

Research Article

Unbalanced Threshold Based Distributed Data Collection Scheme in Multisink Wireless Sensor Networks

Guorui Li,¹ Ying Wang,² Cong Wang,¹ and Yiyang Liu¹

¹College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

²Department of Information Engineering, Qinhuangdao Institute of Technology, Qinhuangdao, Hebei 066100, China

Correspondence should be addressed to Guorui Li; lgr@mail.neuq.edu.cn

Received 18 September 2015; Revised 19 November 2015; Accepted 1 December 2015

Academic Editor: Yao Liang

Copyright © 2016 Guorui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In multisink wireless sensor networks, synchronized data collection among multiple sinks is a significant and challenging task. In this paper, we propose an unbalanced threshold based distributed data collection scheme to reconstruct the synchronized sensed data of the whole sensor network in all sinks. The proposed scheme includes the unbalanced threshold based distributed top- $|K|$ query algorithm and the distributed iterative hard thresholding algorithm. By computing unbalanced thresholds and pruning unnecessary element exchanging, each sink can synchronize the top- $|K|$ aggregated values efficiently via the unbalanced threshold based distributed top- $|K|$ query algorithm. After that, the synchronized sensed data of the whole sensor network can be reconstructed through the distributed iterative hard thresholding algorithm in a distributed and cooperative manner. We show through experiments that the proposed scheme can reduce the interaction times and decrease the number of transmitted data and that of computed data compared to the existing schemes while maintaining the similar data reconstruction accuracy. The communication and computational performances of the proposed scheme are also analyzed in detail in the paper.

1. Introduction

Wireless sensor network (WSN) is an autonomous wireless network consisting of a large number of tiny, inexpensive, and spatially distributed wireless sensor nodes. It has emerged as one of the most promising technologies for the future [1]. The typical applications of WSN include environment monitoring, industrial process control, intelligent transportation, military surveillance, and health care monitoring [2].

However, wireless sensor nodes have severe resource constraints which include limited computational, memory, communication capacities and nonrenewable energy supply. Therefore, data collection schemes in wireless sensor network should be light weight and energy efficient. Two representative types of data collection schemes in wireless sensor networks are spatial-temporal correlation based data predication schemes [3–5] and distributed source coding schemes [6–8]. In the first type of data collection scheme, a series of spatial-temporal correlation based data prediction algorithms are adopted to prolong the system lifetime by enabling the sink to predict the sensed data based on some historical samples.

In the second type of data collection scheme, distributed source coding techniques, such as Slepian-Wolf coding and Wyner-Ziv coding, are designed to compress the multiple correlated sensor data without the need of communication among sensor nodes. Nevertheless, the computational and communication cost of the above two types of data collection schemes are still high. Meanwhile, the parameters of predication model and the distributed coded data are still required to be computed and transmitted within the network frequently.

Recently, compressive sensing (CS) theory has attracted considerable attention in areas of signal processing, computer science, and applied mathematics [9]. It is based on the principle that a sparse signal can be reconstructed from far fewer samples than that required by the classic Shannon-Nyquist sampling theory. Through nonlinear optimization, we can recover many natural signals which can be represented by a few nonzero coefficients under a suitable basis [10]. The inherent characteristics, such as compressibility, robustness, versatility, and stability, have made compressive sensing theory widely applied to communication [11], photography [12], magnetic resonance imaging [13], astronomy [14], and so forth.

As for the data collection problem in wireless sensor networks, a number of compressive sensing based schemes have been proposed in the literature. In 2009, Luo et al. proposed the first complete compressive sensing based data gathering (CDG) scheme for large scale wireless sensor networks [15]. By adapting a weighted sum of all the readings along the routing tree, CDG scheme can achieve substantial communication cost reduction and balanced energy consumption simultaneously. In order to achieve both energy efficiency and recovery fidelity, Xiang et al. proposed the Compressive Data Aggregation (CDA) scheme in WSN [16]. After abstracting the minimum energy compressed data aggregation problem as a NP-completeness problem, they solve it using a mixed integer programming formulation method. Moreover, they also proposed the Dual-Level Compressed Aggregation (DECA) framework to recover the in-network aggregated data in the widespread monitoring area by utilizing both the low-rank nature of real-world events and the redundancy in sensed data [17]. By exploiting the spatial and temporal correlation among sensed data, Xu et al. proposed the Spatio-Temporal Hierarchical Data Aggregation Scheme Using Compressive Sensing (ST-HDACS) in [18]. After collecting the sensed data from a subset of randomly selected sensor nodes through a hierarchical routing structure, the fusion center recovers the whole sensed data by solving a matrix completing problem. Furthermore, Tang et al. proposed a robust compressive data gathering scheme to identify outlying sensor readings and derive the corresponding accurate values and further infer broken links in the sensor network [19]. Motivated by the fact that the variance among all the solutions of the blind iterative hard thresholding algorithm with different sparse level can indicate the best sparse level, Xiong and Tang proposed the blind 1-bit compressive sensing reconstruction algorithm for wireless sensor network [20]. Similar research works include the sparse random scheduling based data gathering scheme [21], the Resource Efficient Data Gathering (REDG) scheme [22], the Compressive Sensing Based Data Aggregation Scheme (CSDAS) [23], and the distributed multichain based compressive sensing scheme [24].

Although the above schemes can achieve energy-efficient data collection based on compressive sensing theory, they can only be applied to single-sink wireless sensor networks in a centralized manner. In the scenario of multisink wireless sensor networks, compressed data should be transmitted to one specific sink before executing the data reconstruction algorithm. Then, the recovered data should be distributed back to other sinks through unicast or broadcast messages. Obviously, this kind of centralized pattern is less energy efficient and time consuming. In this paper, we focus on designing compressive sensing based distributed data collection scheme for multisink wireless sensor networks. The proposed scheme includes two correlated algorithms: the unbalanced threshold based distributed top- $|K|$ algorithm and the distributed iterative hard thresholding algorithm. In the first algorithm, the largest K absolute aggregated values of a series of object-value pairs are queried in a distributed and energy efficient way. By computing unbalanced thresholds, a number of unnecessary object-value pairs exchanging

are pruned in the query process. In the second algorithm, the iterative hard thresholding algorithm [25] is modified to fit the distributed environment. Meanwhile, the first algorithm is integrated into the second one as a subroutine to realize the $H_K()$ operator. Our experiment results indicate that the proposed scheme can reduce the interaction times and decrease the number of transmitted data and that of computed data compared to the existing schemes while maintaining the similar data reconstruction accuracy.

The contributions of this paper are threefold.

(1) We propose an unbalanced threshold based distributed data collection scheme in multisink wireless sensor networks based on compressive sensing theory. By decomposing the task of sparse data reconstruction into several correlated parts, all sinks can obtain the synchronized sensed data of the whole network in a distributed and cooperative manner.

(2) We propose an unbalanced threshold based distributed top- $|K|$ query algorithm to query the top- $|K|$ aggregated values required by the iterative hard thresholding algorithm in the distributed environment. By adopting unbalanced thresholds, the algorithm can avoid transmitting unnecessary object-value pairs among sinks. Furthermore, the correctness of the proposed algorithm is also proved in this paper.

(3) We analyze the data reconstruction performance, communication complexity, and computational complexity of the proposed scheme through experiments. Furthermore, we also compare the performance of the proposed scheme with existing compressive sensing based data collection schemes.

The rest of this paper is organized as follows. Section 2 introduces the basic principles of compressive sensing theory and the iterative hard thresholding algorithm. Sections 3 and 4 describe the unbalanced threshold based distributed top- $|K|$ query algorithm and the distributed iterative hard thresholding algorithm, respectively. Section 5 presents the experimental results and analysis. Finally, Section 6 concludes this paper.

2. Preliminaries

Compressive sensing theory asserts that a sparse signal can be recovered from fewer linear and nonadaptive measurements than the number of samples required by Shannon-Nyquist Sampling Theorem. The compressive sampling process can be presented as

$$y = \Phi f + z, \quad (1)$$

where $f \in R^N$ is a K sparse unknown signal; that is, there are at most K nonzero elements in f . $\Phi \in R^{M \times N}$ is the measurement matrix, $z \in R^M$ is the unknown noise, and $y \in R^M$ is the measurement signal. Although most natural signals are not sparse directly, they can be transformed into sparse signals. In other words, a signal f can be represented

as a sparse signal $x \in R^N$ under an orthonormal basis $\Psi \in R^{N \times N}$; that is, $f = \Psi x$. Hence, (1) can be rewritten as

$$y = \Phi f + z = \Phi \Psi x + z \quad (2)$$

and matrix $A = \Phi \Psi$ is usually referred to as sensing matrix.

One of the most important properties of sensing matrix is the Restrict Isometry Property (RIP). Formally, a matrix A is said to satisfy the RIP of order K with a Restrict Isometry Constant (RIC) δ_K such that

$$(1 - \delta_K) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_K) \|x\|_2^2 \quad (3)$$

holds for all K sparse signal x . Baraniuk et al. have proved that a random matrix $A \in R^{M \times N}$ with i.i.d. entries followed by normal distribution $N(0, 1/M)$ satisfies the RIP with a high probability when $M \geq CK \log(N/K)$ and C is a constant [26].

The Basis Pursuit De-Noise (BPDN) [27] and the Least Absolute Shrinkage and Selection Operator (LASSO) [28] are two typical signal recovery frameworks which are widely used in compressive sensing theory. In BPDN framework, the l_1 norm of the unknown signal is minimized under the constraint of limited energy of reconstruction error; that is,

$$\begin{aligned} \min_x \quad & \|x\|_1, \\ \text{s.t.} \quad & \|Ax - y\|_2 \leq \varepsilon. \end{aligned} \quad (4)$$

In contrast, the energy of reconstruction error is minimized in LASSO framework under the constraint of limited l_1 norm of the unknown signal x ; that is,

$$\begin{aligned} \min_x \quad & \|Ax - y\|_2^2, \\ \text{s.t.} \quad & \|x\|_1 \leq \eta. \end{aligned} \quad (5)$$

Essentially, the BPDN framework and LASSO framework can all be transformed into the following unconstrained regularized optimization framework:

$$\min_x \quad \frac{1}{2} \|Ax - y\|_2^2 + \tau \|x\|_1. \quad (6)$$

A number of data reconstruction algorithms have been proposed in the literature. In general, they can be classified into two categories: optimization algorithms and greedy algorithms. In the optimization algorithms, a number of convex optimization methods have been adopted to search the optimized solution of (6). Basis Pursuit (BP) algorithm [27], Interior Point (IP) algorithm [29], and homotopy algorithm [30] are representatives of this type of reconstruction algorithm. On the contrary, greedy algorithms iteratively select the support sets or the elements of the reconstructed signal based on certain greedy selection strategy. Orthogonal Matching Pursuit (OMP) algorithm [31], Regularized Orthogonal Matching Pursuit (ROMP) algorithm [32], Compressive Sampling Matching Pursuit (CoSmMP) algorithm [33], Subspace Matching Pursuit (SP) algorithm [34], iterative hard thresholding (IHT) algorithm [25], and Iterative Soft Thresholding (IST) algorithm [35] are representatives of greedy algorithm.

Among all proposed data reconstruction algorithms, the iterative hard thresholding algorithm is very easy to implement and can be relatively fast. Meanwhile, it also possesses strong theoretical recovery performance guarantees as convex optimization based algorithms. Essentially, it can be regarded as an iterative method

$$x^{(t+1)} = H_K(x^{(t)} + \mu A^T(y - Ax^{(t)})), \quad (7)$$

where $H_K()$ is the hard thresholding operator that sets all but K largest elements in magnitude to zero and μ is the step size. Formally, we can represent the hard thresholding operator $H_K()$ as

$$H_K(x)_i = \begin{cases} x_i & x_i \in \{x_{\text{top-}|K|}\} \\ 0 & x_i \notin \{x_{\text{top-}|K|}\}, \end{cases} \quad (8)$$

where $x_{\text{top-}|K|}$ represents the largest K elements of signal x in magnitude. We will propose a distributed implementation of hard thresholding operator $H_K()$ in next section and integrate it in the distributed iterative hard thresholding algorithm which is described in Section 4.

3. The Unbalanced Threshold Based Distributed Top- $|K|$ Query Algorithm

In recent years, top- K query, also known as ranking aware query, has received much attention in the areas of relational database system, content distribution network, multimedia retrieval system, and so forth. However, only monotonic aggregation functions can be applied to the top- K query. The largest K absolute elements query in the hard thresholding operator is nonmonotonic. Therefore, the proposed top- K query algorithms, such as thresholding algorithm (TA) [36] and three-phase uniform threshold (TPUT) algorithm [37], cannot be applied to the distributed IHT algorithm directly. We will describe our proposed unbalanced threshold based distributed top- $|K|$ query algorithm in detail in Section 3.1 and illustrate it with an example in Section 3.2.

3.1. Algorithm Design. We assume there are P nodes in the distributed system and each node is assigned with an ID. The node with the lowest ID is designated as the administrator node and other nodes are designated as the member nodes. Each node i maintains a descending ordered list L_i of its object-value pairs $(O_j, V_i(j))$ where index j ranges from 1 to element number N_i . Note that if an object does not appear in the list, its value is set to zero by default. We will select the largest K sums in magnitude of all the objects in the distributed system.

The core idea of the unbalanced threshold based distributed top- $|K|$ query algorithm is to filter out unnecessary exchanging of elements among distributed nodes by adopting unbalanced thresholds. It includes three phases, that is, the unbalanced threshold computing phase, the candidate set computing phase, and the top- $|K|$ elements computing phase.

In the unbalanced threshold computing phase, each member node sends its first K positive elements and last

K negative elements to the administrator node. Then, the administrator node can compute the partial sum $P(j)$ for every received element j according to

$$P(j) = \sum_{i=1}^P V_i'(j), \quad (9)$$

$$V_i'(j) = \begin{cases} V_i(j) & V_i(j) \text{ is known} \\ 0 & V_i(j) \text{ is unknown} \end{cases}$$

and select the K th largest positive partial sum $\bar{\tau}^1$ and the K th smallest negative partial sum $\underline{\tau}^1$. The partial sum $P(j)$ is used to estimate the sum of top- $|K|$ elements approximately. In order to deal with unbalanced value distribution among different nodes and avoid redundant element exchanging between administrator node and member nodes, we compute different thresholds for each node instead of fixing one identical threshold for all nodes. Based on the computed positive weight $\bar{\omega}_i$ and negative weight $\underline{\omega}_i$, the partial sums $\bar{\tau}^1$ and $\underline{\tau}^1$ are divided into P different thresholds \bar{T}_i and \underline{T}_i , respectively.

In the candidate set computing phase, each member node sends unsent object-value pairs $(O_j, V_i(j))$ to the administrator node when $V_i(j) \geq \bar{T}_i$ or $V_i(j) \leq \underline{T}_i$. Then, the administrator node computes the partial sum according to (9) and selects the K th largest positive partial sum $\bar{\tau}^2$ and K th smallest negative partial sum $\underline{\tau}^2$ once again. Furthermore, it computes the upper bound $U(j)$ and lower bound $L(j)$ of the whole sum for every received object j according to

$$U(j) = \sum_{i=1}^P \bar{V}_i(j), \quad (10)$$

$$\bar{V}_i(j) = \begin{cases} V_i(j) & V_i(j) \text{ is known} \\ \bar{T}_i & V_i(j) \text{ is unknown,} \end{cases}$$

$$L(j) = \sum_{i=1}^P \underline{V}_i(j), \quad (11)$$

$$\underline{V}_i(j) = \begin{cases} V_i(j) & V_i(j) \text{ is known} \\ \underline{T}_i & V_i(j) \text{ is unknown,} \end{cases}$$

respectively. The upper bound and lower bound are used to estimate the maximal range of each object. If $U(j) < \bar{\tau}^2$ or $L(j) > \underline{\tau}^2$, we can guarantee that the whole sum of object j

$$W(j) = \sum_{i=1}^P V_i(j) \quad (12)$$

is not in the set of top- $|K|$ aggregated sums. Hence, we can exclude object j from the candidate set with confidence.

In the top- $|K|$ elements computing phase, each member node sends unsent object-value pairs $(O_j, V_i(j))$ in candidate set S to the administrator node. Then, the administrator node

can compute the whole sum for each object according to (12) and select the top- $|K|$ aggregated sums.

The pseudocode of the unbalanced threshold based distributed top- $|K|$ query algorithm is presented in Algorithm 1.

Lemma 1. *The unbalanced thresholds computed in phase I can guarantee that the true top- $|K|$ objects are among the objects in phase II.*

Proof. Assume that O_k is an object which is not sent to the administrator node in phase II. Then its value $V_i(k)$ in node i is less than \bar{T}_i and greater than \underline{T}_i at the same time; that is, $\underline{T}_i < V_i(k) < \bar{T}_i$. Then the sum of object O_k 's value can be bounded by

$$\sum_{i=1}^P \underline{T}_i < \sum_{i=1}^P V_i(k) < \sum_{i=1}^P \bar{T}_i. \quad (13)$$

Since $\sum_{i=1}^P \underline{T}_i = \sum_{i=1}^P (\underline{\tau}^1 \underline{\omega}_i / \sum_{i=1}^P \underline{\omega}_i) = \underline{\tau}^1$ and $\sum_{i=1}^P \bar{T}_i = \sum_{i=1}^P (\bar{\tau}^1 \bar{\omega}_i / \sum_{i=1}^P \bar{\omega}_i) = \bar{\tau}^1$, (13) can be rewritten as

$$\underline{\tau}^1 < \sum_{i=1}^P V_i(k) < \bar{\tau}^1. \quad (14)$$

That is, object O_k cannot be a top- $|K|$ object. Therefore, true top- $|K|$ objects must be among the objects in phase II. \square

Lemma 2. *The candidate set computed in phase II can guarantee that the true top- $|K|$ objects are among the objects in phase III.*

Proof. Assume that O_k is an object which is not included in candidate set S in phase II. Then the sum of object O_k 's value satisfies $\sum_{i=1}^P V_i(k) < U(j) < \bar{\tau}^2$ and $\sum_{i=1}^P V_i(k) > L(j) < \underline{\tau}^2$ at the same time; that is,

$$\underline{\tau}^2 < \sum_{i=1}^P V_i(k) < \bar{\tau}^2. \quad (15)$$

Since $\bar{\tau}^2$ and $\underline{\tau}^2$ are the K th largest positive partial sum and the K th smallest negative partial sum, there are already K objects whose partial sums are greater than $\bar{\tau}^2$ and lesser than $\underline{\tau}^2$. That is, object O_k cannot be a top- $|K|$ object. Therefore, true top- $|K|$ objects must be among the objects in phase III. \square

Theorem 3. *The unbalanced threshold based distributed top- $|K|$ query algorithm can obtain the largest K aggregated sums in magnitude of all objects in the distributed system.*

Proof. Through Lemmas 1 and 2, we can prove that true top- $|K|$ objects must be among the objects in phase III. Furthermore, the administrator node has obtained all the values for each object in the candidate set. Therefore, the proposed algorithm can obtain the largest K aggregated sums in magnitude of all objects in the distributed system. \square

3.2. Example. We will illustrate the unbalanced threshold based distributed top- $|K|$ query algorithm with an example in

Input: The descending list $L_i = \{(O_j, V_i(j)) \mid i \in \{1, \dots, P\}, j \in \{1, \dots, N_i\}\}$

Phase I: The unbalanced thresholds computing phase

If node i is administrator node **Then**

Computes partial sums $P(j)$ for received elements according to (9);

Let $\bar{\tau}^1 = \min\{P_j\}_{\text{top-}K}$ and $\underline{\tau}^1 = -\min\{-P_j\}_{\text{top-}K}$;

Computes weights $\bar{\omega}_i = \sum_{j=1}^K V_i(j)/K$ and $\underline{\omega}_i = \sum_{j=N_i-K+1}^{N_i} V_i(j)/K$;

Sends thresholds $\bar{T}_i = \bar{\tau}^1 \bar{\omega}_i / \sum_{i=1}^P \bar{\omega}_i$ and $\underline{T}_i = \underline{\tau}^1 \underline{\omega}_i / \sum_{i=1}^P \underline{\omega}_i$ to member nodes;

Else

Sends $\{(O_j, V_i(j)) \mid j \in \{1, \dots, K\} \cup \{N_i - K + 1, \dots, N_i\}\}$ to administrator node;

Phase II: The candidate set computing phase

If node i is administrator node **Then**

Computes partial sums $P(j)$ for received elements according to (9);

Let $\bar{\tau}^2 = \min\{P_j\}_{\text{top-}K}$ and $\underline{\tau}^2 = -\min\{-P_j\}_{\text{top-}K}$;

Computes upper bounds $U(j)$ and lower bounds $L(j)$ of whole sums for received elements according to (10) and (11);

Sends candidate set $S = \{j \mid j \in \{U(j) \geq \bar{\tau}^2\} \cup \{L(j) \leq \underline{\tau}^2\}\}$ to member nodes;

Else

Sends unsent $\{(O_j, V_i(j)) \mid j \in \{V_i(j) \geq \bar{T}_i\} \cup \{V_i(j) \leq \underline{T}_i\}\}$ to administrator node;

Phase III: The top- $|K|$ elements computing phase

If node i is administrator node **Then**

Computes whole sums $W(j)$ for received elements according to (12);

Let $S_{\text{top-}|K|} = \{(O_j, V(j)) \mid j \in \{|W_j|\}_{\text{top-}K}\}$;

Else

Sends unsent $\{(O_j, V_i(j)) \mid j \in S\}$ to administrator node;

Output: The top- $|K|$ element set $S_{\text{top-}|K|}$

ALGORITHM 1: The unbalanced threshold based distributed top- $|K|$ query algorithm.

this subsection. There are 3 nodes in the distributed system, one administrator node and two member nodes. The object-value pairs in each node are presented in Table 1.

In the unbalanced thresholds computing phase, member node 1 and member node 2 send $\{(O_4, 34), (O_5, 28), (O_1, -14), (O_8, -18)\}$ and $\{(O_5, 20), (O_4, 17), (O_1, -13), (O_9, -16)\}$ to administrator node, respectively. The administrator node then computes partial sums for all received elements according to (9). For object 5, partial sum $P(5) = 69$ because $V_1(5) = 21$, $V_2(5) = 28$, and $V_3(5) = 20$ are all known. For object 4, partial sum $P(4) = 51$ because $V_2(4) = 34$, $V_3(4) = 17$ are known but $V_1(4)$ is unknown ($V_1(4)$ is located in neither the first 2 nor the last 2 cells in list L_1). Similarly, we can get $P(2) = 17$, $P(8) = -30$, $P(9) = -46$, and $P(1) = -27$. Thus, the top-2 positive partial sum $\bar{\tau}^1 = 51$ and the top-2 negative partial sum $\underline{\tau}^1 = -30$. According to Algorithm 1, the weights and thresholds for each node are $\bar{\omega}_1 = 19$, $\bar{\omega}_2 = 31$, $\bar{\omega}_3 = 18.5$, $\underline{\omega}_1 = -21$, $\underline{\omega}_2 = -16$, and $\underline{\omega}_3 = -14.5$ and $\bar{\tau}_{11} = 14.2$, $\bar{\tau}_{12} = 23.0$, $\bar{\tau}_{13} = 13.8$, $\underline{\tau}_{11} = -12.2$, $\underline{\tau}_{12} = -9.3$, and $\underline{\tau}_{13} = -8.5$, respectively.

In the candidate set computing phase, member node 1 and member node 2 send $\{(O_9, -10)\}$ and $\{(O_3, 14), (O_8, -11)\}$ to administrator node, respectively. The administrator node then computes partial sums according to (9) and gets $P(5) = 69$, $P(4) = 66$, $P(2) = 17$, $P(3) = 14$, $P(8) = -41$, $P(9) = -56$, and $P(1) = -27$. The top-2 positive partial sum $\bar{\tau}^2$ and top-2 negative partial sum $\underline{\tau}^2$ are updated to 66 and -56, respectively. The administrator node then computes upper bounds and lower bounds for each received element and gets $U(5) = L(5) = 69$, $U(4) = L(4) = 66$, $U(2) = 53.8$,

TABLE 1: An example with 3 nodes.

| Administrator node | Member node 1 | Member node 2 |
|--------------------|---------------|---------------|
| $(O_5, 21)$ | $(O_4, 34)$ | $(O_5, 20)$ |
| $(O_2, 17)$ | $(O_5, 28)$ | $(O_4, 17)$ |
| $(O_4, 15)$ | $(O_2, 22)$ | $(O_3, 14)$ |
| $(O_3, 10)$ | $(O_3, 18)$ | $(O_2, 9)$ |
| $(O_6, 8)$ | $(O_7, 11)$ | $(O_7, 7)$ |
| $(O_7, 2)$ | $(O_6, 5)$ | $(O_0, 1)$ |
| $(O_0, -2)$ | $(O_0, -1)$ | $(O_6, -6)$ |
| $(O_1, -7)$ | $(O_9, -10)$ | $(O_8, -11)$ |
| $(O_8, -12)$ | $(O_1, -14)$ | $(O_1, -13)$ |
| $(O_9, -30)$ | $(O_8, -18)$ | $(O_9, -16)$ |

$L(2) = -0.8$, $U(3) = 51.2$, $L(3) = -7.5$, $U(8) = L(8) = -41$, $U(9) = L(9) = -56$, $U(1) = -12.8$, and $L(1) = -39.2$. Therefore, the candidate set $S = \{5, 4, 8, 9\}$.

In the top- $|K|$ elements computing phase, no object-value pair is transmitted because all the object-value pairs in candidate set S are already known. Therefore, the interaction between administrator node and member nodes can be omitted and we can obtain the top- $|2|$ query results $\{(5, 69), (4, 66)\}$.

The total number of transmitted object-value pairs in our proposed unbalanced threshold based distributed top- $|K|$ query algorithm is 11 and twice interactions are required between each member node and administrator node. As for the modified TA based algorithm [38] and the modified

Input: Sparsity K , error threshold θ , sub-sensing matrix A_i , samples $y_i, i \in \{1, \dots, P\}$
 $x^{(0)} = \mathbf{0}, t = 0$;
 Computing step size μ according to (16);
Do
 All sinks compute $z_i^{(t)}$ according to (18);
 All sinks run Algorithm 1 and the administrator sink obtains $x^{(t+1)}$;
 The administrator sink distributes $x^{(t+1)}$ to other sinks;
While $\|x^{(t+1)} - x^{(t)}\|_2 < \theta$
Output: sensed data x

ALGORITHM 2: The distributed iterative hard thresholding algorithm.

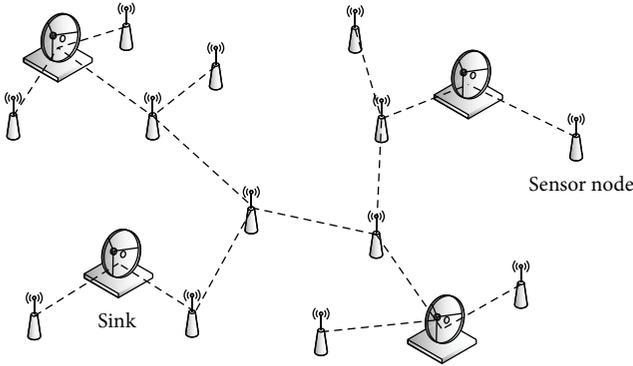


FIGURE 1: The network model.

TPUT based algorithm [39], the total number of transmitted object-value pairs is 16 and 14, respectively. Meanwhile, the interaction times of these two algorithms are 8 and 3, respectively. We can see that both the interaction times and the number of transmitted data of our proposed scheme are the least among these three algorithms. The performance comparison of these algorithms will be discussed in detail in Section 5.

4. The Distributed Iterative Hard Thresholding Algorithm

4.1. Network Model. We assume that P sinks and N sensor nodes are deployed randomly and uniformly in the surveillance area in the multisink wireless sensor network. Each sink is assigned with an ID ranging from 1 to P and each sensor node is assigned with an ID ranging from 1 to N . The sink with the lowest ID 1 is designated as administrator sink and other sinks are designated as member sinks. We assume that an effective multisink routing protocol, such as the dynamic traffic-aware routing protocol [40] or the scalable gradient-based routing protocol [41], is deployed in the network. Each sensor node can transmit its sensed data to the nearest sink in a multihop manner. The network model is shown in Figure 1.

In wireless sensor networks, the spatial correlation among sensed data in the surveillance area implies that any sensed data $x \in R^N$ at a certain time point is of K sparsity. Each sink i takes M_i samples of x using its subsensing matrix $A_i \in R^{M_i \times N}$

and gets $y_i \in R^{M_i \times 1}$; that is, $y_i = A_i x + z_i$, where $z_i \in R^{M_i \times 1}$ is noise. Hence, the set of P sinks will obtain $M = \sum_{i=1}^P M_i$ samples $y = [y_1^T, \dots, y_P^T]^T \in R^{M \times 1}$ using the global sensing matrix $A = [A_1^T, \dots, A_P^T]^T \in R^{M \times N}$. Therefore, the data collection and synchronization among multiple sinks can be ascribed to the compressive sensing problem (6). We will design a distributed iterative hard thresholding algorithm to recover the sensed data x at all sinks in a cooperative manner.

4.2. Algorithm Design. After setting the initial value of the sensed data $x^{(0)}$ and iteration time to $n \times 1$ zero vectors $\mathbf{0}$ and 0 , respectively, each sink computes the step size μ by

$$\mu = (\mu_{MN} + \sigma_{MN} F^{-1}(1 - \alpha))^{-1/2}, \quad (16)$$

where F is the cumulative distribution function of the Tracy-Widom law of order 1, $1 - \alpha$ is the quantile, and parameters μ_{MN} and σ_{MN} can be computed by

$$\mu_{MN} = \left(1 + \sqrt{\frac{(N-1)}{M}}\right)^2, \quad (17)$$

$$\sigma_{MN} = \frac{\sqrt{M} + \sqrt{N-1}}{M} \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N-1}}\right)^{1/3}.$$

According to the random matrix theory, the step size μ is tight when the entries of sensing matrix A are drawn from normal distribution $N(0, 1/M)$ [39].

After that, each sink computes the intermediate result $z_i^{(t)}$ by

$$z_i^{(t)} = \begin{cases} x^{(t)} + \mu A_i^T (y_i - A_i x^{(t)}) & i = 1 \\ \mu A_i^T (y_i - A_i x^{(t)}) & i \neq 1 \end{cases} \quad (18)$$

and executes top- $|K|$ query by running the unbalanced threshold based distributed top- $|K|$ query algorithm. The administrator sink will obtain the $(t + 1)$ th reconstructed result and distribute it to other sinks. The above iteration continues until the difference between two reconstructed results is less than the error threshold θ .

The pseudocode of the distributed iterative hard thresholding algorithm is presented in Algorithm 2.

Since the distributed top- $|K|$ query

$$H_K \left(x^{(t)} + \sum_{i=1}^P \mu A_i^T (y_i - A_i x^{(t)}) \right) = H_K \left(x^{(t)} + \mu [A_1^T, \dots, A_P^T] \left(\begin{bmatrix} y_1 \\ \vdots \\ y_P \end{bmatrix} - \begin{bmatrix} A_1 \\ \vdots \\ A_P \end{bmatrix} x^{(t)} \right) \right) \quad (19)$$

we can assure that the iterative computation in the distributed iterative hard thresholding algorithm equals that in the centralized iterative hard thresholding algorithm (7). Therefore, the recovered sensed data of these two algorithms are identical.

5. Experiment and Analysis

We evaluate the performance of our proposed unbalanced threshold based distributed data collection scheme in this section. The simulation is carried out using Matlab R2014b and Wislab simulator on a MacBook Pro laptop computer with dual core i7 CPU and 8 G memory. There are 50~200 wireless sensor nodes and 4~6 sinks randomly deployed in a $400 \times 400 \text{ m}^2$ surveillance area. Multiple uncorrelated two-dimensional Gaussian distributions have been superposed to simulate the spatial correlated data sources. Note that the Gaussian random matrix is used as the measurement matrix.

We will investigate the data reconstruction accuracy, communication complexity, and computational complexity of our proposed scheme. The signal-to-noise ratio (SNR)

$$\text{SNR} = 20 \log_{10} \left(\frac{x}{|x - \bar{x}|} \right) \quad (20)$$

is used to measure the data reconstruction accuracy, where x and \bar{x} are sensed data and reconstructed data, respectively. The number of transmitted data is the number of accumulated data involved in the communication between administrator sink and member sinks. The number of computed data is the number of accumulated data involved in the computation in administrator sink and member sinks. These two indicators are used to measure the communication complexity and computational complexity, respectively.

5.1. Performance Analysis. Figures 2, 3, and 4 show the SNR of reconstructed data, the number of transmitted data, and the number of computed data, respectively, at different data sampling rates. We can conclude that the data reconstruction performance improves with the increase on the data sampling rate. Meanwhile, the number of transmitted data and that of computed data increase too. The reason behind this rule lies in the fact that the increase on the data sampling rate would cause more data to be transmitted and computed. Therefore, the recovered sensed data will be more accurate.

Figures 5, 6, and 7 show the SNR of reconstructed data, the number of transmitted data, and the number of computed data, respectively, in different surveillance scenarios.

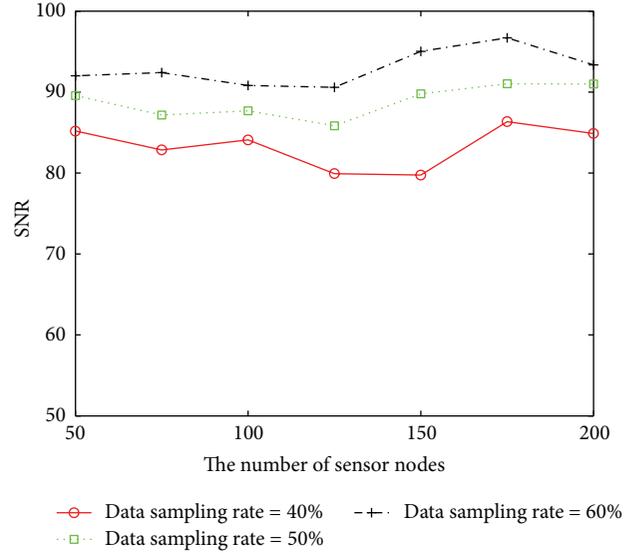


FIGURE 2: The SNR of reconstructed data at different data sampling rates.

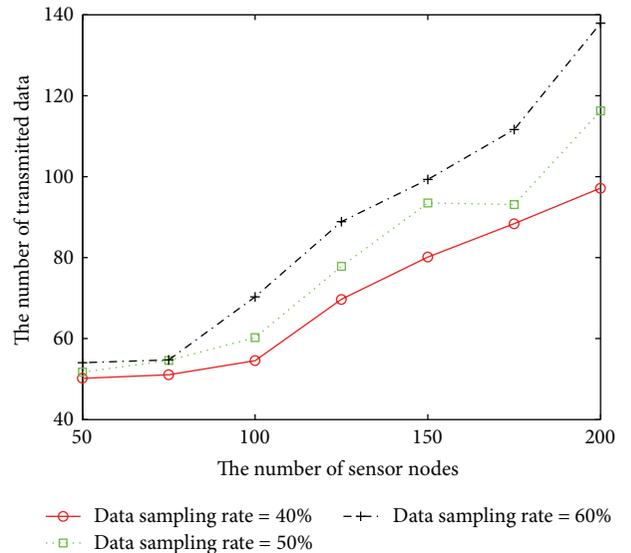


FIGURE 3: The number of transmitted data at different data sampling rates.

The number of uncorrelated two-dimensional Gaussian data sources ranges from 2 to 4. We can conclude that the data reconstruction performance degrades with the increase on the number of uncorrelated data sources. Meanwhile, the number of transmitted data and that of computed data increase at the same time. The reason behind this rule lies in the fact that the increase on the number of uncorrelated data sources leads to the increase on the sparsity of sensed data within the network. Therefore, the data reconstruction performance degrades and more data are required to be transmitted and computed.

Figures 8, 9, and 10 show the SNR of reconstructed data, the number of transmitted data, and the number of computed data, respectively, at different number of sinks.

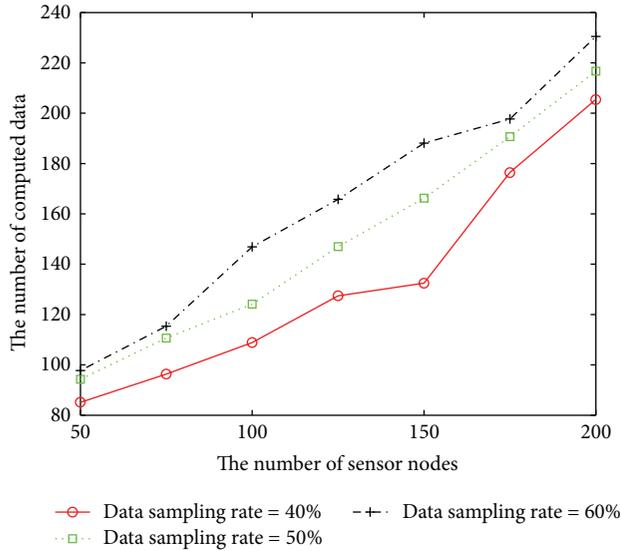


FIGURE 4: The number of computed data at different data sampling rates.

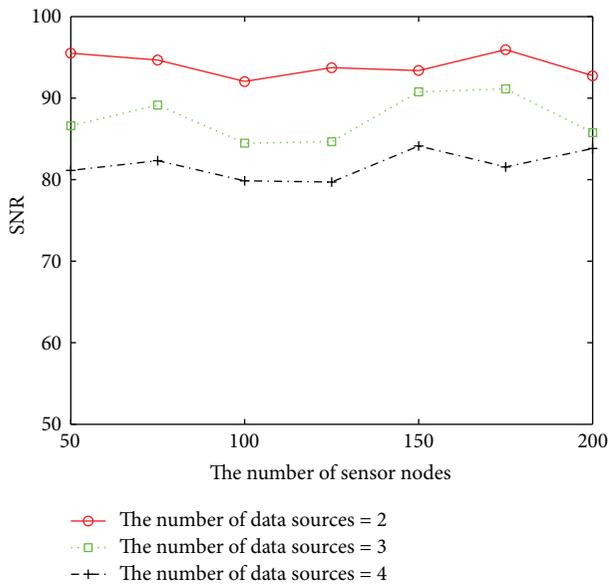


FIGURE 5: The SNR of reconstructed data at different number of data sources.

We can conclude that there are no remarkable differences among the data reconstruction performance with different number of sinks. However, the number of transmitted data and that of computed data increase with the increase of the number of sinks. The reason behind this rule lies in the fact that the increase on the number of sinks has no influence on the data reconstruction performance and only requires more data to be transmitted within the network and computed in the sinks.

5.2. Performance Comparison. We will compare the performance of our proposed scheme with the original iterative hard thresholding algorithm [25], the modified thresholding algorithm base iterative hard thresholding scheme [38],

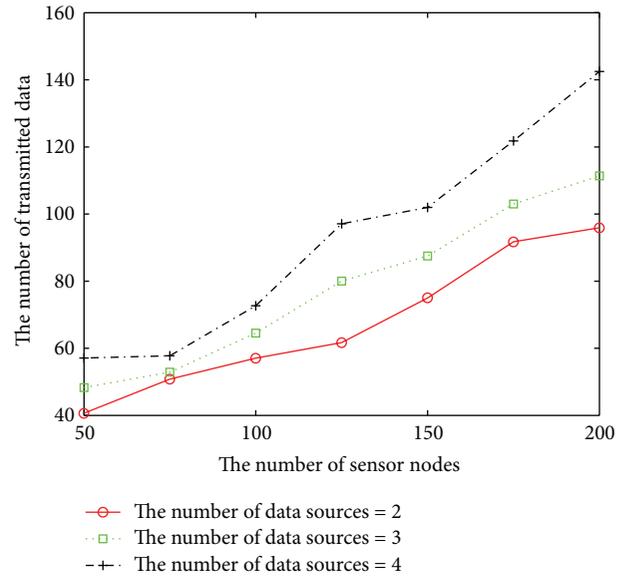


FIGURE 6: The number of transmitted data at different number of data sources.

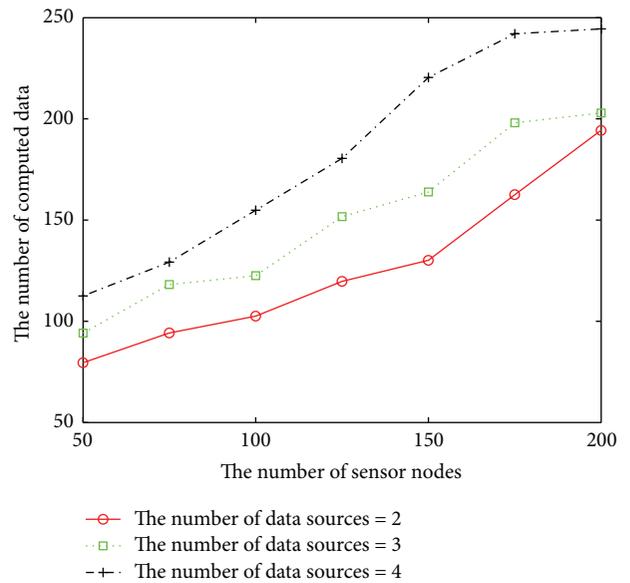


FIGURE 7: The number of computed data at different number of data sources.

and the modified three-phase uniform threshold based iterative hard thresholding scheme [39] in this subsection. Here, the last two schemes are referred to as the modified TA based DIHT scheme and the modified TPUT based DIHT scheme, respectively.

Figure 11 shows the comparison of the SNR of these four schemes when the data sampling rate is 50%, the number of data sources is 3, and the number of sinks is 5. We can conclude that there are no remarkable differences among the data reconstruction performance of these schemes. In other words, they can provide the similar data recovery accuracy regardless of the centralized or distributed reconstruction pattern.

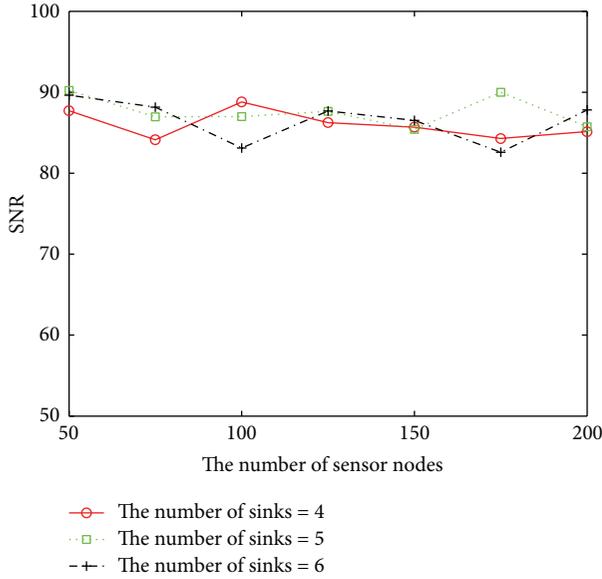


FIGURE 8: The SNR of reconstructed data at different number of sinks.

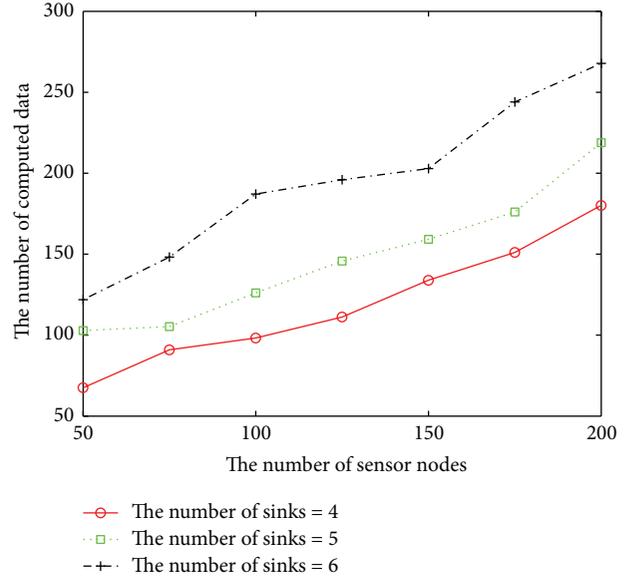


FIGURE 10: The number of computed data at different number of sinks.

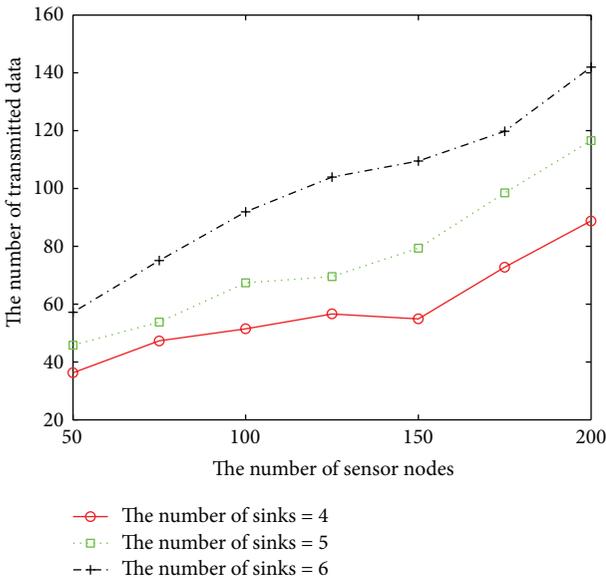


FIGURE 9: The number of transmitted data at different number of sinks.

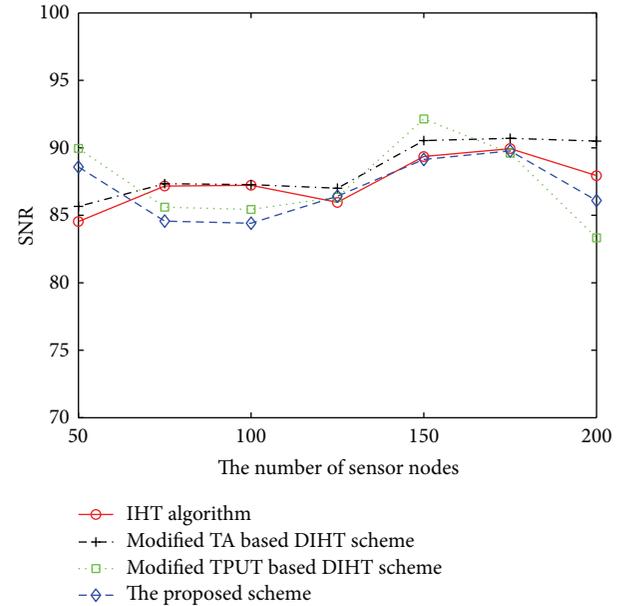


FIGURE 11: The SNR of reconstructed data of different schemes.

Figures 12 and 13 show the number of transmitted data and interaction times of these four schemes in the same experiment settings as in Figure 11. Since the iterative hard thresholding algorithm is a centralized data reconstruction algorithm, all data in the member sinks should be transmitted to the administrator sink for reconstruction and then the recovered data would be resent back to the member sinks for synchronization. Therefore, the number of transmitted data of the IHT algorithm is much higher than that of other three schemes although only one interaction is required.

The interaction times of the modified TA based DIHT scheme dominate that of other three algorithms since only one object-value pair is queried in each interaction during

the top- $|K|$ query process. Meanwhile, the number of transmitted data is relatively high since no efficient data pruning techniques are equipped in this algorithm.

Our proposed unbalanced threshold based distributed data collection scheme requires the fewest data transmission among these four schemes. By adopting unbalanced thresholds, it can adjust the query thresholds adaptively and avoid redundant transmissions between administrator sink and member sinks. Sometimes, all elements in the candidate set S can be obtained in advance in phase II instead of in phase III since reasonable thresholds are computed in the unbalanced threshold based distributed top- $|K|$ query algorithm. The example illustrated in Section 3.2 has shown this property.

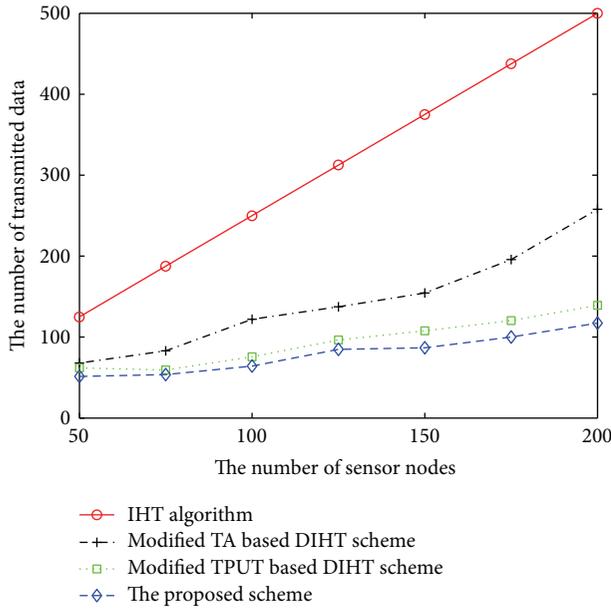


FIGURE 12: The number of transmitted data of different schemes.

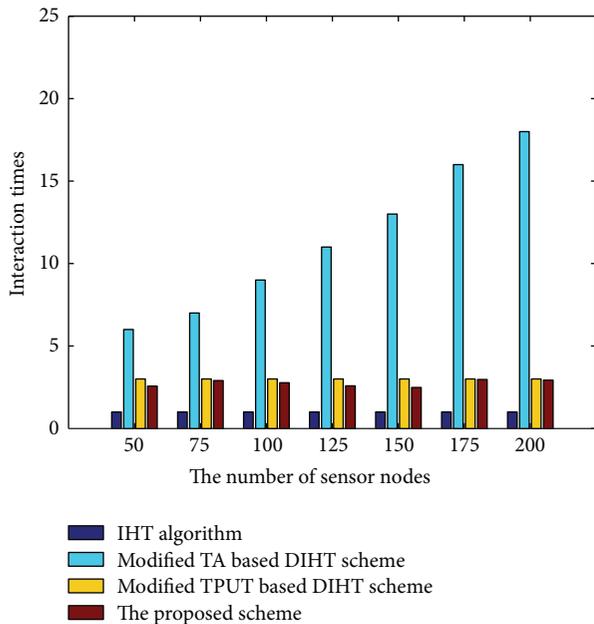


FIGURE 13: The iteration times of different schemes.

Therefore, the interaction times of our proposed scheme are slightly less than that of the modified TPUT based DIHT scheme. The price for the decrease in the number of transmitted data and interaction times is only the computation of unbalanced thresholds, which can be ignored by comparing with the communication energy consumption in wireless sensor nodes.

6. Conclusion

In this paper, we proposed the unbalanced threshold based distributed data collection scheme designed for multisink

wireless sensor networks. All sinks can obtain the synchronized sensed data of the whole network through distributed sparse signal reconstruction. Meanwhile, we also designed a distributed top- $|K|$ query algorithm to reduce the number of transmitted data between administrator sink and member sinks by pruning unnecessary elements exchanging.

The data reconstruction accuracy, communication complexity, and computational complexity of the proposed scheme were analyzed in detail through experiments. Furthermore, we also compare the performance of the proposed scheme with existing centralized and distributed data collection schemes in wireless sensor networks. In the future, we would like to design efficient data collection schemes by utilizing the spatial and temporal correlations among sensed data simultaneously.

Conflict of Interests

The authors declare that they have no competing interests.

Acknowledgments

The work in this paper has been supported by the National Natural Science Foundation of China (Grant nos. 61402094, 61300195), the China Scholarship Council Foundation (Grant no. N201408130105), the Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20120042120009), the Natural Science Foundation of Hebei Province (Grant no. F2014501078), and the Science and Technology Support Program of Northeastern University at Qinhuangdao (Grant no. XNK201401). The authors would also like to thank Professor Wenjing Lou for her kindly supports during Dr. Guorui Li's visiting in Virginia Polytechnic Institute and State University.

References

- [1] P. Rawat, K. D. Singh, H. Chaouchi, and J. M. Bonnin, "Wireless sensor networks: a survey on recent developments and potential synergies," *Journal of Supercomputing*, vol. 68, no. 1, pp. 1–48, 2014.
- [2] R. V. Kulkarni, A. Förster, and G. K. Venayagamoorthy, "Computational intelligence in wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 1, pp. 68–96, 2011.
- [3] G. Li and Y. Wang, "Automatic ARIMA modeling-based data aggregation scheme in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, article 85, 2013.
- [4] U. Raza, A. Camera, A. L. Murphy, T. Palpanas, and G. P. Picco, "Practical data prediction for real world wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2231–2244, 2015.
- [5] H. Zhang, X. Zhang, and D. K. Sung, "Lightweight self-adapting linear prediction algorithms for wireless sensor networks," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 3050–3058, 2015.
- [6] S.-T. Cheng, J.-S. Shih, C.-L. Chou, G.-J. Horng, and C.-H. Wang, "Hierarchical distributed source coding scheme and optimal transmission scheduling for wireless sensor networks,"

- Wireless Personal Communications*, vol. 70, no. 2, pp. 847–868, 2013.
- [7] J. E. Barceló-Lladó, A. M. Pérez, and G. Seco-Granados, “Enhanced correlation estimators for distributed source coding in large wireless sensor networks,” *IEEE Sensors Journal*, vol. 12, no. 9, pp. 2799–2806, 2012.
 - [8] H. Arjmandi and F. Lahouti, “Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks,” *IEEE Sensors Journal*, vol. 11, no. 9, pp. 2094–2101, 2011.
 - [9] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
 - [10] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, “Compressive sensing: from theory to applications, a survey,” *Journal of Communications and Networks*, vol. 15, no. 5, pp. 443–456, 2013.
 - [11] K. Fyhn, T. L. Jensen, T. Larsen, and S. H. Jensen, “Compressive sensing for spread spectrum receivers,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2334–2343, 2013.
 - [12] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, “Compressive light field sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4746–4757, 2012.
 - [13] S. G. Lingala and M. Jacob, “Blind compressive sensing dynamic MRI,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.
 - [14] R. Yao and Y. Zhang, “Compressive sensing for small moving space object detection in astronomical images,” *Journal of Systems Engineering and Electronics*, vol. 23, no. 3, pp. 378–384, 2012.
 - [15] C. Luo, F. Wu, J. Sun, and C. W. Chen, “Compressive data gathering for large-scale wireless sensor networks,” in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking (MobiCom ’09)*, pp. 145–156, Beijing, China, September 2009.
 - [16] L. Xiang, J. Luo, and C. Rosenberg, “Compressed data aggregation: energy-efficient and high-fidelity data collection,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1722–1735, 2013.
 - [17] L. Xiang, J. Luo, C. Deng, A. V. Vasilakos, and W. Lin, “DECA: recovering fields of physical quantities from incomplete sensory data,” in *Proceedings of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON ’12)*, pp. 182–190, IEEE, Seoul, Republic of Korea, June 2012.
 - [18] X. Xu, R. Ansari, and A. Khokhar, “Spatio-temporal hierarchical data aggregation using compressive sensing (ST-HDACS),” in *Proceedings of the 11th International Conference on Distributed Computing in Sensor Systems (DCOSS ’15)*, pp. 91–97, IEEE, Fortaleza, Brazil, June 2015.
 - [19] Y. Tang, B. Zhang, T. Jing, D. Wu, and X. Cheng, “Robust compressive data gathering in wireless sensor networks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2754–2761, 2013.
 - [20] J. Xiong and Q. Tang, “1-Bit compressive data gathering for wireless sensor networks,” *Journal of Sensors*, vol. 2014, Article ID 805423, 8 pages, 2014.
 - [21] X. Wu, Y. Xiong, P. Yang, S. Wan, and W. Huang, “Sparsest random scheduling for compressive data gathering in wireless sensor networks,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5867–5877, 2014.
 - [22] L. Kong, X. Liu, M. Tao et al., “Resource-efficient data gathering in sensor networks for environment reconstruction,” *The Computer Journal*, vol. 58, no. 6, pp. 1330–1343, 2015.
 - [23] C. Zhao, W. Zhang, X. Yang, Y. Yang, and Y.-Q. Song, “A novel compressive sensing based data aggregation scheme for wireless sensor networks,” in *Proceedings of the 1st IEEE International Conference on Communications (ICC ’14)*, pp. 18–23, IEEE, Sydney, Australia, June 2014.
 - [24] A. Salim and W. Osamy, “Distributed multi chain compressive sensing based routing algorithm for wireless sensor networks,” *Wireless Networks*, vol. 21, no. 4, pp. 1379–1390, 2015.
 - [25] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
 - [26] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
 - [27] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
 - [28] R. Tibshirani, “Regression shrinkage and selection via the lasso: a retrospective,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 73, no. 3, pp. 273–282, 2011.
 - [29] K. Fountoulakis, J. Gondzio, and P. Zhlobich, “Matrix-free interior point method for compressed sensing problems,” *Mathematical Programming Computation*, vol. 6, no. 1, pp. 1–31, 2014.
 - [30] M. S. Asif and J. Romberg, “Sparse recovery of streaming signals using l_1 -homotopy,” *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4209–4223, 2014.
 - [31] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
 - [32] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, 2010.
 - [33] D. Needell and J. A. Tropp, “CoSaMP: iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
 - [34] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
 - [35] K. Bredies and D. A. Lorenz, “Linear convergence of iterative soft-thresholding,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5–6, pp. 813–837, 2008.
 - [36] R. Fagin, A. Lotem, and M. Naor, “Optimal aggregation algorithms for middleware,” *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 614–656, 2003.
 - [37] P. Cao and Z. Wang, “Efficient top-K query calculation in distributed networks,” in *Proceedings of the 23rd Annual ACM Symposium on Principles of Distributed Computing (PODC ’04)*, pp. 206–215, ACM Press, Newfoundland, Canada, July 2004.
 - [38] S. Patterson, Y. C. Eldar, and I. Keidar, “Distributed sparse signal recovery for sensor networks,” in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’13)*, pp. 4494–4498, IEEE, Vancouver, Canada, May 2013.
 - [39] P. Han, R. Niu, and Y. C. Eldar, “Modified distributed iterative hard thresholding,” in *Proceedings of the 40th IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP '15), pp. 3766–3770, IEEE Press, South Brisbane, Australia, April 2015.

- [40] D. D. Tan and D.-S. Kim, “Dynamic traffic-aware routing algorithm for multi-sink wireless sensor networks,” *Wireless Networks*, vol. 20, no. 6, pp. 1239–1250, 2014.
- [41] H. Yoo, M. Shim, and D. Kim, “A scalable multi-sink gradient-based routing protocol for traffic load balancing,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, article 85, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

