Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 1763416, 9 pages
http://dx.doi.org/10.1155/2016/1763416

Research Article

A Perturbed Compressed Sensing Protocol for Crowd Sensing

Zijian Zhang, Chengcheng Jin, Meng Li, and Liehuang Zhu

Beijing Engineering Research Center of Massive Language Information Processing and Cloud Computing Application, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Liehuang Zhu; liehuangz@bit.edu.cn

Received 10 December 2015; Revised 28 April 2016; Accepted 10 May 2016

Academic Editor: Tony T. Luo

Crowd sensing network is a data-centric network consisting of many participants uploading environmental data by smart mobile devices or predeployed sensors; however, concerns about communication complexity and data confidentiality arise in real application. Recently, Compressed Sensing (CS) is a booming theory which employs nonadaptive linear projections to reduce data quantity and then reconstructs the original signal. Unfortunately, privacy issues induced by untrusted network still remain to be unsettled practically. In this paper, we consider crowd sensing using CS in wireless sensor network (WSN) as the application scenario and propose a data collection protocol called perturbed compressed sensing protocol (PCSP) to preserve data confidentiality as well as its practicality. At first, we briefly introduce the CS theory and three factors correlated with reconstruction effect. Secondly, a secure CS-based framework using a secret disturbance is developed to protect raw data in WSN, in which each node collects, encrypts, measures, and transmits the sampled data in our protocol. Formally, we prove that our protocol is CPA-secure on the basis of a theorem. Finally, evaluation on real and simulative datasets shows that our protocol could not only achieve higher efficiency than related algorithms but also protect signal's confidentiality.

1. Introduction

Crowd sensing network is a powerful sensor network utilizing the force from crowd. Crowd sensing is a form of network wireless sensing, which can be achieved by exploiting WSN. With enormous sensors deployed, WSN is limited by its relatively weak computational capability and low energy reservation. The primary task of WSN is to sense, transmit, and process packets while maintaining the energy cost to the minimum.

In traditional WSN, where communication is conducted via intranet or private network, bandwidth is severely consumed and certain commands from sensor nodes cannot be timely relayed to information server because great amounts of data collected during collection phase need to be transmitted. On the other hand, since trust management is maintained in public network, data confidentiality may be exposed. Hence, how to reasonably design secure transmission schemes in WSN has become a precondition for applying WSN to many fields extensively.

Without the traditional signal acquisition process constraint, Compressed Sensing (CS), proposed by Candes et al. [1] and Donoho [2] in 2006, is a booming theory that captures and represents compressible signals at a sampling rate significantly lower than the Nyquist rate [3–6]. It first employs nonadaptive linear projections that preserve the structure of the signal, and then the signal reconstruction can be conducted using an optimization process from these projections. Compressive sensing has a wide range of applications such as compressive detection and estimation, DNA microarray, and distributed compressed video sending [7].

Moreover, traditional data compressing method of WSN comes with several disadvantages, including the following. (1) Several important components and corresponding locations need to be preserved after orthogonal transformation in data compressing; otherwise, the original data could not be recovered [7]. (2) In layered multihop WSN, owing to the hardware limitation, sensors' energy storage is constrained to a low level. Intuitively, nodes closer to sink node will die sooner thanks to their faster battery consumption rate, which would result in the imbalance of energy consumption among sensors in different positions. Due to the advantages of CS, more and more CS techniques have been integrated into WSN, but most of them only consider the time relativity

of a single node. In fact, space relativity can also be traced in nodes of WSN, leading to Distributed Compressed Sensing (DCS) which views the raw data as original signal and compress the signal before transmitting. DCS has advantages as follows. (1) The random measurement from DCS is a random linear combination of every element in original signal. Thus, losing part of measurement will not affect the reconstruction of original signal. (2) In DCS-based WSN model, data quantity of each node remains the same, so energy consumption is balanced and network lifetime is prolonged.

Although DCS can effectively solve the problems raised by traditional methods, data security can never be overlooked. Researches on CS security still need to be explored. Some [8–11] tried to modify the measurement matrix but failed to apply their schemes in WSN; others [12] performed encryption (like AES, etc.) after the data is compressed to protect data security, but secure network is required. Notice that most WSN is deployed in remote, unattended, or even hostile environment, meaning node's reliability is difficult to guarantee. Therefore, it is crucial to design a secure model. In this paper, we propose a perturbed compressed sensing protocol (PCSP) to preserve data confidentiality with high practicality. Our contributions are listed as follows.

(i) We propose a perturbed compressed sensing protocol (PCSP) in WSN for crowding sensing and our PCSP can reduce communication complexity explicitly.

(ii) We prove that our PCSP can provide data confidentiality; to be more specific, our PCSP is proved to be chosen-plaintext attack secure.

(iii) We systematically evaluate our PCSP by comparing its performance with existing approaches. Experiments show that our PCSP achieves higher accuracy of recovery.

Organization. The rest of this paper is organized as follows. In Section 2, we review the related work presented in the literature. Then, we briefly introduce the main idea of CS in Section 3. Section 4 illustrates our protocol in detail. While security is discussed in Section 5. We systematically evaluate performance of PCSP by making comparisons with existing approaches in Section 6; in addition, limitations of our protocol and future work are explained in Section 7. At last, we conclude this paper in Section 8.

2. Related Work

Compressed Sensing (CS) is a new method for compressing signal which breaks through the traditional limit of sampling frequency. Through matrix computation at the encoding end, we can compress the original signal from high dimension to low dimension with a small sampling frequency and low computation complexity. At the decoding end, the original signal is reconstructed by solving a convex optimization problem.

Meanwhile, CS is capable of providing a good encryption feature on its interior structure level. Because the projection is a function value of measurement matrix which can be seen as a shared key between encoding end and decoding end.

Researches on CS put focus upon three factors associated with the reconstruction effect: sparse representation, measurement matrix, and reconstruction algorithm improvement. As a precondition for applying CS, common methods for sparse representation are discrete cosine transform basis, fast Fourier transform basis, disperse wavelet transform basis, Curvelet basis, Gabor basis, and redundant dictionary [15]. In particular, redundant dictionary or overcomplete dictionary can adaptively find out the optimal base according to the sparse property of different signal such that the minimum sparsity on this base and the best signal compression degree are both reached. For measurement matrix, Null Space Property (NSP) [16] and Restricted Isometry Property (RIP) [1, 17–19] should be satisfied; these matrixes include Gauss random matrix, Bernoulli measurement matrix, sparse stochastic matrix, toeplitz matrix, and circulant matrix. The work in [1, 2, 15, 20] proved that measurement matrix making up of independent and identical distributed Gauss random variable is irrelevant with any overcomplete redundant dictionaries, and accurate recovery of original signal can be guaranteed even after the signal is compressed. Hence, Gauss random matrix is one of the best options for measurement matrix, but doing so brings high complexity and pseudorandom matrix is an alternative choice in researches. In recent years, researchers have been working on robust pursuit algorithm, such as greedy pursuit (including MP [21], OMP [22], StOMP [23], and ROMP [24]), convex relaxed approach (including BP [25], interior point method [26], gradient projection method [27], and iterative threshold method [28]), and the combination of the former two (including Fourier sampling [29] and HHS [30]).

The classic OMP [22] is a greedy pursuit, the basic idea is transvection computation, and the most related (to compressed value Y) column vector is selected in each iteration, until the reconstruction sparse representation of original signal is found. Then we can retrieve original signal through spares inverse operation and decryption. Its advantage is convenient implementation, whereas the disadvantage is that multiple measurements are required.

As long as CS is proposed, how to use CS to provide data security is also a research hotspot. The work in [30–33] pointed out that the linear projection on measurement matrix is essentially a protection of data secrecy to some extent. The work in [30] analyzed the security of CS under several possible attacks. The work in [31] compared CS with other encryption methods through quantization. The work in [32, 33] designed the measurement matrix as symmetric secret keys such that eavesdroppers cannot obtain original signal. The work in [12] adopted AES and SHA to provide data confidentiality and data integrity after data compression.

Regarding the security problem raised by applying CS to WSN, this paper proposes an encryption method based on existing DCS model. Analysis and experiments show that our approach can provide data confidentiality with high accuracy.

3. Preliminary

First, let us take a review at the basic principles of CS. CS theory suggests that N-dimension original signal X can be

linearly projected into $M \times 1$ matrix by $M \times N$ measurement matrix Φ. If using some orthogonal basis or atomic set Ψ, such as Gabor basis and redundant dictionary [15], which is used in our frame, X can be interpreted as a vector $\theta \in R^N$ with only k nonzero elements which means

$$X = \Psi\theta = \sum_{i=1}^{N} \psi_i \theta_i. \tag{1}$$

We call X k-sparse and the solution to equation above sparse representation or sparse decomposition. To further explain (1), we have

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1N} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N1} & \psi_{N2} & \cdots & \psi_{NN} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix} \tag{2}$$

and (2) can be inferred by substituting $x_i = \psi_{i1}\theta_1 + \psi_{i2}\theta_2 + \cdots + \psi_{iN}\theta_N$, so we have

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \psi_{11} \\ \psi_{21} \\ \vdots \\ \psi_{N1} \end{bmatrix} \theta_1 + \begin{bmatrix} \psi_{12} \\ \psi_{21} \\ \vdots \\ \psi_{N2} \end{bmatrix} \theta_2 + \cdots + \begin{bmatrix} \psi_{1N} \\ \psi_{2N} \\ \vdots \\ \psi_{NN} \end{bmatrix} \theta_N. \tag{3}$$

Then X can be projected on $M \times N$ measurement matric Φ to obtain $M \times 1$ vector:

$$Y = \Phi X = \Phi\Psi\theta = AX, \tag{4}$$

where $A = \Phi\Psi$ is the sensing matrix. Meanwhile, the measurement matrix requires satisfying NSP and RIP. In [18, 34], Gauss random matrix is proved to be appropriate, so it is used in our protocol to measure signal. Then the M-dimension projection is transmitted to receiver for recovering original signal. As introduced in Section 2, in CS field, OMP algorithm is a classical recovery algorithm, which can obtain the sparsity coefficient of data. Therefore, our recovery algorithm is based on OMP algorithm. To further study it, OMP algorithm is described in Algorithm 1.

Input: compressed signal Y, sensing matrix A;
and signal sparsity k:
begin
(1) $\text{res}_0 = Y$, $\Lambda_0 = \emptyset$, $A_0 = \emptyset$, $t = 1$;
(2) **while** $t \le k$ **do**
$\lambda_t = \arg\max |\langle r_{t-1}, a_j \rangle|$, $j \in [1, 2, \ldots, N]$;
$\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$, $A_t = A_{t-1} \cup a_\lambda$;
$Y = A_t\theta_t$, compute the least square solution:
$\widehat{\theta}_t = \arg\min \|Y - At\theta t\| = (A_t^T A_t)^{-1} A_t^T Y$;
$\text{res} = Y - A_t\widehat{\theta}_t = Y - (A_t^T A_t)^{-1} A_t^T Y$;
$t = t + 1$;
(3) **Output** $\widehat{\theta}$;
end

Algorithm 1: Orthogonal matching pursuit (OMP).

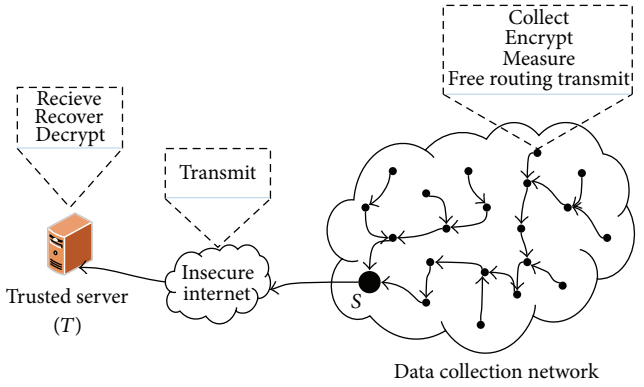


Figure 1: Overview of WSN.

4. Perturbed Compressed Sensing Protocol (PCSP)

4.1. Network Assumption. For simplicity, we denote smart device and sensor as node in the rest of the paper. Also we assume a general multihop network with n nodes and N alive, a sink node S, and a trusted server T. The overview of WSN is shown in Figure 1. Each node is required to register with the trusted registration authority RA to share a secret key with T. Nodes can collect environmental information such as temperature, humidity, and pressure. They can also receive node information from last hop node and forward node information to the next hop node. T can compute each node's corresponding measurement coefficient matrix as sensing matrix for reconstruction.

Each alive node generates a packet. As packets travel towards S, our protocol allows each node to choose the nearest node whose distance to S is smaller as the next hop node to forward the packet. The category of collected information is distinguished by the network layer data packets. The format of a packet (8 bytes) is shown in Figure 2. Where ID is the ID number of current node. Flag represents the category of collected information by nodes (1 is temperature, while 2 indicates humidity, and 3 represents pressure; also, we use 4 and 5 to denote light and salt, resp.). The value of collected environmental information can be read from Value. For ID List, it is NULL if the node is a leaf node; otherwise, we use received ID List with ID of current node appended as the ID List. The number of node information gathering round is stored in Round Number. The checksum of all bits is written in Check. Due to the fact that framework of this paper is independent with network layer protocol, the data is abstracted to pure digital signal in our following discussion.

4.2. Adversarial Model. We consider a setting with a polynomially bounded adversary capable of controlling a certain number of nodes completely. Once the adversary compromises a node, it can obtain all the node's secret keys and

	10 bits	3 bits	8 bits	17 bits	9 bits	1 bit
Packet:	ID	Flag	Value	ID list	Round number	Check

Figure 2: Format of data packet.

modify, forge, or discard messages or simply transmit false aggregation results, and its goal is to launch stealthy attacks [35] where the attacker's goal is to make T accept false aggregation results while not being detected.

4.3. PCSP. We assume that the final result X sensed by nodes is $N \times 1$ matrix. Disturbances $f(k_i, r)$ $(i = 1, 2, \ldots, N)$ are added to correlative element of X to ensure confidentiality, where r is the number of the round. Each node encrypts its sensory data x_i to $\mathrm{Enc}(x_i)$ which is transformed into linear projection Y_i on measurement matrix Φ_i. From the perspective of the whole network, the raw data X is changed to encrypted data $\mathrm{Enc}(X)$, which is transformed into compressed data Y. When final projection Y arrives at T through Internet from S, a perturbed orthogonal matching pursuit algorithm (POMP) is performed to recover the data $\mathrm{Enc}(\widehat{X})$, and then T should decrypt it to obtain original data $\widehat{X}$. Data transformation based on PCSP is shown in Figure 3. Our protocol can be divided into two major components expounded as follows.

4.3.1. Data Compression and Encryption during Free Routing. Before sensing from nodes, the trusted server should do some preparing work, as shown in Algorithm 2.

For node i, its task is to collect raw data, compute linear projection on measurement matrix, and forward message, which are described as follows.

In round r, i first senses raw data (like temperature) x_i and encrypts x_i to ciphertext:

$$\mathrm{Enc}\left(x_i\right) = x_i + f\left(k_i, r\right), \tag{5}$$

where k_i is the secret key of i and f is a hash function. We can see $\mathrm{Enc}(X)$ as

$$\mathrm{Enc}\left(X\right) = \left[\mathrm{Enc}\left(x_1\right), \mathrm{Enc}\left(x_2\right), \ldots, \mathrm{Enc}\left(x_N\right)\right]^T. \tag{6}$$

Then i computes its corresponding measurement coefficient matrix:

$$\Phi_i = \left[\phi_{1i}, \phi_{2i}, \ldots, \phi_{Mi}\right]^T \tag{7}$$

which is ith column in measurement matrix Φ. At last, i forwards signal (message):

$$Y_i = \Phi_i \times \mathrm{Enc}\left(x_i\right) \tag{8}$$

to the next node.

After receiving message Y_i, node j (using the same method to obtain Y_j) only needs to add its measurement Y_j to Y_i and sends the result to next hop until the last one sends data to S. The final compressed data Y ($M \times 1$ matrix) is transmitted to T through unsafe Internet, where

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1} & \phi_{M2} & \cdots & \phi_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}. \tag{9}$$

4.3.2. Data Recovery and Decryption Algorithm. For T, when compressed signal Y is received, it first computes sensing matrix A and utilizes POMP algorithm (see details in Algorithm 3) to reconstruct $\widehat{\theta}$, which is the sparse representation of $\mathrm{Enc}(\widehat{X})$, thereby $\mathrm{Enc}(\widehat{X})$ can be computed as

$$\mathrm{Enc}\left(\widehat{X}\right) = \Psi\widehat{\theta}. \tag{10}$$

Original data $\widehat{X}$ can be recovered by decrypting $\mathrm{Enc}(\widehat{X})$ employing the shared key between nodes and T.

Input: length N, key generation algorithm *keyGen*, original signal X:
begin
(1) round number $r = 1$;
(2) **for** $i \leftarrow 1$ to N **do**
$k_i = \mathrm{keyGen}(i)$;
distribute k_i to node i;
(3) construct Gabor dictionary parameter group
$\langle s, N/2, w, v\rangle$;
(4) residual $\mathrm{res}_d = X$;
(5) **while** $\mathrm{res}_d > threshold$ **do**
$i = 1$;
$\mathrm{res}_i = \mathrm{res}_d$;
while *no* **do**
search with res_d and compute optimal subgroup: $\langle s, w, v\rangle$;
if $\langle s, w, v\rangle$ *has been chosen* **then**
acquire corresponding atomic dictionary;
else
generate new atomic dictionary;
search to find the optimal parameter u;
remove the chosen atoms in subgroup;
store the corresponding parameters;
orthogonal projection
$Pv = \Psi * (\Psi' * \Psi)\hat{}(-1) * \Psi' * \mathrm{res}_i$;
$\mathrm{res}_{i+1} = \mathrm{res}_i - Pv$;
$i++$;
$\Psi = \{g_{ri}(n)\}$;
orthogonal projection
$Pv = \Psi * (\Psi' * \Psi)\hat{}(-1) * \Psi' * \mathrm{res}_d$;
$\mathrm{res}_d = \mathrm{res}_d - Pv$;
(6) sparse representation
$S = (\Psi' * \Psi)\hat{}(-1) * \Psi' * X$;
(7) **Output** sparse matrix Ψ and sparsity k;
end

Algorithm 2: Initialization algorithm.

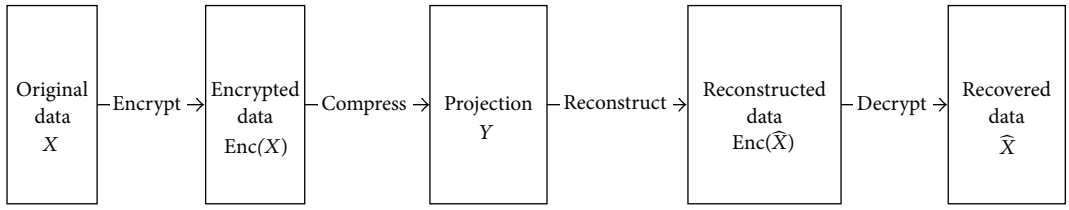


Figure 3: Data transformation based on PCSP.

Input: compressed signal Y, measurement matrix Φ, sparse matrix Ψ, key K, round number r and signal sparsity k:
begin
(1) sensing matrix $A = \Phi\Psi$;
(2) $\widehat{\theta} = \text{OMP}(Y, A, k)$;
(3) $\text{Enc}(\widehat{X}) = \Psi\widehat{\theta}$;
(4) **for** $i \leftarrow 1$ to length(Enc($\widehat{X}$)) **do**
$\widehat{X}_i = \text{Enc}(\widehat{X}_i) - f(k_i, r)$;
Output $\widehat{X}$;
end

Algorithm 3: Perturbed orthogonal matching pursuit (POMP).

5. Security Analysis

Adversaries can compromise a fraction of nodes in sensor network. After a node i is compromised, its private information such as secret key k_i and ID will be leaked to adversary who can launch stealthy attack to make T accept false data without being detected.

We consider the situation where the adversary is trying to forge a valid $\text{Enc}(x_i)$ without the knowledge of k_i. Apparently, the possibility relies on the pseudorandomness of the hash function f we chose and we believe the probability of generating an authentic $\text{Enc}(x_i)$ is approximately $1/2^N$. Formally, our protocol is proved to be a chosen-plaintext attack secure based on Theorem 1.

Theorem 1. *If F is a pseudorandom function, the PCSP scheme is secure under a chosen-plaintext attack.*

Proof. Assume that f is a random function. We construct a new scheme which is exactly same as PCSP scheme, except that the pseudorandom function F is replaced by f. Since f is a random function, the probability that the adversary chooses the correct plaintext from the challenge cipher text is exactly $1/2$.

Now we consider the PCSP scheme in the chosen-plaintext attack. Here we define the probability that the adversary wins the chosen-plaintext attack: that is, $1/2 + \epsilon(n)$, where n is the security parameter. We then construct a distinguisher D to distinguish F and f as below: D runs the adversary to attack PCSP scheme under chosen-plaintext attack experiment.

(1) When a message x needs to be encrypted, D sends the adversary $x + F(k, r)$.

(2) When two plaintexts m_0 and m_1 are received, D flips a coin i, ($i = 0$ or 1), and sends the adversary $x_i + g(k, r)$. Here g is one of pseudorandom functions or random functions.

(3) When the output j of the adversary is received, D outputs $g = F$ if the adversary wins; otherwise, D outputs $g = f$.

From the viewpoint of D, if $g = F$, the probability that the adversary wins is $1/2 + \epsilon(n)$. Otherwise, the probability that the adversary wins is $1/2$, since the challenge cipher text is a random number. Therefore, the probability that D wins is $\epsilon(n)$. Finally, $\epsilon(n)$ must be negligible. □

6. Evaluation

In this section, we attempt to present the performance evaluation results on the real and simulative datasets. To evaluate the efficiency of our protocol, we follow the estimation error used in [36] to compare the accuracy among PCSP and three related algorithms (see details in Experiment 1). Later, we conduct simulation experiments with encryption/decryption and then encryption/decryption is removed in Experiment 2 for proving that our proposed protocol is effective to protect the confidentiality of data while preserving accuracy (as shown in Experiment 2).

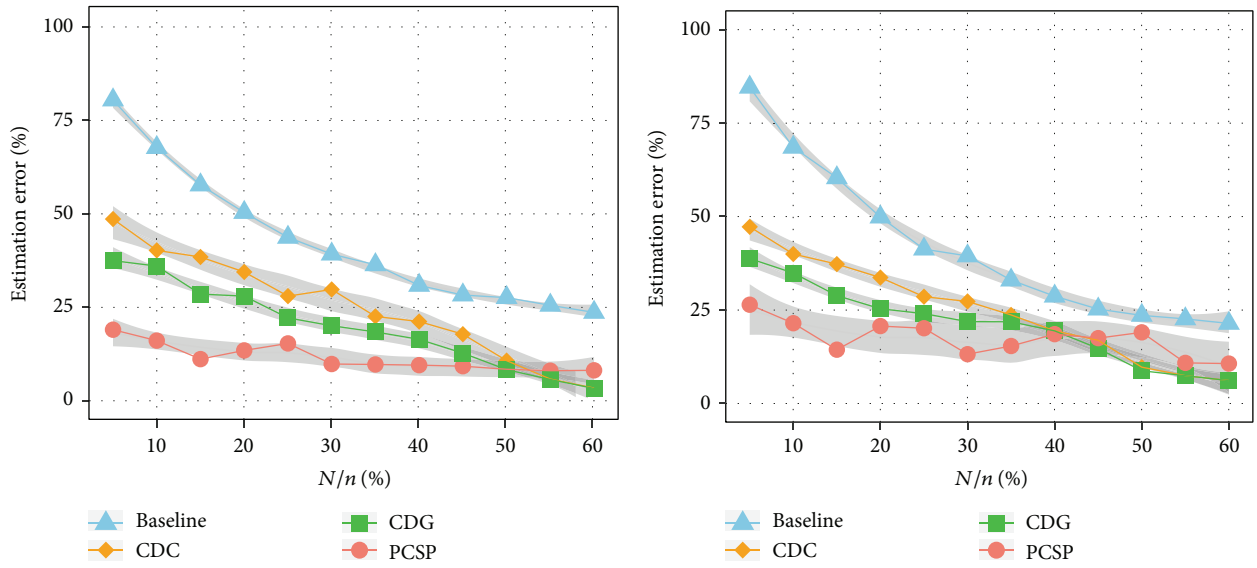
Experiment 1 (comparison with related algorithms on real datasets). Datasets used in this experiment contain NBDC-CTD [14] and InelLab [13], of which attributes are summarized in Table 1. We investigate performance of our method compared with the following state-of-art methods.

(1) *Baseline*. This algorithm uses basic routing and estimation methods, which is seen as baseline in [36]. Sensor node transmits packets to S using the shortest path. When S receives the final packet, it sends the final packet to information server, which takes advantage of the k-Nearest Neighbors (kNN) [37] Algorithm to recover the data.

(2) *CDG [38]*. In this framework, the following tree-based routing and traditional methods of CS for reconstructing the data collected from WSN are used. A sensor node will not send a packet to its parent node until receiving all packets from its children, so it collects all sensor readings to a packet. Convex optimization methods are used by information server to estimate the signal.

Table 1: Datasets for Experiment 1.

Name	Time period	Environment	Physical condition
IntelLab [13]	Feb. 28–Apr. 5, 2004	Indoor	Temperature, humidity, light
NBDC-CTD [14]	Oct. 26–Oct. 28, 2012	Ocean	Temperature, humidity, salt



(a) Estimation error on NBDC-CTD

(b) Estimation error on IntelLab

Figure 4: Comparison with related algorithms on real datasets.

(3) *CDC [36]*. Opportunistic routing with compression and a NSRP-based estimator are utilized in the CDC scheme. The compression scheme adds or subtracts the reading of last hop node as the packet travels towards S. Information server employs random linear projections of the orthonormal basis to estimate the coefficient vector to recover original data, because nonuniform sparse random projections (NSRP) used in compressing can preserve inner products within a small error.

We follow a classic evaluation criterion named as estimation error [36] (EE) defined in (11) and observe the performance of our method compared with CDG, CDC, and baseline algorithms:

$$\mathrm{EE} = \frac{\left\| X - \widehat{X} \right\|_2}{\left\| \widehat{X} \right\|_2}. \tag{11}$$

We run all of these algorithms 50 times and calculate the mean of their EE, respectively. A conclusion that the estimation error of our protocol is robust to the scale of the WSN can be inferred from Figure 4. As shown in Figures 4(a) and 4(b), our PCSP outperforms the competing algorithms when the number of alive nodes is small. In particularly, PCSP achieves estimation error as low as 18.89% and 26.39% on NBDC-CTD and IntelLab whilst results of other approaches are all higher.

Experiment 2 (comparison with encrypted and unencrypted data on simulative datasets). First of all, initialization algorithm (Algorithm 2) is run to start network sparse learning on encrypted and unencrypted data. In encryption process, nodes sense and encrypt data. Then we use pseudo-random Gaussian matrix to generate measurement matrix Φ and final signal Y is arrived at T. T takes advantage of POMP algorithm to obtain $\widehat{X}$. In the process without encryption, nodes just sense data. Then we make use of the measurement matrix Φ generated in encryption process, and then Y arrives at T. Later on, T runs OMP algorithm to reconstruct original signal $\widehat{X}$. If round number r is bigger than the threshold, then reinitialize the whole network. The parameters of two experiments are listed in Table 2.

To estimate the performance of our method compared with unencrypted data method, we employ EE mentioned in Experiment 1 and another criterion E defined in

$$E = \frac{\widehat{X} - X}{X}. \tag{12}$$

We conduct experiments 50 times in which the mean of EE is calculated, also original data, recovered data, encrypted data, compressed data, and estimation error EE as well as error E are recorded, as shown in Figure 5. Figure 5(a) indicates that original signal, recovered encrypted signal, and unencrypted signal keep the same trend. While Figure 5(b) presents the encrypted data, which cannot be utilized to speculate on the original data. Compressed result of encrypted data can be seen in Figure 5(c), whose dimensionality is lower than original data ($125 < 200$). As shown in Figure 5(d), estimation error of encrypted data has small variation with

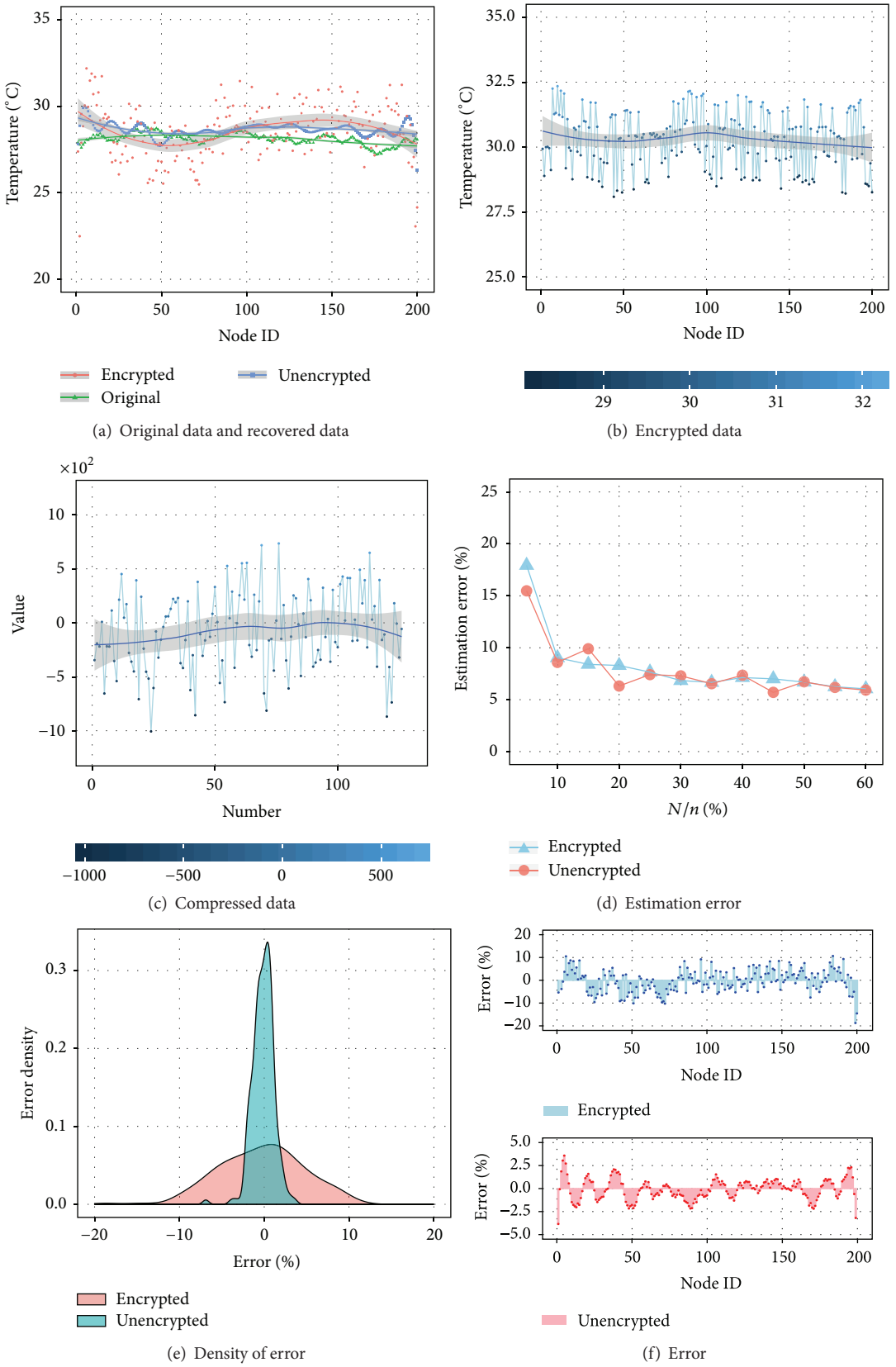


Figure 5: Comparison between encrypted data and unencrypted data.

Table 2: Parameter setting of Experiment 2.

Parameter	$N = 200$, threshold = 4
Data	Encrypting after randomly generating data $X \in [25, 30]$
Ψ	Search on Gabor overcomplete dictionary
Φ	Random Gaussian matrix, expectation = 0, variance = 1
Result of encrypted data	$k = 18$, estimation error = 25.444
Result of unencrypted data	$k = 20$, estimation error = 25.457

that of original data. We demonstrate error density of these two experiments in Figure 5(e) and details in Figure 5(f).

7. Discussion and Future Work

Crowd sensing by applying compressed sensing to WSN is an extremely complex task. Despite the fact that work done in this paper can initially perform the task with sensor energy balanced while preserving data privacy, some challenges still remain to be addressed. Firstly, network synchronization is necessary between WSN and T to obtain the number of rounds and keys for encryption/decryption. Another one is that our work only considers protecting data confidentiality rather than preserving data integrity and availability. Several improvements still need to be considered as follows. (1) The accuracy of reconstruction algorithm can be increased. (2) More security features (such as data integrity and availability) can be further studied.

8. Conclusion

In the context of crowd sensing in WSN, we proposed a perturbed compressed sensing protocol (PCSP) combined with compressed sensing technology to solve the issues about data confidentiality and sensor energy. Our protocol can be summarized into two components, in which encrypted data is obtained by perturbing sensor data gathered by each node; then, data compression by crowd sensing in WSN is enforced by linear projection utilizing compressed sensing. Afterwards, we presented performance analysis and security analysis along with experiments results which demonstrated that our protocol is capable of transmitting signal at a low energy cost while preserving data confidentiality. At last, we described limitations of our protocol with future work followed.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by National Natural Science Foundation of China no. 61272512 and no. 61300177 and Beijing Natural Science Foundation no. 4132054.

References

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

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

[3] L. Palopoli, R. Passerone, and T. Rizano, "Scalable offline optimization of industrial wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 2, pp. 328–339, 2011.

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

[5] J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, 2008.

[6] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, 2011.

[7] YZ, *Design and research on CS-based wireless sensor network spatial sparse signals network models [Ph.D. thesis]*, Nankai University, 2012.

[8] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "On the security and robustness of encryption via compressed sensing," in *Proceedings of the IEEE Military Communications Conference (MILCOM '08)*, pp. 1–7, San Diego, Calif, USA, November 2008.

[9] Y. Rachlin and R. D. Baron, "The secrecy of compressed sensing measurements," in *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 813–817, IEEE, Urbana, Ill, USA, September 2008.

[10] A. M. Abdulghani and E. Rodriguez-Villegas, "Compressive sensing: from 'compressing while sampling' to 'compressing and securing while sampling'," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)*, pp. 1127–1130, Buenos Aires, Argentina, September 2010.

[11] R. Dautov and G. R. Tsouri, "Establishing secure measurement matrix for compressed sensing using wireless physical layer security," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC '13)*, pp. 354–358, San Diego, Calif, USA, January 2013.

[12] M. Zhang, M. M. Kermani, A. Raghunathan, and N. K. Jha, "Energy-efficient and secure sensor data transmission using encompression," in *Proceedings of the 26th International Conference on VLSI Design and 12th International Conference on Embedded Systems (ES '13)*, pp. 31–36, IEEE, Pune, India, January 2013.

[13] http://www.select.cs.cmu.edu/data/labapp3/index.html.

[14] http://tao.ndbc.noaa.gov/.

[15] S.-T. Li and D. Wei, "A survey on compressive sensing," *Acta Automatica Sinica*, vol. 35, no. 11, pp. 1369–1377, 2009.

[16] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[17] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, pp. 1433–1452, Madrid, Spain, August 2006.

[18] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[19] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[20] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

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

[22] 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.

[23] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[24] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[25] 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.

[26] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ1-regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[27] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[28] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[29] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse fourier representations via sampling," in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC '02)*, pp. 152–161, ACM, 2002.

[30] A. C. Gilbert, M. J. Strauss, and R. Vershynin, "One sketch for all: fast algorithms for compressed sensing," in *Proceedings of the 39th ACM Symosium on the Theory of Computing (STOC '07)*, pp. 237–246, San Diego, Calif, USA, June 2007.

[31] A. M. Abdulghani and E. Rodriguez-Villegas, "Compressive sensing: from 'compressing while sampling' to 'compressing and securing while sampling'," *Proceedings of the 32rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1127–1130, 2010.

[32] Y. Rachlin and R. D. Baron, "The secrecy of compressed sensing measurements," in *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 813–817, September 2008.

[33] R. Dautov and G. R. Tsouri, "Establishing secure measurement matrix for compressed sensing using wireless physical layer security," in *Proceedings of the 10th International Conference on Computing, Networking and Communications (ICNC '13)*, pp. 354–358, San Diego, Calif, USA, January 2013.

[34] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, 2006.

[35] B. Przydatek, D. Song, and A. Perrig, "SIA: secure information aggregation in sensor networks," in *Proceedings of the 1st ACM International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pp. 255–265, November 2003.

[36] X.-Y. Liu, Y. Zhu, L. Kong et al., "CDC: compressive data collection for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2188–2197, 2015.

[37] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[38] 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, ACM, Beijing, China, September 2009.