CDC: Compressive Data Collection for Wireless Sensor Networks

Xiao-Yang Liu, Yanmin Zhu, *Member, IEEE*, Linghe Kong, *Member, IEEE*, Cong Liu, Yu Gu, *Member, IEEE*, Athanasios V. Vasilakos, *Senior Member, IEEE*, and Min-You Wu, *Senior Member, IEEE*

Abstract—Data collection is a crucial operation in wireless sensor networks. The design of data collection schemes is challenging due to the limited energy supply and the hot spot problem. Leveraging empirical observations that sensory data possess strong spatiotemporal compressibility, this paper proposes a novel compressive data collection scheme for wireless sensor networks. We adopt a power-law decaying data model verified by real data sets and then propose a random projection-based estimation algorithm for this data model. Our scheme requires fewer compressed measurements, thus greatly reduces the energy consumption. It allows simple routing strategy without much computation and control overheads, which leads to strong robustness in practical applications. Analytically, we prove that it achieves the optimal estimation error bound. Evaluations on real data sets (from the GreenOrbs, IntelLab and NBDC-CTD projects) show that compared with existing approaches, this new scheme prolongs the network lifetime by $1.5\times$ to $2\times$ for estimation error 5-20 percent.

Index Terms—Compressive data collection, wireless sensor networks, compressive sensing, random compression, nonuniform random projection

1 Introduction

Wireless sensor networks (WSNs) are adopted in many military, civilian and commercial applications recent years [1]. A crucial operation of WSNs [2] is to perform data collection, where sensor readings are collected from sensor nodes and then transmitted to the sink through multi-hop wireless communications. Various applications rely on efficient data collection, such as battlefield surveillance [3], [4], [5], habit monitoring [6], infrastructure monitoring [7], and environmental monitoring [8].

A primary challenge of designing data collection schemes lies in prolonging the network lifetime. First, each sensor node, being a micro-electronic device, can only be equipped with a limited power source while in many applications, recharging is impractical (impossible or not worth it). Thus, a WSN can only support limited volume of traffic load. Second, the information that a WSN can effectively transport is even less under multi-hop transmissions since the network capacity decreases as the number of nodes increases [9], i.e., the multi-hop scheme requires lots of packet forwardings. Third, the many-to-one traffic pattern, called *convergecast* [10], of data collection induces load unbalance. It leads to the *hot spot* problem [11], i.e., the sensor nodes closer to the sink will run out of energy sooner. Therefore, the network lifetime of WSNs will be significantly shortened.

Furthermore, designers have to deal with the following constraints: the unreliability of low-power wireless communication, and the limited computational ability of sensor nodes. Low-power transceivers induce poor link quality, therefore packet loss occurs frequently [12], [13]. Actually, real-world WSN projects suffer from serious data loss with loss rates as high as 23-64 percent [12]. To ensure reliable transmission will cost unconscionable amount of energy as it induces an exceptionally huge number of retransmissions, which is not cost-effective. On the other hand, sensor nodes can only support simple computing tasks, therefore the preprocessing or compression of data collection schemes should be easy to implement.

Existing solutions have limitations and thus are unsatisfactory. Generally, data collection in WSNs follows two approaches: raw-data collection and aggregated-data collection. WSNs are typically composed of hundreds to thousands of sensor nodes generating tremendous amount of sensory readings, as the packet loss problem and the hot spot problem surface, raw-data collection is rather inefficient or problematic [11], [13]. This approach will lead to a large number of retransmissions in real-world situations and node failures (cluster headers, or tree node, etc.) as batteries run out. Aggregated-data collection takes advantage of spatiotemporal correlations (or compressibility) within sensory data to reduce communication costs. More specifically, in-network data compression [14] is adopted to reduce global traffic, such as distributed source coding [15] or transform coding [16]. However, they incur significant computation and control overheads, i.e., the former one relies on communications between neighbor nodes, while

- X.-Y. Liu, Y. Zhu, L. Kong, and M.-Y. Wu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. E-mail: {yanglet, yzhu, linghe.kong, mwu}@sjtu.edu.cn.
- C. Liu is with the Department of Computer Science, University of Texas at Dallas, Richardson, TX. E-mail: cong@utdallas.edu.
- Y. Gu is with the Department of ISTD, Singapore University of Technology and Design, Singapore. E-mail: jasongu@sutd.edu.sg.
- A.V. Vasilakos is with the Department of Computer Engineering, University of Western Macedonia, Psychiko, Attica 15410, Greece. E-mail: vasilako@ath.forthnet.gr.

Manuscript received 10 Feb. 2014; revised 4 July 2014; accepted 22 July 2014. Date of publication 31 July 2014; date of current version 6 July 2015.
Recommended for acceptance by H. Jin.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2014.2345257

the later one requires complicated transforming computation, which are not suitable for WSNs.

Compressive sensing theory (CS) [17], [18] exhibits successful application in designing effective data collections [11], [19]. The authors in [11] propose a compressive data gathering (CDG) scheme which reduces global-scale communication costs while achieving tempting load balancing. However, it assumes a perfectly reliable routing tree, thus is vulnerable to packet loss or node malfunctioning. Maintaining such routing infrastructure will cause large control overheads. Although in simulations it behaves well under careful controls, the practical performance is unsatisfactory.

Main contributions. First, we propose a novel compressive data collection scheme for wireless sensor networks. Our scheme compresses sensory readings "on the fly" under an opportunistic routing.

Second, we model the data collection process as a nonuniform sparse random projection (*NSRP*), we propose a NSRP-based estimator which guarantees optimal error bound.

Finally, based on real data sets, we show that our scheme prolongs the network lifetime by $1.5\times$ to $2\times$ for estimation error 5-20 percent, compared with the baseline scheme and the CDG [11] scheme.

2 Related Work

Energy conservation [20] is an important issue in wireless sensor networks. In-network compression is a promising approach to reduce the amount of information to be transmitted by exploiting sensory data's redundancy. We classified existing data collection schemes into three categories: conventional compression, distributed source coding, and compressive sensing.

Conventional compression. Conventional compression techniques assume specific data structures and thus require communication among sensor nodes [20]. In joint entropy coding approach, nodes use relayed data as side information to encode their readings. If the data are allowed to be communicated back and forth during encoding, sensor nodes may cooperatively perform transforms to better utilize the correlation, such as the gossip-based technique used in [19]. There are two main problems with this approach. First, the route heavily influences the compression performance [14]. To achieve high compression ratio, data compression and packet routing are required to be optimized jointly, which is proved to be NP-hard [21], [22]. Second, structure-aware data compression induces computational and communication overheads [14], [23], [24], rendering this kind of data collection schemes to be inefficient.

Distributed source coding. Distributed source coding intends to reduce complexity at sensor nodes and to utilize correlation at the sink [15]. After encoding sensor readings independently, each node simply sends the compressed message along the shortest path to the sink [8]. Distributed source coding performs well for static correlation patterns. However, when correlation pattern changes or anomaly readings show up, the estimation accuracy will be greatly affected.

Compressive sensing. Recently, compressive sensing gains increasing attention in wireless sensor networks [11], [16], [19], In both static and mobile sensor networks [23], the interplay of routing with compressive sensing is a key issue [25]. Some of them conclude that although sparsity exists in the environment, the restricted isometry property (RIP) is required by traditional compressive sensing decoder, and hardly good approximations can be achieved. Then some proposed network-layer compression [26] to avoid this kind of problem. Our scheme adopts opportunistic routing with quite simple compression, therefore the data collection process is dynamic. This dynamic feature leads to energy balancing and finally benefits energy consumption.

Note that there are several solvers that are related with our work. Actually, all compressive sensing solvers are designed for random projections. Few of them use sparse random projections [26], [27]. Our work are motivated by [26], however, that solver cannot be applied directly for wireless sensor networks since uniform sampling is hard to achieve and thus will require complicated control overhead. Belief propagation [27] exploits sparse encoding, however it also works for uniform sampling while our estimator is the first solver that can deal with nonuniform sampling. The probability distribution is also used in their decoding process, which is treated as prior information while we use the sampling distribution in the generation of a projection matrix for decoding.

3 System Model and Design Overview

3.1 Network Model

We consider a wireless sensor network consisting of n sensor nodes and a sink. Sensor nodes are distributed in the target field to sense the physical conditions and then report sensory readings back to the sink through multi-hop transmissions.

Since wireless sensor networks use low-power transceivers, the link quality is bad, as revealed in [12] and [13]. We assume that the wireless channel is lossy. It is the motivation of adopting the opportunistic routing in Section 4.1.

The monitoring period is evenly divided into T time slots, denoted as $\{0, 1, \ldots, t, \ldots, T-1\}$. A record at the sensor node includes sensor reading, node ID, position (longitude and latitude), and time stamp. The format of a record is:

Let $U_{i,t}$ denote the sensor reading of the ith node at slot t, which may be temperature, humidity, illumination, etc. Then, a physical condition, say temperature, can be represented as a data matrix:

$$\mathbf{U} = \begin{bmatrix} U_{0,0} & \ldots & U_{0,t} & \ldots & U_{0,T-1} \\ U_{1,0} & \ldots & U_{1,t} & \ldots & U_{1,T-1} \\ \ldots & \ldots & \ldots & \ldots & \\ U_{n-1,0} & \ldots & U_{n-1,t} & \ldots & U_{n-1,T-1} \end{bmatrix}, \tag{1}$$

where the ith row is the ith node's reading sequence, and the tth column is the whole network's readings at slot t.

Notations. U refers to either a matrix or a vector, depending on the context; throughout the paper, $N = n \times T$; $U^{\mathcal{T}}$ is the transpose of U; $|\theta|$ denotes the magnitude of coefficient θ; $\|U\|_2$ is the ℓ_2 normal of U; $\mathbf{\Psi}$ denotes the transform basis. A packet at the sink is called a measurement throughout the paper.

3.2 Data Model

We consider a data vector $U \in \mathbb{R}^{N\times 1}$, and an orthonormal transform $\mathbf{\Psi} = [\Psi_1, \Psi_2, \ldots, \Psi_N] \in \mathbb{R}^{N\times N}$. $\mathbf{\Psi}$ can be a wavelet or a Fourier transform basis [28]. The coefficients vector $\theta = [\Psi_1 U, \Psi_2 U, \ldots, \Psi_N U]^{\mathcal{T}}$ can be reordered decreasingly in terms of magnitude, such that $|\theta|_{(1)} \geq |\theta|_{(2)} \geq \cdots \geq |\theta|_{(N)}$.

Power-law decaying data model. The coefficients' magnitude decays according to the power law [29], i.e., the ith largest coefficient satisfies

$$|\theta|_{(i)} \leq C i^{-1/\varpi}, \quad i = 1, 2, \ldots, n, \tag{2}$$

where C is a constant and $-1/\varpi$ controls the compressibility of the data.

Optimal error bound. The optimal approximation for the power-law decaying data [29] is keeping the largest K coefficients and setting the others as zero. The optimal estimation error bound is

$$\|U - \hat{U}_{opt}\|_2^2 = \|\theta - \hat{\theta}_{opt}\|_2^2 = \eta_\varpi C K^{-1/\varpi + 1/2}, \tag{3}$$

where η_ϖ is a constant that only depends on $-1/\varpi$.

We verify this data model based on the data set listed in Table 1. Please refer to our online supplementary file, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2014.2345257. Note that our estimator in Section 4.5 is robust to noise, which is an inherent capability from the above data model. However, our model cannot deal with anomaly readings (sensor faults and outliers).

3.3 Design Overview

The framework of our scheme is shown in Fig. 1. It has two major components: an opportunistic routing and an estimator. The opportunistic routing is responsible for data compression and packet relaying. By modeling it as a Markov chain, the compression probability of each node can be calculated. Then, we prove that nonuniform sparse random projection preserves inner product of two vectors and apply this property to design a simple but quite accurate estimator.

At the beginning, data U ($\theta = U^{\mathcal{T}} V$) is stored locally in sensor nodes. Each sensor node generates a TOKEN with probability p. Therefore, initially there are $L = np$ (in expectation) TOKENs across the network.

The compressive data collection scheme works in the following way:

- Each node with a TOKEN generates one packet destinationed to the sink.
- The packets are relayed under an opportunistic routing. Compression is performed at each newly encountered node.
- At the sink, the compression process is modelled as $Y_{L\times 1} = W_{L\times N} U_{N\times 1}$. Projecting the basis V by M, we have V'. Then, $\hat{\theta} = Y^{\mathcal{T}} V'$ is an estimation of θ.
- Finally, $\hat{U} = (\hat{\theta} V^{-1})^{\mathcal{T}}$, where V^{-1} is the inverse of V.

Key issues.

- How to determine p? Since $p = L/n$ while n is known, we will show how to determine L.

record:	reading	ID	position	time stamp

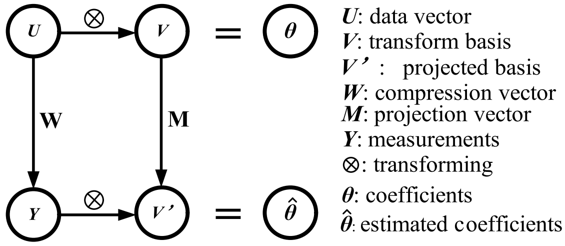


Fig. 1. The framework of CDC.

- The classic result of linear algebra says that $\hat{\theta} = Y^{\mathcal{T}} V'$ is solvable only when $L \geq N$. To make the data collection efficient, we are faced with the question: how to exploit the compressibility of sensory data by designing an estimation algorithm for the case $L \ll N$. Compressive sensing theory [17], [18] points out that this is possible.

In Section 4, we first describe how our data collection scheme works and then model it mathematically. Finally, an estimator and the corresponding results for L are presented in Section 4.5.

4 Design

4.1 Opportunistic Routing with Compression

The opportunistic routing has two tasks: packet forwarding and data compression. We first describe a *data collection path*, and then the compression process along this path.

Packet forwarding. For node s_i, we define a *nearer-to-sink neighbor set* as the one-hop neighbors that are closer to the sink than itself, i.e., $\mathcal{N}(i) = \{j | d(j, sink) \leq d(i, sink) \& d(i,j) \leq R_c\}$ where R_c is the communication range. When a packet arrives at node s_i, s_i compresses its sensory reading into the packet and then sends it out according to the opportunistic routing [30], [31], [32], i.e., forwarding the packet to a randomly selected one of its nearer-to-sink neighbors $s_j \in \mathcal{N}(i)$. In this way, each packet is guaranteed to be successfully delivered to the sink.

Data collection path. The trajectory of the lth packet from a source node to the sink is called a *data collection path*, denoted as $P_l = \langle p_0, p_1, \ldots, p_{\rho_l} \rangle$ with $p_{\rho_l} = sink$, i.e., the packet travels across ρ_l sensor nodes before it reaches the sink.

Since opportunistic routing is adopted, the data collection paths are dynamic and random. It turns out to bring about two good features: energy balancing and security. These non-deterministic data collection paths will balance energy consumption among nodes, as well as preventing possible attacks.

Data compression. As the packet travels towards the sink, the compression scheme adds or substracts the sensory reading of a newly encountered node, as shown in Fig. 2a. The format of the data packet is

Packet:	value	ID list	coefficient list	time stamp list

Let u_i ($i = 0, 1, \ldots, \rho_l - 1$) be the reading of the ith node along path P_l. The data compression is performed as following:

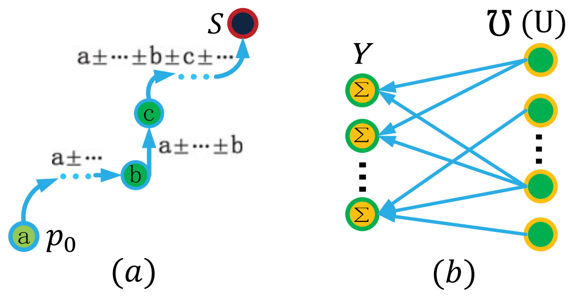


Fig. 2. (a) Along P_l, the packet adds or subtracts the sensor reading of a newly encountered node. (b) The left process is modelled as: A sampling process ʊ randomly selects a subset of sensor readings, and a compression process sums them up with random coefficients chosen from the set $\{-1, +1\}$ to get one measurement.

Step 1. A node with the lth TOKEN becomes node p_0 of P_l. It generates a packet containing data $y_0 = \pm u_0$, then transmits the packet to one of its nearer-to-sink neighbors according to the opportunistic routing.

Step 2. The packet arrives at sensor s_i, who adds or substracts its own sensor ready with probability $1/2$ as

$$y_i = y_{i-1} + r_i u_i. \tag{4}$$

Sensor ID, the coefficient r_i and the current time slot is added to the packet's header. Then s_i transmits it to one of its neighbors closer to the sink according to the opportunistic routing.

Step 3. The encoding process continues along path P_l until the packet reaches the sink.

From the proof of Lemma 1 (in Section 4.2), we know that the random coefficients r_i can be real values in $[-1, 1]$ or chosen from $\{-1, +1\}$. We use the set $\{-1, +1\}$ because the nodes will only need to perform addition or substraction operations.

In the end of the data compression process, $L = np$ packets are collected by the sink. Next, we consider the sink's strategy to estimate the sensor readings.

4.2 Problem Formulation for Estimation

Traditional compressive sensing approach. In traditional compressive sensing approach [11], [23], [24], the sink establishes the following equation:

$$Y_{L\times 1} = A_{L\times N} U_{N\times 1}, \tag{5}$$

where A is an matrix with elements corresponding to r_i in Eqn. (4). The sink can extract A from the collected packets' headers.

Assuming that data U is *sparse*. To be exact, there exists a transform basis $V_{N\times N}$ under which $U_{N\times 1}$ can be represented using $K \ll n$ nonzero coefficients, i.e., $\theta = U^{\mathcal{T}} V$ with $\|\theta\|_0 = K$. Compressive sensing theory [17], [18] claims that, with probability at least $1 - n^{-\gamma}$ (γ is set to be large), U can be reconstructed exactly as the solution to the following ℓ_1-minimization problem, i.e., $U = (\hat{\theta} V^{-1})^{\mathcal{T}}$, where

$$\begin{aligned} &\hat{\theta} = \min_{\theta} \sum_{i=1}^{N} |\theta_i|, \quad s.t. \quad Y_{L\times 1} = A_{L\times N} U_{N\times 1} \\ &L = O(K\mu^2_{(A,V)} logN/K), \quad \mu_{(A,V)} = \max_{1\le i,j\le N} |A_i^{\mathcal{T}} V_j|. \end{aligned} \tag{6}$$

Existing approaches suggest to densely compress WSNs' data to get random measurements as in [11], [19]. Compressive sensing allows one to use convex optimization to estimate sensory readings under the condition of the *RIP* property [17], [18], i.e., to decouple the matrix A and the basis V to have small value of μ. However, the *RIP* property does not hold for opportunistic routings, which postpones its utilization [24], [25].

Problem formulation. We take another approach by regarding the compression process as *nonuniform sparse random projections* (shown in Fig. 2b), modeled as two mutually independent processes *nonuniform sampling process* $\mho(\cdot)$ and *linear encoding* A as

$$\mho(U_j) = \begin{cases} U_j, & prob.\ \pi_j \\ 0, & prob.\ 1-\pi_j \end{cases}, \quad A_{ij} = \begin{cases} 1, & prob.\ \frac{1}{2} \\ -1, & prob.\ \frac{1}{2} \end{cases}, \tag{7}$$

where $\pi_j \neq 0, j \in 1, 2, \ldots, N$ corresponds to the chance of U_j being compressed in the collected packets. Thus, we model the compression scheme as

$$Y_{L\times 1} = A_{L\times N} \mho(U_{N\times 1}). \tag{8}$$

Our problem becomes:

$$\begin{aligned} &\hat{\theta} = \min_{\theta} \|U - \hat{U}\|_2^2, \\ &s.t.\ Y_{L\times 1} = A_{L\times N} \mho(U_{N\times 1}), \end{aligned} \tag{9}$$

where $\hat{U} = (\hat{\theta} V^{-1})^{\mathcal{T}}$.

The opportunistic routing provides load balancing in the cost of nonuniform compression probability of each node [24], [25]. This leads to the failure of the traditional compressive sensing approach. In the following section, we provide a method to calculate the compression probability for each node. Then in Section 4.5, we introduce a new estimation algorithm by using the information of compression probability of each node.

4.3 Compression Probability Estimation

Modeling the opportunistic routing as a Markov chain. The packet forwarding process can be modeled as a Markov Chain: the states are the nodes, and the forwarding probability constitutes the transition probability matrix (each entry specifies the probability that a packet is transmitted from one node to one of its nearer-to-sink neighbors).

First, we estimate the transition matrix P based on the "incomplete observation" version of maximum likelihood estimation (MLE). Once the transition matrix P is known, the compression probability distribution π can be derived.

The incomplete observation problem. For each pair of nodes, assume that there is a *transition link* with two states: *ON* and *OFF*. If we have complete observations of all links' states in each time slot, then *the estimation of the transition matrix is to maximize the log-likelihood of the posterior probability*, i.e., $logP((P_1, P_2, \ldots P_L) \mid P)$ with $(P_1, P_2, \ldots P_L)$ denotes the data collection paths of the L collected packets. Maximum likelihood estimation is the most used routine to solve this problem [33].

However, the observation of a transition link's state is to try a "packet-transmitting test," which is contradict to our goal as the data collection aims to minimize the number of packets transmitted. Therefore, with "incomplete" or undersampled observations, the traditional MLE scheme can not be used here.

MLE with incomplete observation. Here, we adopt an "incomplete observation" version of MLE [34]. Let O_{ijt} denote the number of observed transition from node s_i to node s_j occurring over t time slots and $(P^t)_{ij}$ the ijth element of the matrix P^t (the probability of a packet in node s_i arrives at node s_j after t time slots), this new MLE is defined as

$$\begin{aligned}\hat{P} &= \max_P \ log\, P((P_1, P_2, \ldots P_L) \mid P) \\ &= \max_P \sum_i \sum_j \sum_t O_{ijt}\ log\,(P^t)_{ij}.\end{aligned} \tag{10}$$

It is proved in [34] that the expectation maximization algorithm converges to the global maximum. Which means that with high probability, the above maximizer can produce an accurate transition matrix.

Estimation of the compression probability π. The compression probability closely relates with the Markov chain-like occurrence, except that we should only count one time if the packet stays at a node waiting for transmission. The estimation algorithm is described as following:

Step 1. Initially, $\pi_0 = \{L/n, \ldots, L/n\}$.

Step 2. Get $\bar{P}$ by setting P's diagonal elements to 0.

Step 3. Calculate the nodes's occurrence frequency in data collection paths during T time slots:

$$O_i(T) = \sum_{i=0}^{T-1} \pi_0 \bar{P}^i. \tag{11}$$

Step 4. Average the occurrence frequency by $\sum \pi_0$, we get the probability distribution

$$\pi_i = \frac{O_i(T)}{\sum \pi_0} = \frac{\sum_{i=0}^{T-1} \pi_0 \bar{P}^i}{\sum \pi_0}, \tag{12}$$

where the sum and division are element-wise operations on row vectors.

4.4 Nonuniform Sparse Random Projection

Eq. (8) equals to the following linear equations:

$$Y_{L\times 1} = W_{L\times N} U_{N\times 1}\,, \tag{13}$$

$$W_{ij} = \begin{cases} +1, & prob.\ \ \frac{1}{2}\pi_j \\ 0, & prob.\ \ 1-\pi_j\,. \\ -1, & prob.\ \ \frac{1}{2}\pi_j \end{cases} \tag{14}$$

Sparse and *nonuniform* raise from the fact that the opportunistic routing will neither pass all sensor nodes nor pass them with equal probability, which is the case for most existing routings. *Sparse* allows the collected packets to compress sensor readings from a random selected subset, while traditional compressive sensing approaches require to compress all sensory readings together, or the subset are randomly selected with equal probability [11], [19].

Correspondingly, we construct a projection matrix $M \in \mathbb{R}^{L\times N}$ (where $L \ll N$) containing entries

$$M_{ij} = \frac{1}{\pi_j}\begin{cases} +1, & if\ W_{ij} = +1 \\ 0, & if\ W_{ij} = 0 \\ -1, & if\ W_{ij} = -1 \end{cases}. \tag{15}$$

The entries within each row are mutual-independent, while the entries across different rows are fully independent. In expectation, each row contains $\sum_{j=1}^n \pi_j$ nonzero elements, i.e., there are on average $\sum_{j=1}^n \pi_j$ sensor readings encoded in one collected packet.

Next, we prove in Lemma 1 that with high probability, nonuniform sparse random projections preserve inner products with predictable error. Therefore, using only their random projections, we are able to estimate the inner product of two vectors. The detailed proof is in the next section

Lemma 1. *For any vectors* $U, V \in \mathbb{R}^{N\times 1}$, *and* $W, M \in \mathbb{R}^{L\times N}$ *in Eqs. (14) and (15). The random projections* $Y = \frac{1}{\sqrt{L}} WU$, $V' = \frac{1}{\sqrt{L}} MV$, *with the expectation and variance satisfying:*

$$E[Y^T V'] = U^T V, \tag{16}$$

$$\begin{aligned} &Var(Y^T V') \\ &\le \frac{1}{L}\left((U^T V)^2 + \xi \|U\|_2^2 \|V\|_2^2 + (\kappa - 2 - \xi)\sum_{j=1}^N U_j^2 V_j^2 \right), \end{aligned} \tag{17}$$

where $\xi = max(\frac{\pi_l}{\pi_m})(l, m \in \{1, 2, \ldots n\}), \kappa = \frac{1}{min(\pi)}$ *denote the degree of* nonuniform *and the expected times to sample the "rarest" node, respectively.*

4.5 NSRP-Based Estimator

The intuition for our estimator design is that *nonuniform sparse random projections* preserve inner products within a small error. Hence we can use random linear measurements $Y = WU$ of the original data, and random linear projections $V' = MV$ of the orthonormal basis, to estimate the coefficients vector θ, as in the following way:

Step 1. Extract packets' headers at the sink, get $W_{L\times N}$, $Y_{L\times 1}$, and the projection matrix M.

Step 2. Set $L_1 = C_1 \frac{1+\xi+\kappa H^2}{\epsilon^2}$ and $L_2 = C_2(1+\gamma)\log N$ such that $L = L_1 L_2$.

Step 3. Partition $Y_{L\times 1}$ into L_2 column vectors $\{Y_1, Y_2, \ldots, Y_{L_2}\}$ with each of size $L_1 \times 1$, partition M into $\{M_1, M_2, \ldots, M_{L_2}\}$ with each of size $L_1 \times N$, then project the basis V to get $\{V_1' = \frac{1}{\sqrt{L_1}} M_1 V, \ldots, V_{L_2}' = \frac{1}{\sqrt{L_1}} M_{L_2} V\}$.

Step 4. Compute $\zeta_l = Y_l^T V_l', l = 1, 2, \ldots, L_2$. Set each element of $\hat{\theta}$ as the median value of each column vector $\zeta_1, \ldots, \zeta_{L_2}$.

Step 5. Keep the K largest coefficient in $\hat{\theta}$ and set the remaining to zero.

Step 6. Return $\hat{U} = (\hat{\theta} V^{-1})^T$.

The following two theorems hold for the above estimator. Please refer to the next section for the detailed proof for Theorem 1. Since the proof of Theorem 2 is similar to that in [26], for completeness, we include it in the supplementary file, available online.

Theorem 1. *For data vector* $U \in \mathbb{R}^{N\times 1}$ *satisfying*

$$\frac{\|U_\infty\|}{\|U\|_2} \le H. \tag{18}$$

Let $V = \{V_1, \ldots, V_N\}$ be the transform basis with each vector $\mathbb{R}^{N\times 1}$, $W, M \in \mathbb{R}^{L\times N}$ as in Eqs. (14) and (15) with the compression probability π,

$$L = \begin{cases} O\Big(\frac{1+\gamma}{\epsilon^2}(\xi+\kappa H^2)\, log\, n\Big), & if\ (\xi+\kappa H^2) > \Omega(1), \\ O\Big(\frac{1+\gamma}{\epsilon^2}\, log\, n\Big), & if\ (\xi+\kappa H^2) \le O(1), \end{cases} \tag{19}$$

Then, with probability at least $1 - N^{-\gamma}$, the random projections $Y = \frac{1}{\sqrt{L}}WU$ and $V' = \frac{1}{\sqrt{L}}MV_i$ can produce an estimate $\hat{\theta}_i$ for $U^T V_i$ (Step 4 in the above estimator), satisfying

$$|\hat{\theta}_i - U^T V_i| \le \epsilon \|U\|_2^2 \|V_i\|_2^2, \tag{20}$$

for all $i = 1, 2, \ldots, N$.

Theorem 2. *Suppose data $U \in \mathbb{R}^{N\times 1}$ satisfies condition (18), $W, M \in \mathbb{R}^{L\times N}$ in Eqs. (14) and (15) with probability distribution π, and*

$$L = \begin{cases} O\Big(\frac{1+\gamma}{\epsilon^2\eta^2}(\xi+\kappa H^2)K^2\, log\, n\Big), & if\ \ (\xi+\kappa H^2) \ge \Omega(1), \\ O\Big(\frac{1+\gamma}{\epsilon^2\eta^2}K^2\, log\, n\Big), & if\ \ (\xi+\kappa H^2) \le O(1). \end{cases} \tag{21}$$

Let $Y = \frac{1}{\sqrt{L}}WU$, consider an orthonormal transform $\mathbf{\Psi} \in \mathbb{R}^{L\times N}$ and the corresponding transform coefficients $\theta = \mathbf{\Psi}U$. If the K largest transform coefficients in magnitude give an approximation with error $\|U - \hat{U}_{opt}\| \le \eta\|U\|_2^2$, then given only Y, W, M, and $\mathbf{\Psi}$, the above NSRP-based estimator produces an estimate $\hat{U}$ with error

$$\|U - \hat{U}\| \le (1+\epsilon)\eta\|U\|_2^2, \tag{22}$$

with probability at lest $1 - N^{-\gamma}$.

For the above estimator, Theorem 1 states that with high probability, the nonuniform sparse random projections of data vector and any projected basis vector can produce estimates of their inner products within a small error. Thus Theorem 1 guarantees small estimation error for each coefficient. Theorem 2 shows that with high probability, nonuniform sparse random projections can approximate compressible data with error being comparable to the optimal error bound (by setting ϵ to be small, i. e, $\epsilon = o(1)$.). Thus, with high probability, the above estimator produces an estimation of the original data within small error.

Note that compared with uniform sampling [26], nonunifrom sampling requires bigger L. For the extra $(\xi + \kappa H^2)$ component, ξ can be regarded an indicator of the degree of nonuniform, and κ is a factor to guarantee successful decoding for the "rarest" node.

5 Proofs

5.1 Lemma 1

Proof. The pair of nonuniform sparse random projections $W, M \in \mathbb{R}^{L\times n}$ satisfies:

$$E[W_{ij}] = 0, E[M_{ij}] = 0; E[W_{ij}M_{ij}] = 1,$$
$$E[W_{il}M_{im}] = E[W_{il}]E[M_{im}] = 0 \quad if\ l \ne m,$$
$$E[W_{il}M_{il}W_{im}M_{im}] = E[W_{il}M_{il}]E[W_{im}M_{im}] = 1 \quad if\ l \ne m,$$
$$E\big[W_{ij}^2\big] = \pi_j, E\big[M_{ij}^2\big] = \frac{1}{\pi_j}, E\big[W_{ij}^2M_{ij}^2\big] = \frac{1}{\pi_j},$$
$$E\big[W_{il}^2M_{im}^2\big] = E\big[W_{il}^2\big]E\big[M_{im}^2\big] = \frac{\pi_l}{\pi_m} \quad if\ l \ne m.$$

The above results will be used in the following process without explicit mention.

Define the random variables

$$\omega_i = \left(\sum_{j=1}^n u_j W_{ij}\right)\left(\sum_{j=1}^n v_j M_{ij}\right)$$
$$\zeta = \alpha^T\beta = \frac{1}{L}\sum_{i=1}^L \omega_i$$

so that $\omega_1, \omega_2, \ldots, \omega_M$ are independent.

$$\begin{aligned} E(\omega_i) &= E\left[\sum_{j=1}^n u_j v_j W_{ij}M_{ij} + \sum_{l\ne m} u_l v_m W_{il}M_{im}\right] \\ &= \sum_{j=1}^n u_j v_j E[W_{ij}M_{ij}] + \sum_{l\ne m} u_l v_m E[W_{il}M_{im}] \\ &= \mathbf{u}^T\mathbf{v} \\ E[\zeta] &= \mathbf{u}^T\mathbf{v}. \end{aligned}$$

Similarly, we can compute the second moments and variance as following:

$$\begin{aligned} E\big[\omega_i^2\big] &= E\Bigg[\left(\sum_{j=1}^n u_j v_j W_{ij}M_{ij}\right)^2 + \left(\sum_{l\ne m} u_l v_m W_{il}M_{im}\right)^2 \\ &\quad + 2\left(\sum_{j=1}^n u_j v_j W_{ij}M_{ij}\right)\left(\sum_{l\ne m} u_l v_m W_{il}M_{im}\right)\Bigg] \\ &= \sum_{j=1}^n u_j^2 v_j^2 E\big[W_{ij}^2M_{ij}^2\big] + 2\sum_{l<m} u_l v_l u_m v_m E[W_{il}M_{il}W_{im}M_{im}] \\ &\quad + \sum_{l\ne m} u_l^2 v_m^2 E\big[W_{il}^2M_{im}^2\big] + 2\sum_{l<m} u_l v_m u_m v_l E[W_{il}M_{il}W_{im}M_{im}] \\ &= \sum_{j=1}^n u_j^2 v_j^2 \frac{1}{\pi_j} + 2\sum_{l\ne m} u_l v_l u_m v_m + \sum_{l\ne m} u_l^2 v_m^2 \frac{\pi_l}{\pi_m} \\ &\quad \left(Let\ \xi = max\left(\frac{\pi_l}{\pi_m}\right), \kappa = \frac{1}{min(\pi)}\right) \\ &\le 2\left(\sum_{j=1}^n u_j^2 v_j^2 + 2\sum_{l\ne m} u_l v_l u_m v_m\right) \\ &\quad + \xi\left(\sum_{j=1}^n u_j v_j + \sum_{l\ne m} u_l^2 v_m^2\right) + (\kappa - 2 - \xi)\sum_{j=1}^n u_j v_j \\ &= 2(u'v)^2 + \xi\|u\|_2^2\|v\|_2^2 + (\kappa - 2 - \xi)\sum_{j=1}^n u_j^2 v_j^2 \end{aligned}$$

$$\begin{aligned} Var(\omega_i) &= E\big[\omega_i^2\big] - E[\omega_i]^2 \\ &\le (u'v)^2 + \xi\|u\|_2^2\|v\|_2^2 + (\kappa - 2 - \xi)\sum_{j=1}^n u_j^2 v_j^2 \end{aligned}$$

TABLE 1
Data Sets for Evaluation

Name	Environment	Time period	Physical conditions	Selected Sub-Matrix	Time interval
GreenOrbs [35]	Forest	Aug. 03 $\sim$ 05, 2011	Temperature, light, humidity	326×750	5 minutes
IntelLab [36]	Indoor	Feb. 28 $\sim$ Apr. 5, 2004	Temperature, light, humidity	54×500	30 seconds
NBDC CTD [37]	Ocean	Oct. 26 $\sim$ 28, 2012	Temperature, salt, conductivity	216×300	10 minutes

$$
\begin{aligned}
Var(\zeta) &= \frac{1}{L^2}\sum_{i=1}^{L} Var(\omega_i) \\
&\le \frac{1}{L}\left((u'v)^2 + \xi\|u\|_2^2\|v\|_2^2 + (\kappa - 2 - \xi)\sum_{j=1}^{n} u_j^2 v_j^2\right).
\end{aligned}
$$

□

5.2 Theorem 1

Proof. Fix any two vectors $u, v \in \mathbb{R}^n$, with $\|u\|_\infty / \|u\|_2 \le H$. Set $L = L_1 L_2$, with L_1, L_2 are positive integers. Partition the $L \times n$ matrix W and M into L_2 matrices $\{W_1, W_2, \ldots, W_{L_2}\}$ and $\{M_1, M_2, \ldots, M_{L_2}\}$, each of size $L_1 \times n$. The corresponding random projections are $\{\alpha_1 = \frac{1}{\sqrt{L_1}} W_1 u, \ldots, \alpha_{L_2} = \frac{1}{\sqrt{L_1}} W_{L_2} u\}$, and $\{\beta_1 = \frac{1}{\sqrt{L_1}} W_1 v, \ldots, \beta_{L_2} = \frac{1}{\sqrt{L_1}} W_{L_2} v\}$.

Define the independent random variables $\zeta_l = \alpha_l' \beta_l$, $l = 1, 2, \ldots, L2$. Applying Lemma 1 to each ζ_l, we derive that $E[\zeta_l] = u'v$ and

$$
Var(\zeta) \le \frac{1}{L_1}\left((u'v)^2 + \xi\|u\|_2^2\|v\|_2^2 + (\kappa - 2 - \xi)\sum_{j=1}^{n} u_j^2 v_j^2\right).
$$

By the Chebyshev inequality and

$$
\begin{aligned}
P(|\zeta_l - u'v| \ge \epsilon\|u\|_2\|v\|_2) &\le \frac{Var(\zeta_l)}{\epsilon^2\|u\|_2^2\|v\|_2^2} \\
&\le \frac{1}{\epsilon^2 L_1}\left(\frac{(u'v)^2}{\|u\|_2^2\|v\|_2^2} + \frac{\xi\|u\|_2^2\|v\|_2^2}{\|u\|_2^2\|v\|_2^2} + (\kappa - 2 - \xi)\frac{\sum_{j=1}^{n} u_j^2 v_j^2}{\|u\|_2^2\|v\|_2^2}\right) \\
&(By\ the\ Cauchy-Schwarz\ inequality\ and\ Eq.\ (18)) \\
&\le \frac{1}{\epsilon^2 L_1}\left(1 + \xi + \kappa\frac{H^2\|u\|_2^2\sum_{j=1}^{n} v_j^2}{\|u\|_2^2\|v\|_2^2}\right) \\
&\le \frac{1}{\epsilon^2 L_1}\left(1 + \xi + \kappa H^2\right) \triangleq p.
\end{aligned}
$$

Thus, we can obtain a constant probability p by setting $L_1 = O(\frac{1+\xi+\kappa H^2}{\epsilon^2})$.

We define the estimate $\hat{a}$ as the median of the independent random variables $\zeta_1, \ldots, \zeta_{L_2}$, each of which lies outside of the tolerable approximation interval with probability p. Formally, let I_l be the indicator random variable of the event that $\{|\zeta_l - u'v| \ge \epsilon\|u\|_2\|v\|_2\}$, which occurs with probability p.

Let $I = \sum_{l=1} L_2 I_l$ be the number of ζ_l's that lie outside of the tolerable interval, where $E[I] = L_2 p$. When the event that at least half of the ζ_l's are outside the tolerable interval occurs with arbitrarily small probability, then the median $\hat{a}$ is within the tolerable interval. So, if we set $p < 1/2$, say $p = 1/4$, and apply the Chernoff bound, we get

$$
P\left(I > (1+c)\frac{L_2}{4}\right) < e^{-c^2 L_2/12},
$$

where $0 < c < 1$ is some constant.

Thus, for u and $v_i \in v_1, \ldots, v_n \subset \mathbb{R}^n$, and the corresponding random projections α and β produce an estimate $\hat{a}$ for $u'v$ that lies outside the tolerable approximation interval with probability at most $e^{-c^2 L_2/12}$. For the n estimates $\hat{a}_i, i = 1, 2, \ldots n$, the probability that at least one lies outside the tolerable interval is upper bounded by $P_e \le n e^{-c^2 L_2/12}$.

Setting $L_1 = O(\frac{1+\xi+\kappa H^2}{\epsilon^2})$ obtains $p = 1/4$, and setting $L_2 = O((1+\gamma)\log n)$ obtains $P_e \le n^{-\gamma}$ for some constant $\gamma > 0$. Therefore, with $L = L_1 L_2 = O(\frac{1+\gamma}{\epsilon^2}(1 + \xi + \kappa H^2)\log n)$, the nonuniform random projection pair W, M can produce inner products of vectors with probability at least $1 - n^{-\gamma}$. If $(\xi + \kappa H^2) > \Omega(1)$, then $L = O(\frac{1+\gamma}{\epsilon^2}(1 + \xi + \kappa H^2)\log n)$. If $(\xi + \kappa H^2) \le O(1)$, then $L = O(\frac{1+\gamma}{\epsilon^2}\log n)$. □

6 Evaluation

6.1 Simulation Settings

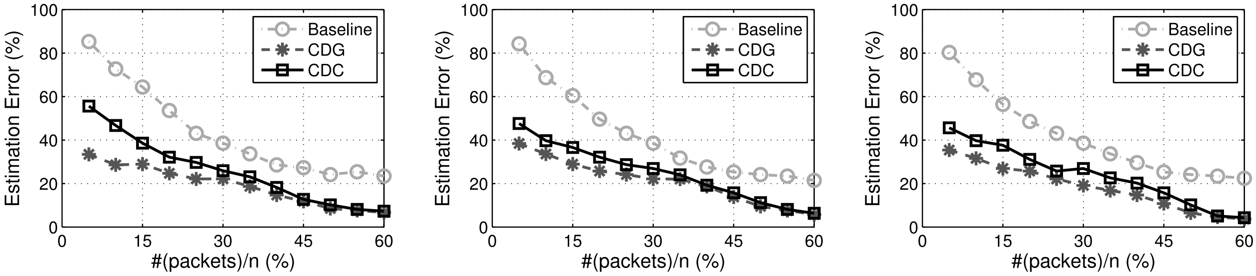
Ground truth. We use data sets collected from the GreenOrbs [35], IntelLab [36] and NBDC-CTD [37] projects, as summarized in Table 1. In our simulations, time is divided into slots of equal length, and the sensory data extracted from these data sets are used in the corresponding slots.

Network topology. Nodes' positions in these three WSNs are also provided. More accurate link quality model can provide more realistic simulation results. The *RSSI* value is the best indicator. However, this information may not be always known to the network protocol designer, then the distance will serve an acceptable choice. We adopt the following link quality models of the opportunistic routing for simulation:

- For the GreenOrbs project, the *RSSI* value between two neighbor nodes is given. We use a *RSSI-Link-Quality* model [30] to control the successful probability of transmitting a packet.
- For the IntelLab project, the aggregate connectivity information is provided. The probability of successful delivery in each link is averaged over all time. Note that this is not a symmetric relationship, i.e., sensor A may hear B better than B hears A.
- For the NBDC-CTD project, such information is absent, we use a *Distance-LinkQuality* model [31] instead.

6.2 Compared Algorithms

Baseline. Packets are transmitted back to the sink along the shortest path. Then the sink applies the k-nearest neighbors



(a) Estimation Error in GreenOrbs (b) Estimation Error in IntelLab (c) Estimation Error in NBDC-CTD

Fig. 3. The estimation error for Basline, CDG, and CDC in the GreenOrbs, IntelLab and NBDC-CTD projects.

(KNN) [38] method to estimate the readings, i.e., by averaging the k-nearest neighbors' values. Since both the routing and estimation are the most basic ones, therefore we use it as the baseline algorithm.

CDG (MobiCom'09). The CDG scheme compresses all sensor readings together in each collected packet. It uses the following tree-based routing: a node waits for all its children's packets, performs random linear compression, and then sends the packet to its parent node. The estimation uses the convex optimization method of traditional compressive sensing theory. It is a bit different from [11] since the link quality is not perfectly reliable as we allow the transmission to fail through a controlled probability and introduce the retransmission mechanism.

6.3 Metrics

Based on the network topology and sensory data sets of these three wireless sensor networks, we run the above three schemes to collect sensor readings back to the sink. For the baseline algorithm and the CDC scheme, we vary the probability p of generating TOKENs to get different number of random measurements. For equality, the CDG scheme will collect accordingly the same number of random measurements.

Estimation error. Each algorithm estimates a $\hat{U}$ for the original data U. The estimation error is defined as

$$E = \frac{\|U - \hat{U}\|_2}{\|\hat{U}\|_2}. \tag{23}$$

Delay. The data collection delay is defined as the time when the last packet arrives at the sink, which is measured in terms of the number of slots.

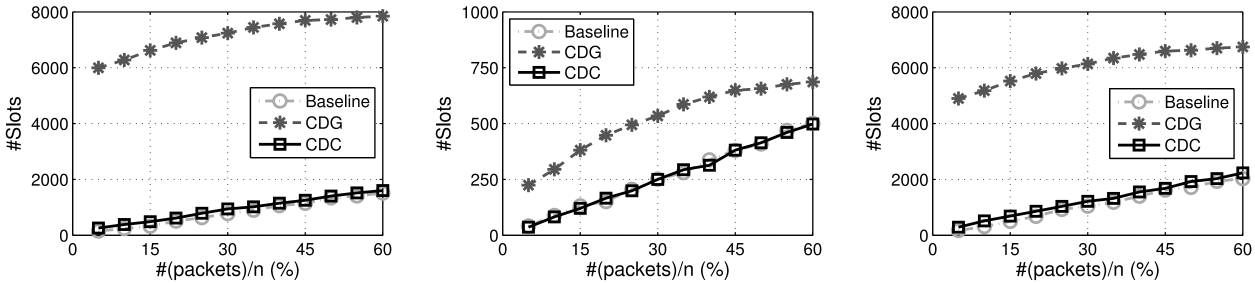
Network lifetime. The energy consumption is set according to the energy consumption model [39]. At the beginning, each node have initial energy of $1{,}000{,}000$ units which can support the sensor node to run about a month. The network lifetime is defined as the time when the first node runs out of energy, which is also measured in terms of the number of slots.

6.4 Results

From Figs. 3a, 3b, and 3c, CDG and CDC perform much better than the baseline algorithm and can reach estimation errors as low as 5 percent. This is because they both exploit the compressibility nature of the sensor readings and use random compression techniques. However, CDG behaves better in situations where less number of packets are collected. Possibly, less collected packets can lead to (1) stronger nonuniform nature of the compression probability of sensor nodes, or (2) less observations of the routing process, then less accurate of probability estimation in Section 4.3.

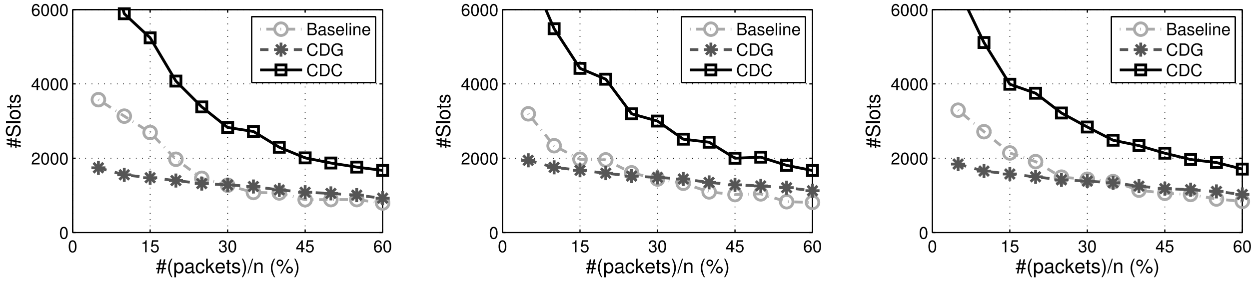
From Figs. 4a, 4b, and 4c, it is quite unexpected that the delay of the CDG scheme is several or even hundreds of times longer than the other two. The reason maybe that: CDG tries to encode every node's packets, and a parent node has to wait for all its children's packets before transmitting the compressed packet to its own parent node. Because the network size of the IntelLab project is smaller, the delay performance is closer for these three schemes. The baseline algorithm exhibits a good, stable, and moderate growth in delay since it used the shortest path routing. The CDC's routing strategy is quite similar with the baseline scheme, therefore it experiences quite similar performance in terms of delay.

From Figs. 5a, 5b, and 5c, we find that our scheme has the best performance. For estimation error within 20 percent, CDC prolongs the network lifetime by $1.5\times \sim 2\times$. This is because that CDC requires fewer measurements, even fewer number of sensory readings in each measurement, thus greatly reduces the energy consumption.



(a) GreenOrbs (b) IntelLab (c) NBDC-CTD

Fig. 4. The delay for Basline, CDG, and CDC in the GreenOrbs, IntelLab, and NBDC-CTD projects.



(a) GreenOrbs (b) IntelLab (c) NBDC-CTD

Fig. 5. The network lifetime for Basline, CDG, and CDC in the GreenOrbs, IntelLab, and NBDC-CTD projects.

7 Conclusion and Future Work

We have proposed a novel compressive data collection scheme for wireless sensor networks. This scheme leverages the fact that raw sensory data have strong spatiotemporal compressibility. Our scheme consists of two parts: the opportunistic routing with compression, and the nonuniform random projection based estimation. The proposed scheme agrees with Braniuk's [1] suggestion that sensory data acquisition should be more efficient, and the new techniques that combine sensing and network communication together is a promising approach. We prove that this scheme can achieve optimal approximation error, and trace based evaluations show that its error is comparable with the existing method [11]. More important, our scheme exhibits good performance for energy-conservation.

In the future, we will apply this nonuniform random projection based estimator for adaptive data gathering [40] and the multiple attributes scenario [41].

Acknowledgments

This research was supported by the 973 Program (2014CB340303), National 863 Program (2013AA01A601), NSFC (No. 61170238, 60903190, 61303202, 61073158, 61100210), STCSM Project No. 12dz1507400, FQRNT grant 131844, Singapore-MIT IDC IDD61000102a, SUTD-ZJU/RES/03/2011, and NRF2012EWT-EIRP002-045, Singapore NRF (CREATE E2S2), iTrust IGDSi1305013, and Singapore-MIT International Design Center IDG31000101. This work was also supported by the Program for Changjiang Scholars and Innovative Research Team in the University (IRT1158, PCSIRT), China. Xiao-Yang Liu gratefully acknowledges financial support from China Scholarship Council.

References

[1] R. G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717–719, 2011.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Elsevier Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[3] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, N. Gordon, T. Bessell, M. Rutten, and S. Jha, "Wireless sensor networks for battlefield surveillance," presented at the Land Warfare Conf., Brisbane, Australia, 2006.

[4] L. Kong, M. Zhao, X.-Y. Liu, J. Lu, Y. Liu, M.-Y. Wu, and W. Shu, "Surface coverage in sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 234–243, Jan. 2014.

[5] X.-Y. Liu, K.-L. Wu, Y. Zhu, L. Kong, and M.-Y. Wu, "Mobility increases the surface coverage of distributed sensor networks," *Elsevier Comput. Netw.*, vol. 57, no. 11, pp. 2348–2363, 2013.

[6] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Proc. ACM 2nd Int. Conf. Embedded Netw. Sensor Syst.*, 2004, pp. 214–226.

[7] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon, "Health monitoring of civil infrastructures using wireless sensor networks," in *Proc. 6th Int. Conf. Inf. Process. Sensor Netw.*, 2007, pp. 254–263.

[8] L. Selavo, A. Wood, Q. Cao, T. Sookoor, H. Liu, A. Srinivasan, Y. Wu, W. Kang, J. Stankovic, D. Young, and J. Porter, "LUSTER: Wireless sensor network for environmental research," in *Proc. 5th Int. Conf. Embedded Netw. Sensor Syst.*, 2007, pp. 103–116.

[9] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[10] L. Fu, Y. Qin, X. Wang, and X. Liu, "Converge-cast with MIMO," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 649–657.

[11] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 145–156.

[12] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2013, pp. 1654–1662.

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

[14] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: A survey," *IEEE Wireless Commun.*, vol. 14, no. 2, pp. 70–87, Apr. 2007.

[15] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proc. IEEE Conf. Comput. Commun.*, 2004, pp. 2571–2582.

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

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

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

[19] H. Zheng, F. Yang, X. Tian, X. Gan, X. Wang, and S. Xiao, "Data gathering with compressive sensing in wireless sensor networks: A random walk rased approach," *IEEE Trans. Parallel Distrib. Syst.*, 2014, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6748092

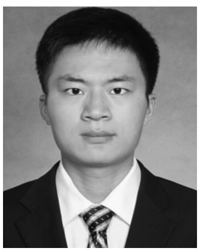
[20] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 7. no. 3, pp. 537–568, 2009.

[21] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Network correlated data gathering with explicit communication: NP-completeness and algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 41–54, Feb. 2006.

[22] X. Liu, J. Luo, and A. V. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *Proc. 8th Annu. IEEE Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, 2011, pp. 46–54.

[23] L. Guo, R. Beyah, and Y. Li, "SMITE: A stochastic compressive data collection protocol for mobile wireless sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 1611–1619.

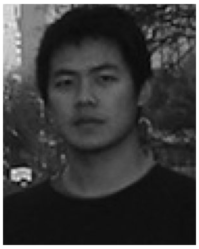
[24] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 2087–2095.
[25] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the interplay between routing and signal representation for compressive sensing in wireless sensor networks," in *Proc. IEEE Inf. Theory Appl. Workshop*, 2009, pp. 206–215.
[26] W. Wang, M. Garofalakis, and K. W. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proc. 6th IEEE Int. Symp. Inf. Process. Sensor Netw.*, 2007, pp. 331–339.
[27] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
[28] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, Mar. 2008.
[29] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.
[30] M. H. Lu, P. Steenkiste, and T. Chen, "Design, implementation and evaluation of an efficient opportunistic retransmission protocol," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 73–84.
[31] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 69–74, 2004.
[32] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," *ACM SIGCOMM Comput. Comuun. Rev.*, vol. 37, no. 4, pp. 169–180, 2007.
[33] S. Johansen and K. Juselius, "Maximum likelihood estimation and inference on cointegration—with applications to the demand for money," *Oxford Bull. Econ. Statist.*, vol. 52, no. 2, pp. 169–210, 1990.
[34] C. Sherlaw-Johnson, S. Gallivan, and J. Burridge, "Estimating a Markov transition matrix from observational data," *J. Oper. Res. Soc.*, vol. 46, pp. 405–410, 1995.
[35] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, and X.-Y. Li, "Does wireless sensor network scale? A measurement study on GreenOrbs," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 873–881.
[36] IntelLab data. (2004). [Online]. Available: http://www.select.cs.cmu.edu/data/labapp3/index.html
[37] NBDC-CTD data. (2013). [Online]. Available: http://tao.noaa.gov/refreshed/ctd_delivery.php
[38] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
[39] X. Jiang. M. V. Ly, J. Taneja, P. Dutta, and D. Culler, "Experiences with a high-fidelity wireless building energy auditing network," in *Proc. 7th ACM Conf. Embedded Netw. Sensor Syst.*, 2009, pp. 113–126.
[40] L. Kong, X.-Y. Liu, M. Tao, M.-Y. Wu, Y. Gu, L. Cheng, and J. Niu, "Resource-efficient data gathering in sensor networks for environment reconstruction," *The Comput. J.*, 2014, http://comjnl.oxfordjournals.org/content/early/2014/06/28/comjnl.bxu054.full.pdf+html
[41] G. Chen, X.-Y. Liu, L. Kong, J.-L. Lu, and M.-Y. Wu, "Multi-attribute compressive data gathering," *IEEE WCNC*, pp. 1–6, 2014.



Xiao-Yang Liu received the BEng degree in computer science from the Huazhong University of Science and Technology, China, in 2010. He is currently working toward the PhD degree in the Department of Computer Science and Engineer at Shanghai Jiao Tong University. His research interests include wireless communication, sensor networks, and distributed systems.

Yanmin Zhu (M'02) received the PhD degree in computer science from the Hong Kong University of Science and Technology in 2007. He is currently an associate professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. His research interests include vehicular networks, sensor networks, and mobile computing. Prior to joining Shanghai Jiao Tong University, he was a research associate with the Department of Computing at the Imperial College London. He is a member of the IEEE and IEEE Communication Society.

Linghe Kong (S'09-M'13) received the BE degree in automation from Xidian University in 2005, the Dipl. Ing degree in telecommunication from TELECOM SudParis in 2007, and the PhD degree in computer science from Shanghai Jiao Tong University in 2012. He is currently a research assistant professor with Shanghai Jiao Tong University and also a postdoctoral researcher in the Singapore University of Technology and Design. His research interests include wireless sensor networks and mobile computing. He is a member of the IEEE.



Cong Liu received the PhD degree in computer science from the University of North Carolina at Chapel Hill in July 2013. He is currently an assistant professor in the Department of Computer Science at the University of Texas at Dallas. His research interests include real-time and embedded systems, operating systems, cloud computing, and wireless sensor networks. He has published more than 30 papers in premier conferences and journals such as RTSS, PACT, ECRTS, and the *Journal of Parallel and Distributed Computing*. He received the Best Student Paper Award at the 30th IEEE Real-Time Systems Symposium, the premier real-time and embedded systems conference. He also received the Best Papers Award at the 17th RTCSA and the UNC's dissertation fellowship.

Yu Gu received the PhD degree from the University of Minnesota, Twin Cities, in 2010. He is currently an assistant professor in the Pillar of Information System Technology and Design at the Singapore University of Technology and Design. His research includes networked embedded systems, wireless sensor networks, cyber-physical systems, wireless networking, real-time and embedded systems, distributed systems, vehicular ad-hoc networks, and stream computing systems. He is a member of the IEEE.

Athanasios V. Vasilakos (M'00-SM'11) is currently a professor with the University of Western Macedonia, Greece. He was/is an editor for many technical journals, such as the *IEEE Transactions on Network and Service Management*, *IEEE Transactions on Cloud Computing*, *IEEE Transactions on Information Forensics and Security*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Nanobioscience*, *IEEE Transactions on Information Technology in Biomedicine*, *ACM Transactions on Autonomous and Adaptive Systems*, and the *IEEE Journal on Selected Areas in Communications*. He is also the general chair of the European Alliances for Innovation (www.eai.eu). He is a senior member of the IEEE.

Min-You Wu (S'84-M'85-SM'96) received the MS degree from the Graduate School of Academia Sinica, Beijing, China, in 1981, and the PhD degree from Santa Clara University, Santa Clara, CA, in 1984. He is a professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University, and a research professor with the University of New Mexico. He is the chief scientist at the Grid Center of Shanghai Jiao Tong University. His research interests include grid computing, wireless networks, sensor networks, multimedia networking, parallel and distributed systems, and compilers for parallel computers. He is a senior member of the IEEE.