

# Data Aggregation and Recovery for the Internet of Things: A Compressive Demixing Approach

Evangelos Zimos<sup>1</sup>, João F. C. Mota<sup>2</sup>, Evaggelia Tsiligianni<sup>1</sup>, Miguel R. D. Rodrigues<sup>3</sup>, and Nikos Deligiannis<sup>1</sup>

<sup>1</sup>Dept. Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium  
imec, Kapeldreef 75, B-3001 Leuven, Belgium  
{ezimos, etsiligi, ndeligia}@etrovub.be

<sup>2</sup>Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, UK  
j.mota@hw.ac.uk

<sup>3</sup>Dept. Electronic and Electrical Engineering, University College London, London, UK  
m.rodrigues@ucl.ac.uk

**Abstract**—Large-scale wireless sensor networks (WSNs) and Internet-of-Things (IoT) applications involve diverse sensing devices collecting and transmitting massive amounts of heterogeneous data. In this paper, we propose a novel compressive data aggregation and recovery mechanism that reduces the global communication cost without introducing computational overhead at the network nodes. Following the principles of compressive demixing, each node of the network collects measurement readings from multiple sources and mixes them with readings from other nodes into a single low-dimensional measurement vector, which is then relayed to other nodes; the constituent signals are recovered at the sink using convex optimization. Our design achieves significant reduction in the overall network data rates compared to prior schemes based on (distributed) compressed sensing or compressed sensing with (multiple) side information. Experiments using real large-scale air-quality data demonstrate the superior performance of the proposed framework against state-of-the-art solutions, with and without the presence of measurement and transmission noise.

**Index Terms**—Compressive demixing, wireless sensor networks, Internet of things, air-pollution monitoring, smart cities.

## I. INTRODUCTION

**W**IRELESS sensor network (WSN) technology lies in the crux of various Internet-of-Things (IoT) applications [1], where myriads of wireless nodes monitor ambient conditions in an area, such as temperature, humidity, and concentration of air pollution particles. Consider a WSN comprising  $N$  nodes, each measuring  $L$  data types; for instance, the concentration of CO<sub>2</sub>, SO<sub>2</sub> and NO in the air. *How can we efficiently gather all the data from the distributed nodes and recover it at the sink?* The design needs to minimize the data rates, maintain the power consumption at the sensor nodes, and account for the effect of measurement and transmission noise. Power savings can be achieved by decreasing the encoding complexity at the node and/or by reducing the radio emission. Furthermore, in such large-scale setups, *multi-hop*

transmission over small distances is preferable so as to keep the transmission power as low as possible [1].

Prior studies on data gathering propose collaborative wavelet transform coding [2] and clustered data aggregation [3], [4] techniques that require excessive transmission of overhead information—and hence additional encoding complexity. Distributed source coding (DSC) [5] and multiterminal coding [6] provides an alternative strategy that leverages inter-sensor data correlation at the decoder side, but performs well only for a limited number of nodes. Haupt *et al.* focussed on large-scale WSNs proposing an intelligent design where compressed sensing (CS) is used to balance the power consumption of the sensing devices [7]. A similar gathering scheme that considers multi-hop routing and includes a network capacity analysis was presented in [8]. This method provides significant rate reductions compared to traditional multi-hop solutions. Both [7], [8] leverage spatial dependencies among sensor data to obtain sparse signal representations; independent recovery of single modal data is performed at the sink. Leveraging spatiotemporal compressibility, a dynamic compression scheme followed by a recovery algorithm was proposed in [9]. In [10], [11], another compressive gathering scheme based on compressed sensing with (multiple) side information [12] was introduced. This work leverages both the spatial dependencies between sensor readings of the same source, as well as the dependence structure among sensor readings of different sources. The number of transmissions  $M_l$  required for each source  $l$  is smaller than in [7], [8] as recovery accounts for multiple side information signals at the data recovery stage. Finally, distributed compressed sensing (DCS) [13] provides an alternative solution that uses spatiotemporal dependencies among sensor data to succeed joint recovery. Using a sophisticated model to describe the joint sparsity among sensor signals, DCS requires a total number of transmissions similar to [7] for high mean-squared error (MSE) and similar to [10] when improved reconstruction quality is needed.

Although these compressive gathering designs focus on improving the reconstruction quality at the sink, they actually

We acknowledge the support of the Fonds Wetenschappelijk Onderzoek (project no. G0A2617) and the VUB-Duke-UCL-UGent International Joint Research Group on Big Data.

disregard the significant amount of energy consumed due to gathering readings of different sources separately. In this paper, we follow a different approach: instead of developing another recovery algorithm, we introduce a mechanism that jointly aggregates sensor readings from all data sources into a *single* measurement vector. To this end, we use the compressive demixing paradigm [14], [15], which allows reconstructing a set of structured signals only from their sum. This enables us to perform both *joint gathering and joint recovery* of the sensor signals, and thus significantly reduce the total amount of communication. Moreover, the proposed algorithm is robust against measurement and communication noise and requires little computation from the sensor nodes.

The paper is organized as follows. Section II presents an overview of compressed sensing and compressive demixing theory, and the design for single modal data collection in [8]. Section III includes the proposed scheme for multiple source heterogeneous data collection. In Section IV, we demonstrate experimental results employing large-scale air-pollution data.

## II. BACKGROUND

### A. Compressed Sensing

Compressed Sensing (CS) builds upon the fact that many signals  $\mathbf{x} \in \mathbb{R}^N$  have sparse representations, i.e., they can be written as  $\mathbf{x} = \Psi \mathbf{s}$ , where  $\Psi \in \mathbb{R}^{N \times N}$  is an orthogonal basis, and  $\mathbf{s} \in \mathbb{R}^N$  is a  $k$ -sparse vector (it has at most  $k$  nonzero entries). Suppose we observe  $M \ll N$  linear measurements from  $\mathbf{x}$ :  $\mathbf{y} = \Phi \mathbf{x} = \mathbf{A} \mathbf{s}$ , where  $\Phi \in \mathbb{R}^{M \times N}$  is a sensing (or encoding) matrix, and  $\mathbf{A} := \Phi \Psi$ ,  $\mathbf{A} \in \mathbb{R}^{M \times N}$ . CS theory states that if  $\mathbf{A}$  satisfies the mutual coherence property [16], the Restricted Isometry Property [17], or the Null Space Property [18], then  $\mathbf{s}$  (and thus  $\mathbf{x}$ ) can be recovered by solving

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A} \mathbf{s}. \quad (1)$$

In particular,  $\mathbf{s}$  is the only solution to (1) whenever the number of measurements  $M$  is sufficiently large. When the measurements are noisy, i.e.,  $\mathbf{y} = \mathbf{A} \mathbf{s} + \boldsymbol{\eta}$ , where  $\boldsymbol{\eta} \in \mathbb{R}^M$  represents additive noise,  $\mathbf{s}$  can be recovered by solving instead

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{y} - \mathbf{A} \mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1, \quad (2)$$

where  $\lambda > 0$  controls the trade-off between sparsity and reconstruction fidelity. Instead of assuming that  $\mathbf{s}$  is strictly sparse (i.e.,  $\|\mathbf{s}\|_0 = k$ ), several works, e.g. [19] (including this one) focus on compressible signals, i.e., signals whose coefficients decay exponentially, when sorted in order of decreasing magnitude.

### B. Compressive Data Gathering for WSNs

The compressive data gathering (CDG) approach in [7], [8] adheres to a multi-hop communication scenario in which each node relays a weighted sum of sensor readings to a neighboring node. Specifically, consider a network of  $N$  nodes and let  $x_i \in \mathbb{R}$  denote a scalar reading of node  $i \in \{1, 2, \dots, N\}$ . As shown in Fig. 1, node 1 generates a

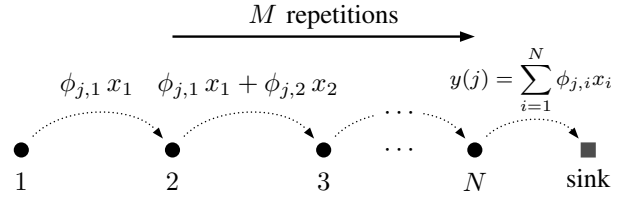


Fig. 1. Multi-hop transmission in a large-scale WSN using CS [8].

pseudorandom number  $\phi_{j,1}$  and transmits the value  $\phi_{j,1} x_1$  to node 2. Subsequently, node 2 generates  $\phi_{j,2}$ , computes the weighted sum  $\phi_{j,1} x_1 + \phi_{j,2} x_2$  and sends it to node 3. In sum, node  $n$  generates  $\phi_{j,n}$ , computes the value  $\phi_{j,n} x_n$ , adds it to the sum of the previous relayed values, and sends  $\sum_{i=1}^n \phi_{j,i} x_i$  to node  $n+1$ . The sink node thus receives  $y(j) = \sum_{i=1}^N \phi_{j,i} x_i$ . After repeating the procedure  $M$  times, for  $j = 1, \dots, M$ , the sink obtains

$$\mathbf{y} = [\phi_1 \quad \dots \quad \phi_i \quad \dots \quad \phi_N]^T \mathbf{x} = \Phi \mathbf{x}, \quad (3)$$

where  $\mathbf{y} = (y(1), \dots, y(j), \dots, y(M))^T$  is the vector of measurements,  $\phi_i = (\phi_{1,i}, \dots, \phi_{j,i}, \dots, \phi_{M,i})$  is the row vector of pseudorandom numbers generated by node  $i$ , and  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_N)^T$  is the vector of the node readings. Considering the spatial correlation of the sensor readings in a densely deployed sensor network, the source signal  $\mathbf{x}$  can be represented by a sparse signal with respect to a suitable basis. Then, the sink can recover  $\mathbf{x}$  using standard CS recovery algorithms [19]–[21].

In a multi-hop network comprising  $N$  nodes, a message is delivered to the sink after  $\mathcal{O}(N)$  transmissions. The conventional approach in which each node conveys the previous nodes' messages together with its own results in an overall communication cost of  $\mathcal{O}(N^2)$ . Employing CDG, each node sends exactly  $M$  measurements, reducing the overall communication cost to  $\mathcal{O}(NM)$ . The number of measurements  $M$  is dictated by CS theory; for an  $N$ -dimensional signal with sparsity level equal to  $\kappa$ ,  $M$  is  $\mathcal{O}(\kappa \log(N/\kappa))$  [17], [18], [22], and the overall number of transmissions is  $\mathcal{O}(N\kappa \log(N/\kappa))$ .

CDG can be extended to transmit multiple signals acquired from  $L$  different sources in two ways. We can either apply it separately for every source signal or treat the measurements at the  $i$ -th node as a vector  $\mathbf{x}_i = (x_i^{(1)}, \dots, x_i^{(L)})^T$ ,  $i \in \{1, 2, \dots, N\}$ . If we assume a similar sparsity level for all source signals, i.e.,  $\kappa_l \approx \kappa$ ,  $l \in \{1, 2, \dots, L\}$ , in both cases, the overall number of transmissions is  $\mathcal{O}(LN\kappa \log(N/\kappa))$ .

### C. Compressive Demixing

Demixing or source separation is the process of extracting multiple components entangled within a signal [14], [15], [23], [24]. According to the basic model for mixed signals, we observe a signal  $\mathbf{z} \in \mathbb{R}^N$

$$\mathbf{z} = \mathbf{x}_1 + \mathbf{x}_2, \quad (4)$$

and wish to determine the component signals  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$ . Demixing is based on prior information about the signal structures, many of which can be encoded by a convex function.

For example, assume  $\mathbf{x}_1$  (resp.  $\mathbf{x}_2$ ) is sparse in the basis  $\Psi^{(1)} \in \mathbb{R}^{N \times N}$  (resp.  $\Psi^{(2)} \in \mathbb{R}^{N \times N}$ ). Then,  $\mathbf{x}_1, \mathbf{x}_2$  can be demixed by solving

$$\begin{aligned} [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2] &= \arg \min_{\mathbf{x}_1, \mathbf{x}_2} \|\Psi^{(1)} \mathbf{x}_1\|_1 + \alpha \|\Psi^{(2)} \mathbf{x}_2\|_1 \\ \text{s.t.} \quad &\mathbf{z} = \mathbf{x}_1 + \mathbf{x}_2. \end{aligned} \quad (5)$$

Specifically, if  $\Psi^{(1)}$  and  $\Psi^{(2)}$  are incoherent in the sense that  $\Psi^{(1)}(\Psi^{(2)})^{-1}$  [or  $\Psi^{(2)}(\Psi^{(1)})^{-1}$ ] can be modeled as a uniform random orthogonal basis, and the sum of the statistical dimensions of the  $\ell_1$ -norm at  $\Psi^{(1)} \mathbf{x}_1$  and  $\Psi^{(2)} \mathbf{x}_2$  is smaller than  $N$  then, with high probability, there exists  $\alpha > 0$  such that (5) has a unique solution [15, Thm. 1], [14]. This means that  $\mathbf{x}_1$  and  $\mathbf{x}_2$  can be perfectly separated via (5).

When the mixed signal  $\mathbf{x}_1 + \mathbf{x}_2$  is partially observed, then we obtain the paradigm of compressive demixing [15], that is,

$$\mathbf{y} = \Phi(\mathbf{x}_1 + \mathbf{x}_2), \quad (6)$$

where  $\Phi \in \mathbb{R}^{M \times N}$ ,  $M \ll N$ , is a matrix representing the linear subsampling mechanism. The strategy used for demixing uncompressed signals can be easily extended to the compressed case. Assuming again that the signals have sparse representations in  $\Psi^{(1)}$  and  $\Psi^{(2)}$ , we would solve

$$\begin{aligned} [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2] &= \arg \min_{\mathbf{x}_1, \mathbf{x}_2} \|\Psi^{(1)} \mathbf{x}_1\|_1 + \alpha \|\Psi^{(2)} \mathbf{x}_2\|_1 \\ \text{s.t.} \quad &\mathbf{y} = \Phi(\mathbf{x}_1 + \mathbf{x}_2). \end{aligned} \quad (7)$$

Clearly, the above models in (4), (6) can be extended to more than two signals. The reader is referred to [14], [15], [25] and the references therein for more details.

### III. PROPOSED SCHEME

Existing compressive data gathering techniques [7], [8], [26], [27] have successfully employed CS by leveraging the spatiotemporal correlation among sensor readings collected from a single source. CS solutions are popular as they reduce transmission cost without incurring additional computation or communication control overhead over the network nodes. However, current WSN and IoT setups involve diverse sensing devices gathering heterogeneous data; for instance, different air pollution measurements (CO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>) are collected in an environmental monitoring setup. While CS is a well-suited solution to homogeneous data compression and transmission, efficient sensing of multiple heterogeneous signals naturally leads to compressive demixing. The basic model of compressively mixed signals allows the formulation of a sensing mechanism that encodes multiple signals into a single low-dimensional measurement vector, and a recovery process that extracts the constituent signals using convex optimization. Our experimental results show that the proposed scheme significantly reduces the acquired sensor data measurements and minimizes the transmission data rates, while being robust to measurement and transmission noise.

Consider a large-scale WSN comprising  $N$  wireless devices (nodes), each of which equipped with  $L$  sensors that monitor diverse, but statistically dependent, data types; for example, if the sensors measure the concentration of CO,

NO<sub>2</sub>, O<sub>3</sub>, and SO<sub>2</sub>, then  $L = 4$ . The wireless nodes form a multi-hop route to the sink, as depicted in Fig. 2. Each node  $i \in \{1, 2, \dots, N\}$  observes realizations of the correlated sources  $X_1, X_2, \dots, X_L$ , which take values in their corresponding continuous alphabets  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_L$ . Let  $x_i^{(l)}$  denote the reading at node  $i \in \{1, 2, \dots, N\}$  of data type  $l \in \{1, 2, \dots, L\}$ . We denote by  $\mathbf{x}_i = (x_i^{(1)}, \dots, x_i^{(L)})^\top$  the  $L$ -dimensional vector collecting the readings of all data types at node  $i$ , and  $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_i^{(l)}, \dots, x_N^{(l)})^\top$  the  $N$ -dimensional vector collecting the readings from the source  $X_l$  at all nodes.

#### A. Joint Data Aggregation

The proposed design encodes the sensor readings of  $L$  sources collected by  $N$  nodes in the network into a *single*  $M$ -dimensional measurement vector. More particularly, the readings of different sources are aggregated using random weights and transmitted as a single mixed measurement; the process is repeated  $M$  times until the sink receives  $M$  mixed measurements.

In the  $j$ -th transmission, the gathering procedure is initiated by node 1, which collects  $\mathbf{x}_1 = (x_1^{(1)}, \dots, x_1^{(L)})^\top$ , generates the pseudo-random numbers  $(\phi_{j,1}^{(1)}, \dots, \phi_{j,1}^{(L)})$ , and transmits the sum  $\phi_{j,1}^{(1)} x_1^{(1)} + \dots + \phi_{j,1}^{(L)} x_1^{(L)}$  to node 2. Subsequently, node 2 collects  $\mathbf{x}_2 = (x_2^{(1)}, \dots, x_2^{(L)})^\top$ , generates the pseudo-random numbers  $(\phi_{j,2}^{(1)}, \dots, \phi_{j,2}^{(L)})$  and sends the sum  $\sum_{i=1}^2 \phi_{j,i}^{(1)} x_i^{(1)} + \dots + \sum_{i=1}^2 \phi_{j,i}^{(L)} x_i^{(L)}$  to node 3. To generalize, each device  $n$  collects  $L$  readings  $\mathbf{x}_n = (x_n^{(1)}, \dots, x_n^{(L)})^\top$  that correspond to different sources, then generates the numbers  $(\phi_{j,n}^{(1)}, \dots, \phi_{j,n}^{(L)})$ , and transmits the value  $\sum_{i=1}^n \phi_{j,i}^{(1)} x_i^{(1)} + \dots + \sum_{i=1}^n \phi_{j,i}^{(L)} x_i^{(L)}$  to device  $n+1$ .

The procedure continues until the  $N$ -th node sends its information to the sink; therefore, in the  $j$ -th transmission, the sink receives the weighted sum

$$\begin{aligned} y(j) &= \sum_{i=1}^N \phi_{j,i}^{(1)} x_i^{(1)} + \dots + \sum_{i=1}^N \phi_{j,i}^{(L)} x_i^{(L)} \\ &= \phi_j^{(1)} \mathbf{x}^{(1)} + \dots + \phi_j^{(L)} \mathbf{x}^{(L)}, \end{aligned} \quad (8)$$

where  $\phi_j^{(l)} = (\phi_{j,1}^{(l)}, \dots, \phi_{j,i}^{(l)}, \dots, \phi_{j,N}^{(l)})$  is the row vector with the randomly generated numbers from all devices regarding source  $X_l$ . Recall that  $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_i^{(l)}, \dots, x_N^{(l)})^\top$  is the signal that contains readings of the same source  $X_l$  collected by all nodes.

The aforementioned procedure is repeated  $M$  times, each indexed by  $j = 1, \dots, M$ , acquiring  $M$  measurements

$$\begin{aligned} y(1) &= \phi_1^{(1)} \mathbf{x}^{(1)} + \dots + \phi_1^{(L)} \mathbf{x}^{(L)}, \\ y(2) &= \phi_2^{(1)} \mathbf{x}^{(1)} + \dots + \phi_2^{(L)} \mathbf{x}^{(L)}, \\ &\vdots \\ y(M) &= \phi_M^{(1)} \mathbf{x}^{(1)} + \dots + \phi_M^{(L)} \mathbf{x}^{(L)}. \end{aligned}$$

Let us denote by  $\Phi^{(l)} = [\phi_1^{(l)}, \phi_2^{(l)}, \dots, \phi_M^{(l)}]^\top$  the  $M \times N$  sensing matrix that corresponds to the source  $X_l$ ; the  $j$ -th row

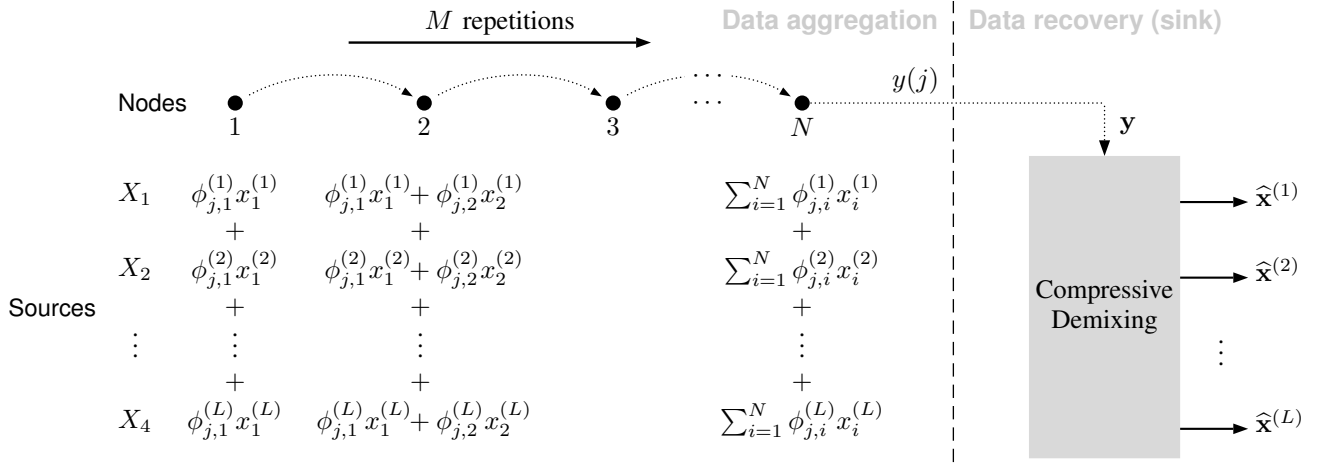


Fig. 2. Proposed scheme for joint data aggregation and recovery via compressive demixing.

of this matrix contains the random weights generated for the  $j$ -th transmission of the related sensor signal collected by  $N$  nodes. Then, the above equations can be written in a matrix form as

$$\mathbf{y} = \Phi^{(1)} \mathbf{x}^{(1)} + \dots + \Phi^{(L)} \mathbf{x}^{(L)}, \quad (9)$$

where  $\mathbf{y} = (y(1), y(2), \dots, y(M))^T$  is the vector containing the  $M$  mixed measurements. This procedure results in  $\mathcal{O}(MN)$  transmissions in the network. Similar to [8], the generation of the sensing matrices at each node is based on a random seed broadcasted by the sink. The sink can reproduce  $\Phi^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$ , using the nodes' identifications.

Although the proposed data gathering strategy is based on a direct multi-hop route, it can readily address networks organized in a tree-based structure, similarly to [7], [8], [10].

### B. Joint Data Recovery via Compressive Demixing

The second stage deals with joint recovery of the sensor signals  $\mathbf{x}^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$  at the sink, using the measurements  $\mathbf{y}$  and the sensing matrices  $\Phi^{(l)}$ . We assume that every signal  $\mathbf{x}^{(l)}$  has a sparse representation  $\mathbf{s}^{(l)}$  on the discrete cosine transform (DCT) domain, that is,  $\mathbf{x}^{(l)} = \Psi \mathbf{s}^{(l)}$ , where  $\Psi$  is the DCT matrix<sup>1</sup>. Defining  $\mathbf{A}^{(l)} := \Phi^{(l)} \Psi$ , the measurement vector given by (9) can be written as

$$\mathbf{y} = \mathbf{A}^{(1)} \mathbf{s}^{(1)} + \dots + \mathbf{A}^{(L)} \mathbf{s}^{(L)}. \quad (10)$$

In order to apply compressive demixing to recover the constituent sparse signals  $\mathbf{s}^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$ , these should be incoherent. Assuming the same sparsifying basis  $\Psi$  for every source signal  $\mathbf{x}^{(l)}$ , incoherence is satisfied as long as the columns of the employed measurement matrices  $\Phi^{(1)}$ ,  $\dots$ ,  $\Phi^{(L)}$  are weakly correlated [14]. The use of random Gaussian measurement matrices ensures that the structures in the involved signals are oriented generically with respect to

each other. Then, the sink can recover  $\mathbf{s}^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$ , by solving the compressive demixing problem

$$\begin{aligned} [\hat{\mathbf{s}}^{(1)}, \dots, \hat{\mathbf{s}}^{(L)}] &= \arg \min_{\{\mathbf{s}^{(l)}\}} \sum_{l=1}^L \alpha_l \|\mathbf{s}^{(l)}\|_1 \\ \text{s.t. } \mathbf{y} &= \sum_{l=1}^L \mathbf{A}^{(l)} \mathbf{s}^{(l)}, \end{aligned} \quad (11)$$

where  $\alpha_l > 0$  are regularization parameters that trade the relative sparsity of solutions. Here, we assume that  $\alpha_l = 1$ , for all  $l \in \{1, 2, \dots, L\}$ . After solving Problem (11), the final estimates of the signal vectors are calculated as:  $\hat{\mathbf{x}}^{(l)} = \Psi \hat{\mathbf{s}}^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$ .

To the best of our knowledge, there are no sharp recovery bounds for Problem (11). A related problem, however, has been studied in [25]. More particularly, the authors of [25] consider a constrained compressive demixing problem of the form

$$\begin{aligned} [\hat{\mathbf{s}}^{(1)}, \dots, \hat{\mathbf{s}}^{(L)}] &= \arg \min_{\{\mathbf{s}^{(l)}\}} \left\| \sum_{l=1}^L \mathbf{A}^{(l)} \cdot \mathbf{s}^{(l)} - \mathbf{y} \right\|_2^2 \\ \text{s.t. } \|\mathbf{s}^{(l)}\|_1 &\leq \|\tilde{\mathbf{s}}^{(l)}\|_1, \text{ for all } l \in \{1, \dots, L\}, \end{aligned} \quad (12)$$

where  $\mathbf{A}^{(l)} \in \mathbb{R}^{M \times N}$  are sub-sampling matrices with elements drawn *i.i.d.* from a Gaussian distribution, and  $\tilde{\mathbf{s}}^{(l)}$ ,  $l \in \{1, 2, \dots, L\}$  are the signals we aim to reconstruct. We refer to (12) as the oracle problem. Note that (12) requires access to the  $\ell_1$ -norms of these signals, not necessarily access to the signals themselves. Theorem A in [25] states that (12) succeeds in recovering the signal, if the number of measurements  $M$  is slightly larger than the sums of the statistical dimensions of the  $\ell_1$ -norms evaluated at  $\tilde{\mathbf{s}}^{(1)}, \tilde{\mathbf{s}}^{(2)}, \dots, \tilde{\mathbf{s}}^{(L)}$ . Using results from [18], it can be shown that the total statistical dimension for  $L$  sparse signals with similar sparsity level equal to  $\kappa$  is  $\mathcal{O}(L\kappa \log(N/\kappa))$ . In a multi-hop routing scenario, this would result in an overall number of transmissions of the order  $\mathcal{O}(NL\kappa \log(N/\kappa))$ . Although this number is of the same order as the communication cost of CDG [8] when applied to

<sup>1</sup>DCT is chosen so as to align with prior work [8], [10]. As it was shown in [11], DCT is a good sparsifying basis for air pollution data resulting in compressible data representations.

multiple signals (see Section II-B), our experiments show that (12) significantly outperforms CDG.

Recall that the above bound on the number of measurements applies to the oracle problem, and not exactly to the problem we propose to solve, (11). Nevertheless, it can be used as an indication of the number of measurements for (11). In the next section, however, our experiments will show not only that this bound can be quite loose, but also that our method outperforms the prior state of the art.

#### IV. EXPERIMENTS

To evaluate the performance of our framework, we consider the problem of air-pollution monitoring based on actual air-pollution sensor readings taken from a database of the United States Environmental Protection Agency (EPA) [28]. In particular, we use  $6 \times 10^5$  actual sensor readings of three air pollutants, namely, carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>) and sulfur dioxide (SO<sub>2</sub>), measured during 2015. We consider a multi-hop network architecture with  $N = 1000$  nodes<sup>2</sup>. The transmission of the sensor values is assumed to be conducted via the Long Range Wide Area Network (LoRaWAN) protocol [29], the most recent low-power wireless networking protocol, specifically designed for IoT architectures, which allows for extremely low-rate data transmission to long ranges. In addition, we assume that the signals sent between sensors are discretized using an analog-to-digital converter, where the bit-depth is 16 bits, and all signals are assumed sparse in the DCT domain (see [11, Sec. VI-B.1]).

##### A. Performance Comparison in the Noiseless Case

We compare the proposed design against (i) independent signal recovery proposed in [8] (baseline system), (ii) the extended  $\ell_1 - \ell_1$  reconstruction algorithm that uses (multiple) side information developed in [10], (iii) the solution given by the DCS system [13], and (iv) the solution of the oracle problem (12). As mentioned in Section III-B, the latter system cannot be practically applied in our setting; it is included in our comparison just for illustration.

Figure 3 shows the results of our experiments and depicts the achieved reconstruction quality versus the total number of transmissions. The reconstruction quality is computed as the total relative error<sup>3</sup>  $\sum_{l=1}^L \frac{\|\mathbf{x}^{(l)} - \hat{\mathbf{x}}^{(l)}\|_2}{\|\mathbf{x}^{(l)}\|_2}$ , and, for the multi-hop routing scenario, the number of transmissions equals the number of measurements at every node multiplied by the number of nodes ( $N = 1000$ ). The maximum possible number of measurements for the proposed method and the oracle problem then equals  $M = N = 1000$ .

We compare the performance of the tested methods by examining the number of the transmissions required to achieve a similar reconstruction quality. According to Fig. 3, the proposed method outperforms the state of the art, when high quality reconstruction is considered, i.e., the total relative error is smaller than 2.2. In this case, compared to the DCS and the

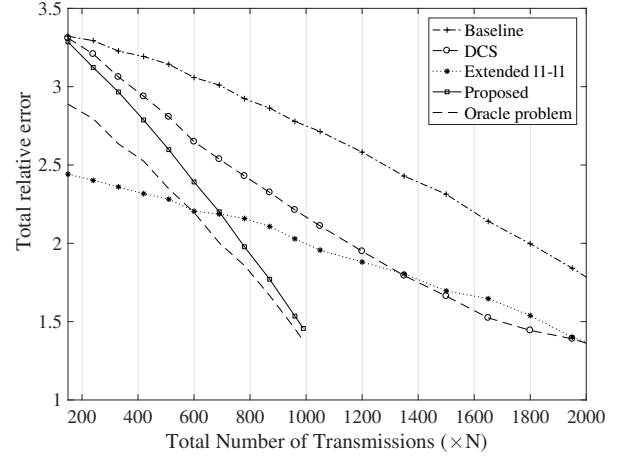


Fig. 3. Performance evaluation of the proposed method against the baseline and the DCS systems using actual measurements from the EPA database.

baseline approach, the proposed method achieves a reduction in the total number of transmissions up to 45% and 60%, respectively. This means that the transmissions among the nodes are also reduced by the same factor, thereby leading to significant power savings and, hence, a prolonged WSN lifetime. The results obtained for the proposed method are similar to the results of the oracle problem and significantly better than the results of the baseline system [8]. For low quality reconstruction requirements (total relative error  $> 2.2$ ), the extended  $\ell_1 - \ell_1$  algorithm requires less transmissions than its competitors. The better performance of the  $\ell_1 - \ell_1$  method when the number of measurements is low has also been reported in [10], [12]. However, when high reconstruction quality is required, the proposed method dramatically improves and outperforms all other methods.

##### B. Performance Comparison under Noise

We perform the same comparison under the assumption that the measurements are corrupted by noise. The noise can be attributed to both the sensing and the transmission steps, and can be modelled as additive white Gaussian noise (AWGN):  $\mathbf{y} = \sum_{l=1}^L \mathbf{A}^{(l)} \mathbf{s}^{(l)} + \boldsymbol{\eta}$ , where  $\boldsymbol{\eta} \in \mathbb{R}^M$  is drawn *i.i.d.* from  $\mathcal{N}(0, \sigma_\eta)$ , with  $\sigma_\eta$  denoting standard deviation of the noise. We varied  $\sigma_\eta = 2, 5, 10$  so as to include weak, moderate and strong noise corruption on the measurements. Results are presented in Fig. 4. For a similar reconstruction quality, the proposed method provides a significant reduction in the number of total transmissions compared to the baseline and the DCS system, even when the noise level increases. Recovery via the extended  $\ell_1 - \ell_1$  method still performs better for low reconstruction quality. It is, however, less robust compared to all others methods [10]. Moreover, unlike the DCS and baseline systems, the proposed scheme provides more robust recovery against measurement and communication noise of various levels. Compared to the noiseless case, the relative error increases on average by 0.5% ( $\sigma_\eta = 2$ ), 2.1% ( $\sigma_\eta = 5$ ) and 4.2% ( $\sigma_\eta = 10$ ).

<sup>2</sup>A typical large-scale WSN comprises thousands of sensor nodes [7], [9].

<sup>3</sup>The error floor in Fig. 3 is relatively large because pollution signals are not exactly sparse in the DCT domain, only compressible.

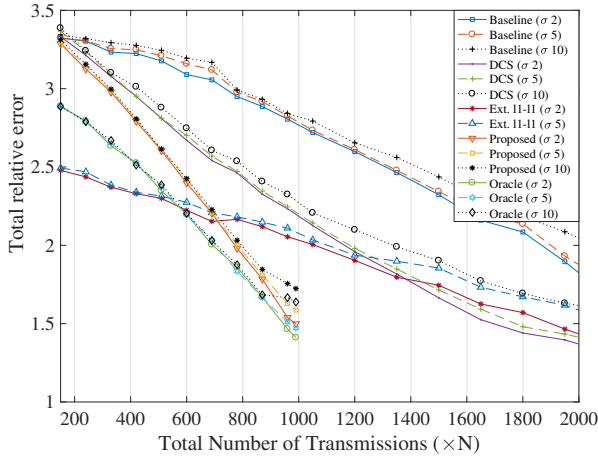


Fig. 4. Performance evaluation of the proposed method, the baseline and the DCS systems for various noise standard deviations:  $\sigma_\eta = 2, 5, 10$ .

## V. CONCLUSIONS

We have proposed a novel design for joint acquisition and recovery of multiple heterogeneous signals, well-suited for large-scale Internet-of-Things applications. Specifically, in this work, we have addressed the efficient transmission of diverse types of data collected from different sources by all nodes in the network. Heterogeneous data are encoded into a *single* low-dimensional measurement vector and recovered at the sink using the compressive demixing paradigm [14], [15]. Compared to existing designs that are based on compressed acquisition of multi-modal data, our method achieves a significant reduction of the total number of transmissions in the network. This gain translates into power savings at the wireless nodes or IoT devices. The proposed scheme was evaluated on a real dataset concerning an air-pollution monitoring application. The experimental results showed that our method significantly reduces (up to 60%) the required transmission rate against prior art [7], [10], [13], for high quality reconstruction. For increased levels of noise, our design is more robust than the prior state of the art.

## REFERENCES

- [1] I. F. Akyildiz and M. C. Vuran, *Wireless sensor networks*. John Wiley & Sons, 2010, vol. 4.
- [2] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Annu. Joint Conf. IEEE Comput. and Commun. (INFOCOM)*, vol. 3, 2003, pp. 1848–1857.
- [3] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for WSNs by exploiting spatiotemporal correlation," *IEEE Trans. Parallel and Distributed Syst.*, vol. 18, no. 7, pp. 1010–1023, 2007.
- [4] D. I. Sacaleanu, R. Stoian, D. M. Ofriim, and N. Deligiannis, "Compression scheme for increasing the lifetime of wireless intelligent sensor networks," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 709–713.
- [5] N. Deligiannis, E. Zimos, D. Ofriim, Y. Andreopoulos, and A. Munteanu, "Distributed joint source-channel coding with copula-function-based correlation modeling for wireless sensors measuring temperature," *IEEE Sensor J.*, vol. 15, no. 8, pp. 4496–4507, 2015.
- [6] E. Zimos, D. Toumpakaris, A. Munteanu, and N. Deligiannis, "Multiterminal source coding with copula regression for wireless sensor networks gathering diverse data," *IEEE Sensors Journal*, vol. 17, no. 1, pp. 139–150, 2017.
- [7] 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, 2008.
- [8] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Efficient measurement generation and pervasive sparsity for compressive data gathering," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3728–3738, 2010.
- [9] X. Liu, Y. Zhu, L. Kong, C. Liu, Y. Gu, A. V. Vasilakos, and M. Wu, "CDC: compressive data collection for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 8, pp. 2188–2197, 2015.
- [10] E. Zimos, J. F. Mota, M. R. Rodrigues, and N. Deligiannis, "Internet-of-things data aggregation using compressed sensing with side information," in *Int. Conf. Telecomm.(ICT)*. IEEE, 2016.
- [11] N. Deligiannis, J. F. C. Mota, E. Zimos, and M. R. D. Rodrigues, "Heterogeneous networked data recovery from compressive measurements using a copula prior," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5333–5347, 2017.
- [12] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Strategies, geometry, and bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4472–4496, 2017.
- [13] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [14] M. B. McCoy and J. A. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Computational Mathematics*, vol. 14, no. 3, pp. 503–567, 2014.
- [15] M. B. McCoy, V. Cevher, Q. T. Dinh, A. Asaei, and L. Baldassarre, "Convexity in source separation: Models, geometry, and algorithms," *IEEE Sign. Process. Mag.*, vol. 31, no. 3, pp. 87–95, 2014.
- [16] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [17] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [18] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [19] 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, 2010.
- [20] J. Tropp, A. C. Gilbert *et al.*, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Academy Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [22] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
- [23] N. Deligiannis, J. F. C. Mota, B. Cornelis, M. R. D. Rodrigues, and I. Daubechies, "Multi-modal dictionary learning for image separation with application in art investigation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 751–764, Feb 2017.
- [24] —, "X-ray image separation via coupled dictionary learning," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 3533–3537.
- [25] M. B. McCoy and J. A. Tropp, "The achievable performance of convex demixing," *arXiv:1309.7478*, 2013.
- [26] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for wsns: Sparse signal modeling and monitoring framework," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3447–3461, 2012.
- [27] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conf. Signals, Syst., Comput.*, 2005, pp. 1537–1541.
- [28] [Online]. Available: <http://www3.epa.gov/airdata/>.
- [29] [Online]. Available: <https://www.lora-alliance.org/>.