quad (6)$$

In addition, for sparse signals whose random projections are contaminated with noise, reconstruction can be achieved through solving a relaxed l_1-minimization problem, where ϵ is a predefined error threshold:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_{l_1} \quad s.t. \quad \|\mathbf{y} - \Phi\mathbf{d}\|_{l_2} < \epsilon,\ \mathbf{d} = \Psi\mathbf{x} \quad (7)$$

Suppose $\tilde{\mathbf{x}}$ is the solution to this convex optimization problem, then the proposed reconstruction of the original signal is $\tilde{\mathbf{d}} = \Psi\tilde{\mathbf{x}}$. It has been shown that the above l_1-minimization problem can be solved with linear programming (LP) techniques [13]. Although the reconstruction complexity of LP based decoder is polynomial, it goes pretty high when N is too large. While there is a large body of on-going work looking for low-complexity reconstruction techniques [25][5], this topic is beyond the scope of our paper. With the current LP based decoder, we would suggest that the size of N does not exceed one thousand.

In (6) and (7), the Ψ matrix describes the correlation pattern among sensor readings. It is utilized only in data recovery process, and is not required to be known to sensors. In this way, most of the computations are shifted from sensors to the sink. Such asymmetry of computation complexity makes CDG an appealing choice for WSNs.

3.2.2 Recover data with abnormal readings

One of the main purposes of sensor network is to monitor abnormal events. However when abnormal events take place, the sparsity of sensor readings is compromised. As an example, Fig. 6(a) differs with Fig. 5(a) only by two abnormal readings, as outlined by the ovals. The corresponding DCT coefficients shown in Fig. 6(b) are not sparse any more. Therefore, the signal in Fig. 6 is not sparse in either time domain or transform domain. In this situation, conventional compression techniques need to transmit significantly more data in order to reconstruct the original signal. Distributed source coding techniques will have a big degradation.

We have a better solution in compressive data gathering. Sensor data with abnormal readings can be decomposed into two vectors:

$$\mathbf{d} = \mathbf{d_0} + \mathbf{d_s} \quad (8)$$

Where $\mathbf{d_0}$ contains the normal readings which are sparse in a certain transform domain, and $\mathbf{d_s}$ contains the deviated values of abnormal readings. Since abnormal readings are sporadic, $\mathbf{d_s}$ is a sparse signal in the time domain. Suppose the normal readings are sparse in Ψ domain, then (8) can be rewritten into:

$$\mathbf{d} = \Psi\mathbf{x_0} + I\mathbf{x_s} \quad (9)$$

Where I is the identical matrix, and both $\mathbf{x_0}$ and $\mathbf{x_s}$ are sparse. We can see that signal $\mathbf{d}$ is decomposed into two signals which are sparse in different domains. We can construct an overcomplete basis $\Psi' = [\Psi\ I]$, then $\mathbf{d}$ should be sparse in Ψ' domain:

$$\mathbf{d} = \Psi'\mathbf{x},\ \mathbf{x} = [\mathbf{x_0}^T \mathbf{x_s}^T]^T \quad (10)$$

Incorporating (10) into (6) or (7), the signal recovery with abnormal readings can be solved similarly by the l1-norm optimization. Donoho et al. [14] showed the possibility of stable recovery under a combination of sufficient sparsity and favorable structure of the overcomplete system. Moreover, they also proved that stable recovery of sparse signal in an overcomplete dictionary also works for noisy data, and the optimally-sparse approximation to the noisy data, to within the noise level, differs from the optimally-sparse decomposition of the ideal noiseless signal by at most a constant multiple of the noise level.

Suppose $\tilde{\mathbf{x}}$ is a vector of length $2N$, and is the solution to the l_1-minimization problem defined in (7) when an overcomplete dictionary is used. Similarly, the original sensor readings can be reconstructed by $\tilde{\mathbf{d}} = \Psi'\tilde{\mathbf{x}}$. Denote $\tilde{\mathbf{x_s}}$ as

an N-dimensional vector composed of the last N elements of $\tilde{\mathbf{x}}$, then the non-zero values in $\tilde{\mathbf{x}}_\mathbf{s}$ indicate the positions of abnormal readings.

4. NETWORK CAPACITY OF COMPRESSIVE DATA GATHERING

The previous section illustrated how to gather and recover sensor readings acquired in one time instance. This section will investigate the benefit of CDG from the viewpoint of network capacity, i.e. how frequent CDG allows sensors to acquire data while ensuring all readings can be transmitted to the sink. The capacity of a data gathering network is defined as follows.

DEFINITION 1 (NETWORK CAPACITY). *We shall define that a rate λ is achievable in a data gathering sensor network, if there exists a time instance t_0 and duration T such that during $[t_0, t_0+T)$ the sink receives λT bits of data generated by each of the sensors s_i, $i = 1, 2, ...N$. Then, network capacity C is defined as the supremum of the achievable rate, or $C = sup\{\lambda\}$.*

Different from the pioneering work on network capacity analysis [16], the traffic pattern in our study is many-to-one. We let all sensors generate data at the same rate, and assume that sensor readings acquired at the same time instance are K-sparse.

4.1 Network Capacity Analysis

We assume a discal sensing area in which N sensor nodes are uniformly distributed, and the sink is located in the middle of the disk. All sensor nodes and the sink communicate over single frequency shared radio channel, accessed through time-division multiple access control (TDMA). We denote W as the amount of data a node transmits in one time slot, and restrict that a node cannot transmit and receive at the same time.

Let $\{X_k,\ k \in V\}$ be the subset of nodes simultaneously transmitting over the shared channel in a specific time slot. Then a successful transmission from X_i, $i \in V$ to X_j can be defined under two interference models.

DEFINITION 2 (PROTOCOL MODEL). *Transmission from node X_i to X_j is successful under protocol model if and only if the following two conditions are satisfied:*

- $\|X_i - X_j\| \leq r$
- $\|X_k - X_j\| > (1+\delta)r$, $\delta > 0$ *for* $k \in V - \{i\}$

The first condition requires that the two communicating nodes are within a distance r. The second condition requires that the receiving node is at least $(1+\delta)r$ away from any other transmitting nodes.

DEFINITION 3 (PHYSICAL MODEL). *Transmission from node X_i to X_j is successful under physical model if and only if:*

$$\frac{\frac{P_i}{\|X_i - X_j\|^\alpha}}{N_G + \sum_{k \in V, k \neq i} \frac{P_k}{\|X_k - X_j\|^\alpha}} \geq \beta$$

where P_i is the transmission power for X_i, α is the fading parameter and N_G is noise power level. The expression on the left is the signal to interference and noise ratio (SINR) at the receiving node. A successful transmission under physical model requires the SINR to be greater than a predefined threshold β.

4.1.1 Capacity under Protocol Model

The capacity under protocol model can be analyzed in a similar way as Marco et al. [19]. Let us first recall the following lemma.

LEMMA 1. *N nodes are uniformly distributed in a region of area A. When N is large, the number of nodes n within a sub-region R of area A_R can be bounded with high probability.*

$$Pr\left(\frac{NA_R}{A} - \sqrt{\alpha_N N} \leq n \leq \frac{NA_R}{A} + \sqrt{\alpha_N N}\right) \to 1 \quad as\ N \to \infty.$$

Sequence α_N is chosen such that $\alpha_N \to \infty$ as $N \to \infty$, and $\lim_{N\to\infty} \frac{\alpha_N}{N} = \varepsilon$, where ε is positive but arbitrarily small.

PROOF. Each of the N nodes has the same probability A_R/A to fall in region R. Therefore, n follows binomial distribution with the mean being $\mu_n = \frac{NA_R}{A}$ and the variance being $\delta_n^2 = \frac{NA_R}{A}(1 - \frac{A_R}{A})$. According to Chebychev's inequality:

$$Pr(|n - \mu_n| \geq \sqrt{\alpha_N N}) \leq \frac{\delta_n^2}{\alpha_N N} = \frac{\frac{A_R}{A} \cdot (1 - \frac{A_R}{A})}{\alpha_N} \quad (11)$$

The probability goes to 0 as $N \to \infty$ and $\alpha_N \to \infty$. □

THEOREM 1. *In a wireless sensor network with N uniformly distributed nodes, compressive data gathering can achieve network capacity of $\lambda \geq \frac{W}{M} \frac{\pi r^2 - \sqrt{\varepsilon}}{\pi(2+\delta)^2 r^2 + \sqrt{\varepsilon}}$ with a probability close to 1 as $N \to \infty$, where ε is arbitrarily close to 0, and M is the number of random measurements. Usually $M = c_1 K$, and c_1 is a constant in the range of $[1, 4]$.*

PROOF. Consider a node in transmission. According to definition 2, the distance from any interfering source to this node is at most $(2+\delta)r$. In other words, all the interfering sources are contained in a disk of area $A_{R_1} = \pi(2+\delta)^2 r^2$. Based on Lemma 1, the number of nodes in this region, denoted by n_{itf}, is less than n_1 with high probability:

$$n_{itf} \leq n_1 = \frac{NA_{R_1}}{A} + \sqrt{\alpha_N N} \quad (12)$$

Next we build the contention graph of the network by connecting interfering nodes. With high probability, the maximal node degree in the contention graph is $n_1 - 1$. According to graph coloring theory, all nodes can be colored with at most n_1 different colors. If we associate each color with a transmission slot, every node gets one chance to transmit in n_1 slots. Therefore, the average transmission rate of each node is:

$$\gamma = \frac{W}{n_1} \quad (13)$$

Then consider the one-hop neighbors of the sink. They are contained in a disk centered at the sink and with a radius of r. The area of the disk is $A_{R_2} = \pi r^2$. According to Lemma 1, the number of nodes in this region, denoted by n_2, can be bounded with high probability:

$$\frac{NA_{R_2}}{A} - \sqrt{\alpha_N N} \leq n_2 \leq \frac{NA_{R_2}}{A} + \sqrt{\alpha_N N} \quad (14)$$

Recall that compressive data gathering is performed on subtree basis. We shall adopt an appropriate routing protocol such that all subtrees are roughly of equal size. For simplicity, we consider the size of each subtree is $N_p = \frac{N}{n_2}$. Since the sensor data from the entire network is K-sparse, when $N \to \infty$, we can consider that each subset of the nodes are proportionally sparse, i.e. K/n_2-sparse. The number of random measurements needed to reconstruct data is M/n_2 per subtree. To achieve the rate λ, the transmission rate of the subtree root should be $M\lambda/n_2$. Take (13) into account, we have:

$$\frac{W}{n_1} = \frac{M\lambda}{n_2} \tag{15}$$

Substituting (12) and (14) into (15), we have:

$$\begin{aligned}\lambda = \frac{W}{M}\frac{n_2}{n_1} &\geq \frac{W}{M}\frac{\frac{NA_{R_2}}{A} - \sqrt{\alpha_N N}}{\frac{NA_{R_1}}{A} + \sqrt{\alpha_N N}} = \frac{W}{M}\frac{\frac{A_{R_2}}{A} - \sqrt{\varepsilon}}{\frac{A_{R_1}}{A} + \sqrt{\varepsilon}} \\ &= \frac{W}{M}\frac{\pi r^2 - \sqrt{\varepsilon}}{\pi(2+\delta)^2 r^2 + \sqrt{\varepsilon}}\end{aligned} \tag{16}$$

As $N \to \infty$, $\sqrt{\varepsilon} \to 0$, the lower bound of achievable capacity is arbitrarily close to $\frac{W}{M(2+\delta)^2}$. □

4.1.2 Capacity under Physical Model

Without loss of generality, we assume the following constraints for the physical model:

- All nodes transmit with equal and finite power P_0.
- All noises are of the same variance. Therefore, for a given small positive number η, there exist a noise level N_0 such that $prob(N_0 > N_G) < \eta$.
- Given α and β, P_0 is chosen such that the network is a connected graph when the noise level is N_0.

THEOREM 2. *In a wireless sensor network with N uniformly distributed nodes, compressive data gathering can achieve network capacity of $\lambda \geq \frac{W}{M}\frac{\pi r_0^2 - \sqrt{\varepsilon}}{\pi(2+\delta_0)^2 r_0^2 + \sqrt{\varepsilon}}$ with a probability close to 1 as $N \to \infty$, given $r_0 < \sqrt[\alpha]{\frac{P_0}{\beta N_0}}$ and $\delta_0 > \sqrt[\alpha-1]{\frac{2\pi\beta c_2}{1-\beta r^\alpha N_0/P_0}} - 1$.*

PROOF. Theorem 1 gives the network capacity under protocol model. We will prove that when $r = r_0$ and $\delta = \delta_0$, a feasible transmission schedule under protocol model is also feasible under physical model.

First, we restrict the communication to nodes within a distance of r_0. When node X_i transmits data to node X_j, the SINR at X_j is:

$$SINR_j = \frac{\frac{P_0}{|X_i - X_j|^\alpha}}{N_0 + \sum_{k \in V, k \neq i} \frac{P_0}{|X_k - X_j|^\alpha}} \tag{17}$$

Denote P_s as the received signal strength and P_f as the interference strength at node X_j. Since $|X_i - X_j| < r_0$, we have:

$$P_s = \frac{P_0}{|X_i - X_j|^\alpha} > \frac{P_0}{r_0^\alpha}. \tag{18}$$

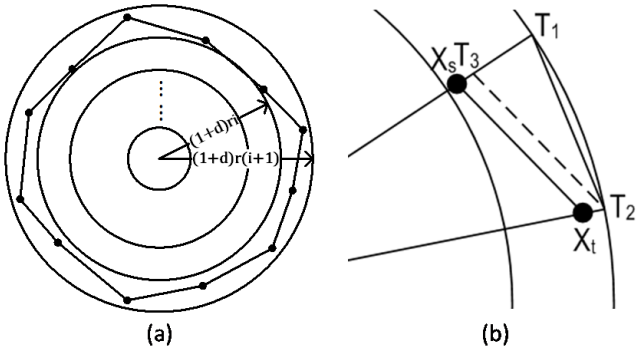


Figure 7: Connecting adjacent transmitting nodes in an annulus

A necessary condition is that when $P_f = 0$, the selection of r_0 should ensure $SINR_j > \beta$. This can be satisfied if $r_0 < \sqrt[\alpha]{\frac{P_0}{\beta N_0}}$.

Next, let us look at the interference part P_f. A feasible schedule under protocol model ensures that there is no other simultaneous transmitter in the circular area centered at X_j and with a radius of $(1+\delta)r$. The interference comes from the transmitters outside this region. Divide the sensing region by concentric circles C_i, $i = 1, 2, ...$ centered at X_j. The radius of circle C_i is $r_i = (1+\delta)ri$,. Denote A_i as the annulus formed by C_i and C_{i+1}. Next, we quantify the interference to X_j caused by the transmitters in each annulus.

Denote a_i as the number of simultaneous transmitters within a particular annulus A_i. Since the distance from X_j to any node in this annulus is larger than $(1+\delta)ri$, the interference from this annulus is:

$$P_f(A_i) = \sum_{k, X_k \in A_i} \frac{P_0}{|X_k - X_j|^\alpha} < \frac{a_i P_0}{((1+\delta)ri)^\alpha} \tag{19}$$

$P_f(A_i)$ can be bounded once a_i is bounded. In doing so, we connect adjacent transmitters clockwise with line segments, as Fig. 7(a) shows. Fig. 7(b) gives an enlarged view of two adjacent transmitting nodes X_s and X_t. Connect the center of circle with the two nodes and extend the lines so that they intersect C_{i+1} at points T_1 and T_2. From T_2 draw a line parallel to $\overline{X_s X_t}$ and intersect $\overline{X_s T_1}$ at point T_3. Then we have:

$$|\overline{X_s X_t}| \leq |\overline{T_2 T_3}| < |\overset{\frown}{T_1 T_2}| + (1+\delta)r \tag{20}$$

A feasible schedule under protocol model ensures that the distance of each line segment is at least $(2+\delta)r$. Summing up all the segments in annulus A_i and using the inequality in (20), we have:

$$\begin{aligned}(2+\delta)ra_i &\leq \sum_{s,t} |\overline{X_s X_t}| \\ &< 2\pi(1+\delta)r(i+1) + (1+\delta)ra_i\end{aligned}$$

$$\Rightarrow a_i < 2\pi(1+\delta)(i+1) \tag{21}$$

Substitute (21) into (19), and sum up the interferences from all annuluses, we have:

$$P_f = \sum_{i=1}^{\infty} P_f(A_i) < \sum_{i=1}^{\infty} \frac{2\pi P_0(1+\delta)(i+1)}{((1+\delta)ri)^\alpha}$$
$$= \frac{2\pi P_0}{r^\alpha(1+\delta)^{\alpha-1}} \sum_{i=1}^{\infty} \left(\frac{1}{i^{\alpha-1}} + \frac{1}{i^\alpha} \right)$$
$$= \frac{2\pi P_0 \left(\zeta(\alpha-1)+\zeta(\alpha)\right)}{r^\alpha(1+\delta)^{\alpha-1}} \quad (22)$$

where $\zeta(\cdot)$ is the Riemann Zeta function. When $\alpha > 2$, $\zeta(\alpha) < \frac{\pi^2}{6}$ and $\zeta(\alpha-1)$ converges to a constant. Denote $c_2 = \zeta(\alpha) + \zeta(\alpha-1)$. Then, when $r = r_0$ and $\delta = \delta_0 > \sqrt[\alpha-1]{\frac{2\pi\beta c_2}{1-\beta r^\alpha N_0/P_0}} - 1$, (22) can be written into:

$$P_f < \frac{P_0}{r_0^\alpha \beta} - N_0 \quad (23)$$

Substitute (18) and (23) into (17), we obtain $SINR_j > \beta$. This proves that a feasible scheduling under protocol model with $r = r_0$ and $\delta = \delta_0$ is also feasible under physical model. Therefore, the network capacity achieved under protocol model when $r = r_0$ and $\delta = \delta_0$ can also be achieved under physical model. □

4.1.3 Capacity Gain over Naive Transmission

COROLLARY 3. *In a wireless sensor network with N uniformly distributed nodes, CDG can achieve a capacity gain of N/M over baseline transmission under both interference models, given that sensor readings are K-sparse, and $M = c_1 K$.*

Denote λ_1 as the network capacity of baseline transmission. It is achieved when every node is allowed to transmit once every n_1 slots, and traffic is evenly distributed among n_2 one-hop neighbors of the sink. Then we have $\frac{W}{n_1} = \frac{N\lambda_1}{n_2}$. Denote λ_2 as the network capacity of CDG. If the same transmission schedule and routing structure are adopted, we have $\frac{W}{n_1} = \frac{M\lambda_2}{n_2}$. From these two equations, we can conclude that CDG can achieve a capacity gain of N/M over baseline transmission.

4.2 NS-2 Simulation

The network capacity analysis is based on scheduled medium access control (MAC). In practice, the computational and communication overhead of MAC scheduling is too high. Contention-based MAC is more often adopted in wireless sensor networks. In order to understand how CDG performs in practical settings, we evaluate its performance through ns-2 [20] simulations and compare it with baseline transmission on two typical topologies: chain [8] and grid topologies [24]. Table 1 lists the main parameters used in the simulation. We adopt 802.11 instead of ZigBee because the implementation of 802.11 in ns-2 is well-established.

For simplicity, we will look into the packet rate instead of bit rate. Each packet contains only one message which is assumed to be 20 bytes for both baseline transmission and CDG. Although both approaches can combine multiple messages in one packet and improve transmission efficiency, we do not use large packets because we are only interested in the comparison between them. Data sparsity is assumed to be 5%. For example, when $N = 1000$, $K = 50$, and we assume that the sink can recover the original data from $M = 200$ random measurements. In the best case, CDG should achieve capacity gain of $N/M = 5$.

Table 1: Simulation parameters

MAC protocol	802.11
Physical data rate	2Mbps
Transmission range	15 meters
Interference range	25 meters
Payload size	20 Bytes
RTS/CTS status	OFF
Retry limit	7
IFQ length	200
K/N (data sparsity)	0.05
$c_1 = M/K$	4

4.2.1 Chain Topology

The chain topology is composed of 1000 sensors and one sink locating at one extreme of the chain. The distance between any two adjacent nodes are 10 meters. Under the given transmission and interference range, nodes can only communicate with adjacent nodes, and may cause interfere to two-hop neighbors.

In the simulation, we vary the input interval and evaluate how output interval and packet loss ratio change accordingly. In general, as the input interval decreases, the output interval decreases and the packet loss ratio increases. However, if an input interval is not achievable, the output interval will cease to decrease and may slightly increase as a result of congestion collapse. We may infer the network capacity from the minimum achieved output interval.

Fig. 8(a) shows that the minimum output interval of baseline transmission is 10.6 seconds per message, and it is achieved when the input interval is 10.2 seconds per message. There is a small gap between these two values because of network jitters and packet losses. Fig. 8(b) shows the performance of CDG. The minimum output interval is 2.11 second per message achieved when the input interval is 1.92 second per message. We can see that CDG can achieve a capacity gain of 5 over baseline transmission.

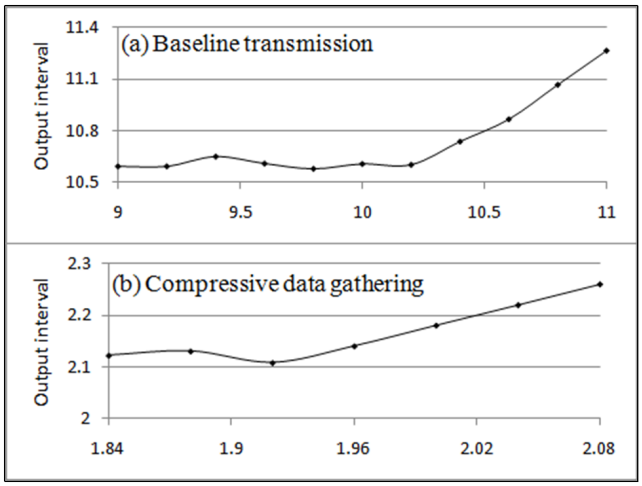


Figure 8: Output-input interval in chain topology

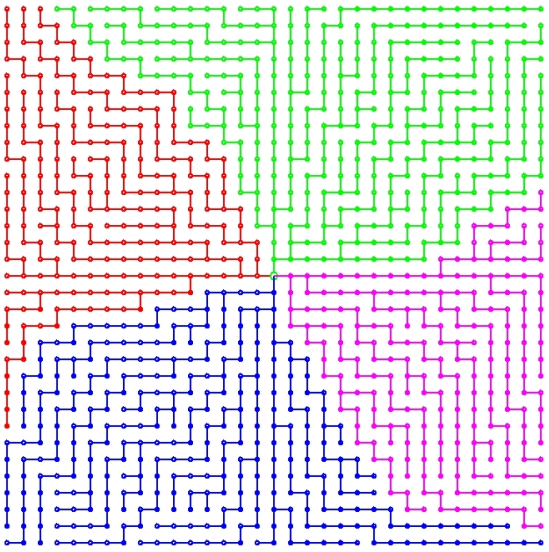


Figure 9: A typical routing tree in grid topology

In addition, the packet loss ratio of CDG is zero when the input interval is 1.92 second per message and above. In contrast, even when the network is not overloaded, baseline transmission incurs a constant packet loss ratio between 3% to 4% as a result of traffic burst.

In this chain topology, CDG introduces an initial delay of 1.80 seconds. This is because transmission starts from the leaf node which is 1000-hop away from the sink. This initial delay does not affect the network capacity because transmitting readings acquired at different time instance can be pipelined.

4.2.2 *Grid Topology*

The grid topology contains 1089 nodes in 33 rows by 33 columns. The distance between adjacent nodes in the same row or column is 14 meters. Therefore, any node not at the border of the network can communicate with four neighbors. Fig. 9 shows a typical tree on the grid topology. The sink is in the middle of the network and four subtrees are represented by four different colors. The subtrees contain similar number of sensor nodes, though not exactly the same. In the simulation, we assume that data from each subtree can be reconstructed from 55 random measurements.

Different from the chain topology where the routing path is deterministic, the grid topology produces changing routing trees in each test run. Therefore, we run three independent tests for each parameter setting and present the average results. In each test run, ten messages per node are collected at given intervals.

Fig. 10(a) shows that baseline transmission achieves the minimum output interval of 5.93 seconds per message when the input interval is 4.7 seconds per message. Fig. 10(b) shows that CDG achieves the minimum output interval of 2.54 seconds per message when the input interval is 2.2 seconds per message. The capacity gain is 2.3 instead of 5. The reason is that in contention based MAC, the transmission slots allocated to each node are not even. Nodes with heavier loads get more time slots to transmit. Therefore, baseline transmission transmits faster than what is assumed in scheduled MAC.

Fig. 11 compares packet loss ratio of the two approaches. Similar to the results in chain topology, CDG achieves near-zero loss ratio when the network is not overloaded. In contrast, the packet loss ratio in baseline transmission is much higher. Even when the input interval is 8 seconds per message, the packet loss ratio is still as high as 20%. This is because in baseline transmission all nodes try to transmit as soon as sensor readings are acquired. In CDG, however, only leaf nodes transmit at the beginning and inner nodes will not transmit until they receive and combine their descendants' readings.

In the grid topology, the initial delay of CDG is neglectable because the tree depth is 32 hops. In our simulations, the average initial delay is less than 0.1 seconds.

5. EXPERIMENTS ON REAL DATA SETS

The previous section has demonstrated the efficiency of CDG under the assumption that data are sparse and can be reconstructed from ideal random measurements. This section will show that sensor data are indeed sparse in reality.

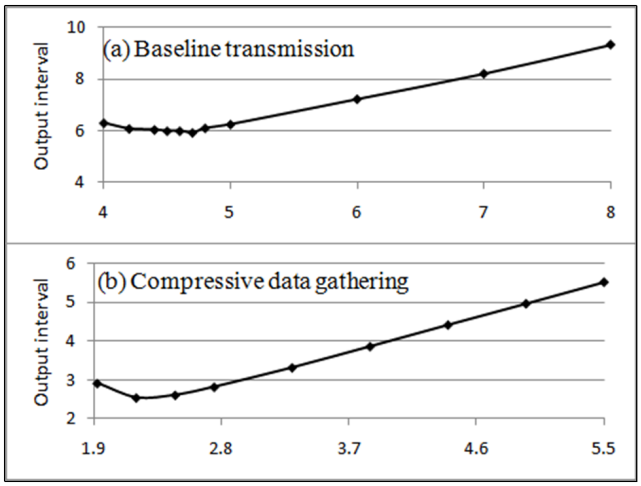


Figure 10: Output-input interval in grid topology

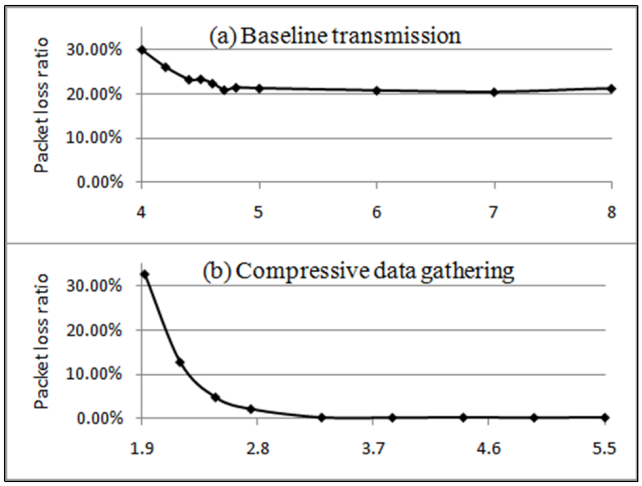


Figure 11: Packet loss ratio in grid topology

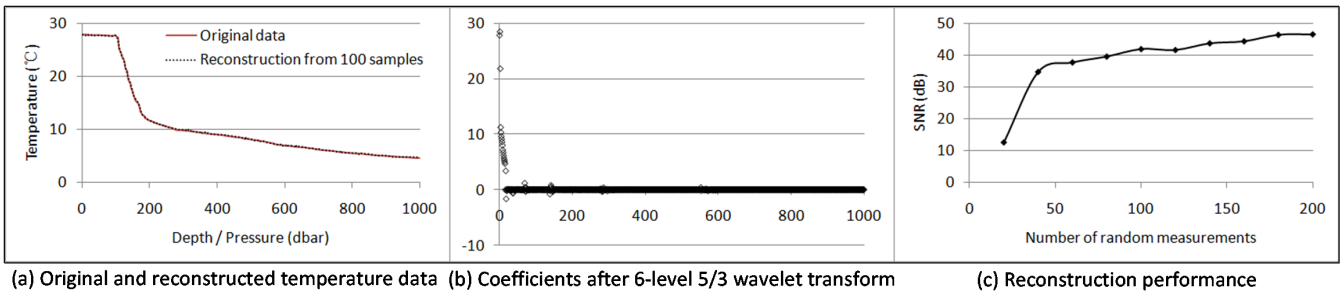


Figure 12: Results on temperature data from the Pacific Sea

Further, data reconstruction is highly robust and efficient although real data are contaminated with noise.

5.1 CTD Data from Ocean

The CTD (Conductivity, Temperature, and Depth) data come from National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center (NDBC). CTD is a shipboard device consisting of many small probes. When collecting data, it is lowered down to the seafloor, and then measures data as it ascends. Although the CTD data are collected by one moving instrument, they demonstrate the same properties as if they were collected by a collection of sensors.

We look into the temperature data collected in the Pacific Sea at (7.0N, 180W) on March 29, 2008 [1]. The data set contains 1000 readings obtained at different depth of sea. We plot the original data by red solid curve in Fig. 12(a). Since the readings are piece-wise smooth, they should be sparse in wavelet domain. Fig. 12(b) shows the 1000 coefficients after 6-level 5/3 wavelet de-correlation. There are only 40 coefficients whose absolute value is larger than 0.2, accounting for only 4.0% of the total coefficients. Although the rest of the coefficients are not strictly zero, we may set $K = 40$.

Compressive sampling theory suggests that data can be reconstructed with high probability from $M = 3K \sim 4K$ random measurements. Fig. 12(c) shows the reconstruction performance with different numbers of random measurements. Each indicated data point is averaged over 10 test runs to avoid fluctuations. Apparently the reconstruction precision increases as M increases. A steep rise is observed in both figures when M becomes greater than K. When $M = K \sim 40$, a reasonable reconstruction SNR of 35dB can be achieved. This translates to a precision over 98%. When $M = 100$ and $M = 200$, the reconstruction precision is 99.2% (41.9dB) and 99.5% (46.5dB). The black dotted curve in Fig. 12(a) shows the reconstructed data when $M = 100$.

5.2 Temperature in Data Center

A contemporary practical application of WSNs is to monitor server temperatures in data centers. The temperature is an indication of server load and abnormal readings in temperature usually sound a note of warning. The sensor data used in this research are collected from a fraction of a data center as shown in Fig. 13. Each rectangular shape represents a rack and the oval shape indicates a sensor placed at the top, middle, and bottom of the rack. As the figure shows, most of the racks are equipped with three sensors while some racks are not monitored and a few others have one or two malfunctioned sensors. There are 498 sensors in total. The data are measured every 30 seconds and transmitted to a sink through baseline scheme. We analyze these data offline to see how much traffic would be reduced if CDG was used. In this network, each node only communicates with adjacent nodes. For simplicity, we assume that all 498 sensors form one subtree to the sink. The energy gain over baseline scheme is similar if sensors form two or more subtrees.

An important observation on this set of data is that sensor readings exhibit little spatial correlations. Although racks are physically close to each other, temperature readings are dominated by server loads instead of ambient temperature. Fig. 14 plots a snapshot of the sensor readings. For clarity, we only show the sensor readings from the bottom of each rack (167 sensors in total) and put the data of each column side by side. Obviously these data are not sparse in any intuitively known domain. We have also checked the entire data set containing sensor readings from all 498 sensors and they are not apparently sparse either. Therefore, conventional compression mechanisms will fail in this situation.

In fact, since the 498 sensors all take values between 10 to 30 degrees centigrade, we have elegantly re-organized d_i into an apparently sparse signal. In particular, we sort d_i in ascending order according to their sensing values at a particular moment t_0. The resulting $\mathbf{d}$ vector is piece-wise smooth and sparse in wavelet domain. Furthermore, since server temperatures do not change violently, sensor readings collected within a relatively short time period can also be regard as piece-wise smooth if organized in the same order. Fig. 15(a) and Fig. 16(a) show the ordered sensor readings 10 minutes and 30 minutes after t_0, respectively. They are

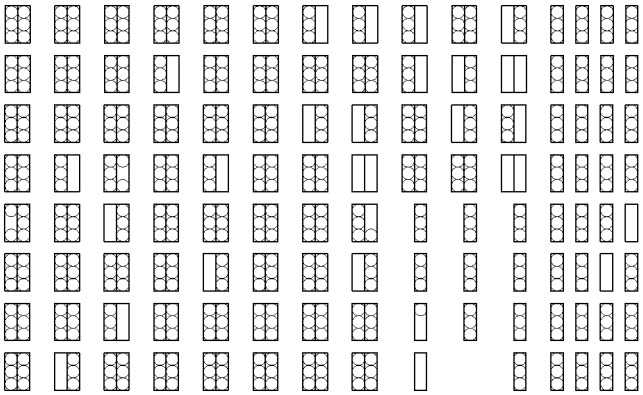


Figure 13: Rack and temperature sensor locations

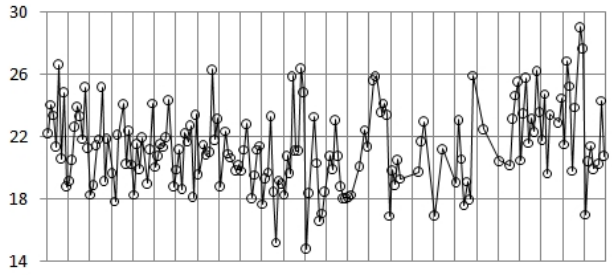


Figure 14: Temperature data of the lowest slot

generally in ascending order with only some small fluctuations. There are also a few significant spikes indicating abnormal temperature readings.

Based on proposed compressive data gathering scheme, we are able to reconstruct such noisy sparse signals with spikes from $M(M < N)$ random measurements. Fig. 15(b)(c) and Fig. 16(b)(c) show the reconstruction results from $M = 0.5N$ and $M = 0.3N$ measurements at two time instances. The average reconstruction precision is over 98%. More importantly, the abnormal readings are accurately captured.

To cope with the situation that temporal correlation becomes weak when the time interval increases, we can refresh the ordering of d_i periodically. In particular, for every one or two hours, the sink requests $M(M = N)$ random measurements in one data gathering process. When $M = N$, the set of equations in (2) is solvable and the sink is able to obtain the exact values of d_i. Then, the sink can re-sort d_i and use this new ordering for data reconstruction in the subsequent hour or two.

We would like to point out that both conventional compression and distributed source coding are unable to exploit this type of sparsity which is observed only at certain reshuffled ordering. In conventional compression, explicit data communication is required between correlated nodes. If correlated nodes are not physically close to each other, the communication between them may take multiple hops. This introduces high overheads and makes compression procedure costly. In distributed source coding, nodes are classified into main nodes and side nodes. The sink allocates appropriate number of bits to each node according to the correlation pattern. However, if the correlation pattern is based on changing sensor ordering, the sink needs to carry out these two tasks and communicate the results to every single node periodically. In contrast, the data gathering process in CDG is unaffected when the ordering of d_i changes. The knowledge of correlation is only used during data reconstruction.

Recall that CDG solves an l_1-minimization problem defined in (7) to reconstruct data. In previous sections, we have discussed how to select the Ψ matrix such that sensor readings $\mathbf{d}$ can be represented by a sparse vector $\mathbf{x}$ in Ψ domain. This section shows how $\mathbf{d}$ can be re-organize to be a sparse signal. This unprecedented flexibility of CDG demonstrates how CDG can achieve a compression ratio of two to three at bottleneck nodes when other conventional mechanisms fail.

6. CONCLUSION AND FUTURE WORK

We have described in this paper a novel scheme for energy efficient data gathering in large scale wireless sensor networks based on compressive sampling theory. We believe this is the first complete design to convert the traditional compress-then-transmit process into a compressive gathering (compress-with-transmission) process to address the two major technical challenges that today's large scale sensor networks are facing. In the development of the proposed scheme, we have carried out the analysis of capacity for wireless sensor network when compressive data gathering is adopted. We have shown that CDG can achieve a capacity gain of N/M over baseline transmission. We have also designed ns-2 simulations to validate the proposed scheme

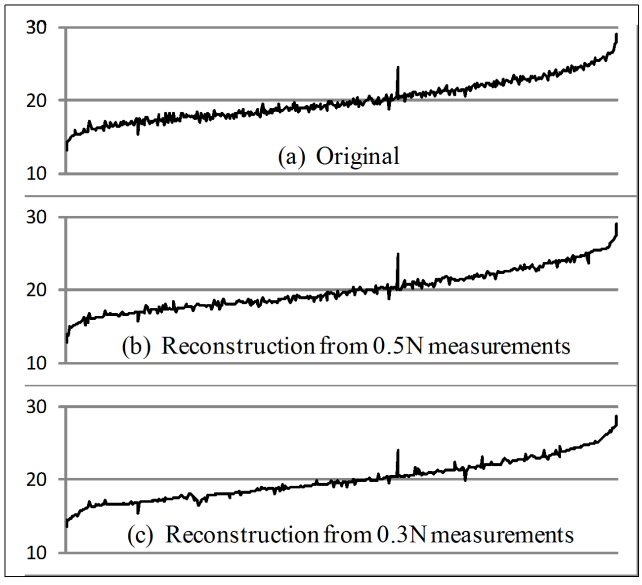


Figure 15: Original and reconstructed sensor readings at $t = t_0 + 10$

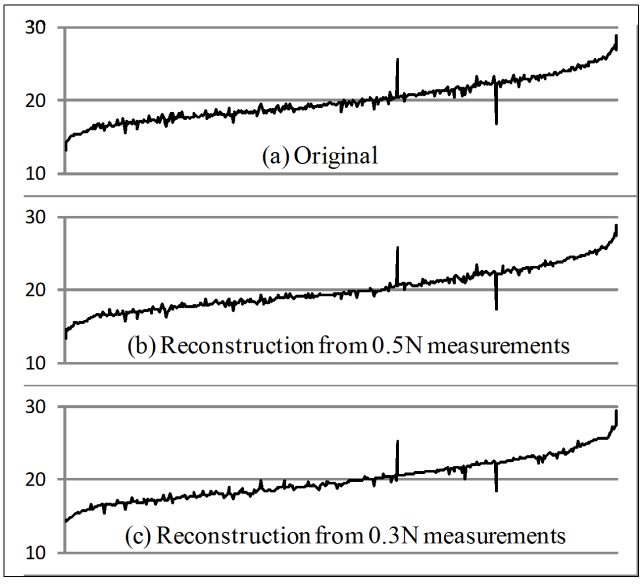


Figure 16: Original and reconstructed sensor readings at $t = t_0 + 30$

when contention-based MAC is used. Furthermore, numerical studies based on real sensor data not only verified data sparsity in practical data acquisition, but also demonstrated the efficiency and robustness of the sensor data reconstruction with and without abnormal readings.

It should be noted that successful application of CDG depends on the properties of sensor field. If sensor readings are not sparse in any known domain and in any proper order, CDG cannot achieve capacity gain because an important prerequisite of compressive sampling theory is missing. At the other extreme, when sensing data are sparse in the original domain, i.e. only a small fraction of sensors acquire non-zero readings, it would be more efficient to directly transmit these non-zero readings through multi-hop forwarding.

CDG is not suitable for small scale sensor networks when signal sparsity may not be prominent enough and the potential capacity gain may be too small. CDG is also more effective for networks with stable routing structure. This is because frequent node failure or dynamic route change will lead to high control overhead that potentially cancel out the gain from data compression. We are currently investigating the extension of CDG to more challenging networking scenarios and the exploitation of fault tolerance of the compressive sampling principles to achieve more robust performance in sensor data gathering.

7. ACKNOWLEDGMENTS

The authors would like to thank Dr. Feng Zhao for providing the data center sensor readings.

8. REFERENCES

[1] *NBDC CTD data*. http://tao.noaa.gov/refreshed/ctd_delivery.php.

[2] J. Aćimović, B. Beferull-Lozano, and R. Cristescu. Adaptive distributed algorithms for power-efficient data gathering in sensor networks. In *Proc. of Intl. Conf. on Wireless Networks, Comm. and Mobile Computing*, pages 946–951, Jun. 2005.

[3] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak. Compressive wireless sensing. In *Proc. of IPSN*, pages 134–142, Apr. 2006.

[4] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118–121, Jul. 2007.

[5] T. Blumensath and M. E. Davies. Gradient pursuits. *IEEE Trans. on Signal Processing*, 56(6):2370–2382, Jun. 2008.

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

[7] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, Mar. 2008.

[8] C. W. Chen and Y. Wang. Chain-type wireless sensor network for monitoring long range infrastructures: architecture and protocols. *International Journal on Distributed Sensor Networks*, 4(4), Oct. 2008.

[9] J. Chou, D. Petrovic, and K. Ramchandran. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. In *Proc. of IEEE Infocom*, pages 1054–1062, Mar. 2003.

[10] 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 *Proc. of IPSN*, pages 309–316, 2006.

[11] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. On network correlated data gathering. In *Proc. of IEEE Infocom*, volume 4, pages 2571–2582, Mar. 2004.

[12] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer. Network correlated data gathering with explicit communication: Np-completeness and algorithms. *IEEE/ACM Trans. on Networking*, 14(1):41–54, Feb. 2006.

[13] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, Apr. 2006.

[14] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, Jan. 2006.

[15] H. Gupta, V. Navda, S. Das, and V. Chowdhary. Efficient gathering of correlated data in sensor network. *ACM TOSN*, 4(1), Jan. 2008.

[16] P. Gupta and P. R. Kumar. The capacity of wireless network. *IEEE Trans. on Inform. Theory*, 46(2):388–404, Mar. 2000.

[17] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak. Compressed sensing for networked data. *IEEE Signal Processing Magazine*, 25(2):92–101, Mar. 2008.

[18] G. Hua and C. W. Chen. Correlated data gathering in wireless sensor networks based on distributed source coding. *Intl. Journal of Sensor Networks*, 4(1/2):13–22, 2008.

[19] D. Marco, E. J. Duarte-Melo, M. Liu, and D. L. Neuhoff. On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data. In *Proc. of IPSN*, pages 1–16, 2003.

[20] S. McCanne and S. Floyd. *Network simulator ns-2*. http://www.isi.edu/nsnam/ns/.

[21] M. Rabbat, J. Haupt, A. Singh, and R. Nowak. Decentralized compression and predistribution via random gossiping. In *IPSN*, pages 51–59, Apr. 2006.

[22] D. Slepian and J. K. Wolf. Noiseless encoding of correlated information sources. 19:471–480, Jul. 1973.

[23] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler. An analysis of a large scale habitat monitoring application. In *Proc. of ACM SenSys*, pages 214–226, Nov. 2004.

[24] S. Tilak, N. B. Abu-Gahazaleh, and W. Heinzelman. Infrastructure tradeoffs for sensor networks. In *Proc. of ACM IWWSNA*, pages 49–58, 2002.

[25] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on IT*, 53(12):4655–4666, Dec. 2007.

[26] S. Yoon and C. Shahabi. The clustered aggregation (cag) technique leveraging spatial and temporal correlations in wireless sensor networks. *ACM Trans. on Sensor Networks*, 3(1), Mar. 2007.

[27] K. Yuen, B. Liang, and B. Li. A distributed framework for correlated data gathering in sensor networks. *IEEE Trans. on Vehicular Technology*, 57(1):578–593, Jan. 2008.