Internet-of-Things Data Aggregation Using Compressed Sensing with Side Information

Evangelos Zimos*§, João F. C. Mota†, Miguel R. D. Rodrigues†, and Nikos Deligiannis*§
*Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.
†Electronic and Electrical Engineering Dept., University College London, Torrington Place, London WC1E 7JE, UK.
§iMinds, Gaston Crommenlaan 8 (b102), 9050 Gent, Belgium.

E-mail: {ezimos, ndeligia}@etro.vub.ac.be, {j.mota, m.rodrigues}@ucl.ac.uk

***Abstract*—The Internet-of-Things (IoT) is the key enabling technology for transforming current urban environments into so-called Smart Cities. One of the goals behind making cities smarter is to provide a healthy environment that improves the citizens' quality of life and wellbeing. In this work, we introduce a novel data aggregation mechanism tailored to the application of large-scale air pollution monitoring with IoT devices. Our design exploits the intra- and inter-source correlations among air-pollution data using the framework of compressed sensing with side information. The proposed method delivers significant improvements in the data reconstruction quality with respect to the state of the art, even in the presence of noise when measuring and transmitting the data.**

I. Introduction

Large cities nowadays suffer from high concentration of harmful substances in the atmosphere, which causes damage to the environment, human health, and citizens' quality of life. Air pollution does not only affect people but also has a negative effect on the ecological system as a whole, including plants and animals. Furthermore, it is responsible for the global rise in temperature, a phenomenon known as "global warming". Internet-of-Things (IoT) applications that leverage the recent technological advances in ubiquitous broadband connectivity, cloud services and big data analytics can revolutionize the way we deal with the air-pollution problem. Typical environmental monitoring scenarios involve a large number of wireless sensor devices—in the order of hundreds or thousands—spread around a geographical area. Data gathering in these large-scale setups is usually accomplished through multi-hop wireless transmission from the sensors to one or more data sinks [1].

A prime constraint in the design of large-scale wireless sensor networks (WSNs) is energy consumption, since wireless sensors are typically powered by small batteries. Power savings can be achieved by decreasing the encoding complexity on the sensor node and by reducing the radio emission. To this end, data gathering schemes should exploit both the intra- and inter-sensor correlations, without increasing the computational effort at the sensor nodes and without requiring inter-sensor communication. Moreover, as information is sent over error-prone wireless channels, effective data protection mechanisms are required to provide reliable data transmission.

Prior works on data aggregation propose predictive compression techniques for WSNs, such as Differential Pulse-Code Modulation (DPCM) followed by entropy coding [2], [3]. Other approaches consider collaborative wavelet transform coding [4] or clustered data aggregation [5]. However, both techniques require excessive transmission of overhead information and, hence, additional encoding complexity due to inefficient handling of abnormal events. Finally, an alternative strategy adheres to distributed source coding (DSC), a paradigm that leverages inter-sensor data correlation at the decoder side. DSC is a promising technique for WSNs as it shifts the computational burden towards the sink node. However, previous works on DSC [6]–[9] perform well only for a limited number of sensors, typically two or three.

Distributed compressed sensing (DCS) [10] is another solution for data aggregation, where random measurements are transmitted from each sensor and the data are recovered at the joint decoder by leveraging the spatiotemporal correlations. As the measurements are sent directly from the sensors to the sink node, this architecture leads to significant and unbalanced battery consumption in large-scale setups. Since the work of Haupt *et al.* [11], [12], compressed sensing (CS) has been used to devise an efficient technique for data gathering and recovery in large-scale WSNs. The technique is tailored to a multi-hop transmission paradigm with the goal to balance the power consumption of the sensing devices. The work in [13] extends this framework by proposing a data recovery algorithm that combines principal component analysis (PCA) with CS for grid network topologies. Finally, an alternative scheme that considers multi-hop routing is presented in [14], where only the spatial correlation of the data is exploited at the sink node to recover the sensor readings.

In this paper, we propose a novel large-scale data aggregation mechanism that is based on an extension of the framework in [15], [16], which addressed the problem of compressed sensing with side information. Our approach follows a multi-hop data transmission scenario, which lies in contrast with the transmission mechanisms in [9], [10]. Furthermore, unlike the works in [11]–[14], which exploit only the intra- or inter-sensor data correlations, our method exploits both types of dependencies. Since the number of correlated sources may be more than one, the algorithm can incorporate multiple correlated signals as side information, resulting in a multi-hypothesis-based CS scenario. To evaluate the performance of our framework, we use real air-pollution data taken from a database of the United States Environmental Protection Agency (EPA) [17]. The experimental results show that, for a given data rate, the proposed method provides significant reductions in the mean squared-error (MSE) of the recovered data compared to the classical CS [11], [12] and DCS [10] methods. Alternately, the proposed method reduces the required data rates for a given data reconstruction quality, thereby resulting in less network traffic and prolonged system lifetime. Furthermore, the proposed design shows robustness against measurement and communication noise without introducing excessive computation on the sensor nodes.

II. Background

In the classical CS framework, the signal of interest $\underline{x} \in \mathbb{R}^N$ can be written in the form $\underline{x} = \mathbf{\Psi}\underline{s}$, where $\underline{s}$ is its K-sparse representation (i.e., $\|\underline{s}\|_0 = K$, with $\|\cdot\|_0$ denoting the ℓ_0 pseudo norm) and $\mathbf{\Psi} \in \mathbb{R}^{N\times N_0}$ is an orthonormal or overcomplete basis, called dictionary. Let $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ be a sensing (or encoding) matrix, such that the measurement matrix $\mathbf{A} = \mathbf{\Phi\Psi}$ satisfies either the mutual coherence property [18], the restricted isometry property [19] or the null-space property [20]. The CS theory states that $\underline{x}$ can be recovered using the measurement matrix $\mathbf{A}$ and $M \ll N$ linear random measurements $\underline{y} = \mathbf{\Phi}\underline{x} = \mathbf{A}\underline{s}$. If the number of measurements is sufficiently large, then $\underline{s}$ is the unique minimizer of the following optimization problem, known as Basis Pursuit [21]:

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

The CS theory can be modified so as to leverage a signal correlated to the signal of interest, called side information (SI), which is assumed to be provided *a priori* at the decoder in order to aid reconstruction [15], [16], [22]–[26]. In particular, the decoder aims at recovering the signal $\underline{x}$ based on the measurements $\underline{y}$, the measurement matrix $\mathbf{A}$ and a side information vector, say $\underline{w}$, which is correlated with $\underline{s}$. The problem of CS with side information can be expressed via the following $\ell_1 - \ell_1$ optimization problem [15], [16], [24]

$$\hat{\underline{s}} = \arg\min_{\underline{s}} \left(\|\underline{s}\|_1 + \|\underline{s} - \underline{w}\|_1\right) \quad \text{s.t.} \quad \underline{y} = \mathbf{A}\underline{s}. \tag{2}$$

In [15], bounds and geometrical interpretations were provided, showing that $\ell_1 - \ell_1$ minimization improves the performance of CS if the side information is of good quality.

III. Data Gathering With CS

Instead of following a classical multi-hop scenario, where each node sends to its neighbour both its own information and the relayed information from other nodes, previous works [12], [14] assumed a different approach in which a weighted sum of readings is transmitted. Let $x(i) \in \mathbb{R}$ denote the reading of sensor $i \in \{1, 2, \ldots, N\}$. As seen in Fig. 1, in the approach of [12], [14], the procedure is initialized by sensor 1, which randomly generates a number $\phi_{j,1}$ and transmits the value $\phi_{j,1}x(1)$ to node 2. Subsequently, node 2 randomly generates another number $\phi_{j,2}$, calculates the weighted sum $\sum_{i=1}^{2} \phi_{j,i}x(i)$ and sends it to node 3. To generalize, each node i generates a random number $\phi_{j,i}$, computes the value $\phi_{j,i}x(i)$, adds it to the sum of the previous relayed values and sends $\sum_{n=1}^{i} \phi_{j,n}x(n)$ to node $i+1$. The procedure repeats until node N sends its information to the sink node, which receives the final weighted sum of the sensors readings, i.e., $y(j) = \sum_{i=1}^{N} \phi_{j,i}x(i)$. The procedure is repeated for M times, each indexed by $j = 1, \ldots, M$. Hence, the sink node obtains M weighted sums $\{y(j)\}_{j=1}^{M}$, which can be expressed as

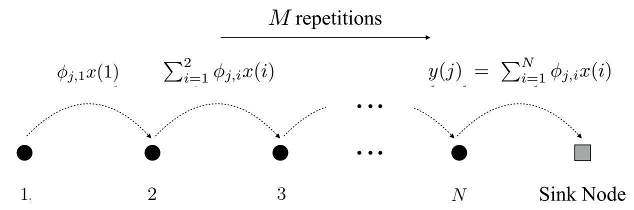


Fig. 1. Multi-hop transmission in a large-scale WSN using CS [12], [14].

$$\begin{aligned} \underline{y} &= [\boldsymbol{\phi}_1 \ldots \boldsymbol{\phi}_i \ldots \boldsymbol{\phi}_N] \cdot \underline{x} \\ &= \mathbf{\Phi} \cdot \underline{x}, \end{aligned} \tag{3}$$

where $\underline{y} = [y(1) \ldots y(j) \ldots y(M)]^T$ is the vector with the weighted sums (a.k.a. measurements), $\boldsymbol{\phi}_i = [\phi_{1,i} \ldots \phi_{j,i} \ldots \phi_{M,i}]^T$ is the column vector with the randomly generated numbers at the node i grouped in the matrix $\mathbf{\Phi}$, and $\underline{x} = [x(1) \ldots x(i) \ldots x(N)]^T$ is the vector with readings from different sensors. The reconstruction of the sensor values $\underline{x}$ at the sink node can be done via CS algorithms, such as OMP [27], CoSaMP [28], AMP [29] or BP-CS [30].

In order to avoid sending the sensing matrix from the sensors to the sink node, the following strategy is adopted: the sink node broadcasts a random seed to the entire network; using this global seed, each sensor generates its own seed based on its id. The sensing coefficients are generated by a pseudo-random number generator that is pre-installed on every sensor. These coefficients can be reproduced at the sink given that the identification numbers of all sensors are known.

IV. Proposed Architecture Using CS with SI

The architecture proposed in [12], [14] exploits the intra-source correlation among the sensor readings using compressed sensing. If the sensors gather values from different pollutants, alias, sources (e.g., CO, NO_2, SO_2), the reconstruction of each source is conducted independently, ignoring the underlying inter-source data dependencies. In this work, we propose a novel data gathering and recovery design that leverages both intra- and inter-source data correlations among sensor readings.

We consider a large-scale WSN comprising N sensors that form a multi-hop route to the sink, as depicted in Fig. 1. Each sensor $i \in \{1, 2, \ldots, N\}$ observes the correlated sources $X_1, X_2, \ldots, X_L$ that take values in their corresponding continuous alphabets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_L$. We denote as $x_l(i)$ a reading produced from the source $X_l, l \in \{1, 2, \ldots, L\}$ and observed from the sensor i. A tuple of readings $\mathbf{x}^{(i)} = [x_1(i), \ldots, x_L(i)]^T$ is assumed to be drawn *i.i.d.* according to the joint probability density function (pdf) $f_{\mathbf{X}}(x_1, x_2, ..., x_L)$.

1) Data Aggregation: During the data aggregation stage, the sensor nodes first acquire the correlated tuples of readings and they proceed to the transmission of the readings of each source separately. Initially, they start transmitting the values of source X_1 using the data aggregation mechanism described in Section III. When all measurements of source X_1 are gathered at the sink node, we repeat the data gathering procedure for the rest of the sources, namely, $X_2, \ldots, X_L$.

For each source $l \in \{1, 2, \ldots, L\}$, the sink node obtains M_l weighted sums $\{y_l(j)\}_{j=1}^{M_l}$, which can be expressed as

$$\begin{aligned} \underline{y_l} &= \left[\boldsymbol{\phi}_1^{(l)} \ldots \boldsymbol{\phi}_i^{(l)} \ldots \boldsymbol{\phi}_N^{(l)}\right] \cdot \underline{x_l} \\ &= \boldsymbol{\Phi}^{(l)} \cdot \underline{x_l}, \end{aligned} \tag{4}$$

where $\underline{y_l} = [y_l(1) \ldots y_l(j) \ldots y_l(M_l)]^T$ is the vector with the weighted sums (a.k.a. measurements), $\boldsymbol{\phi}_i^{(l)} = [\phi_{1,i}^{(l)} \ldots \phi_{j,i}^{(l)} \ldots \phi_{M_l,i}^{(l)}]^T$ is the column vector with the randomly generated numbers at the node i grouped in the matrix $\boldsymbol{\Phi}^{(l)}$, and $\underline{x_l} = [x_l(1) \ldots x_l(i) \ldots x_l(N)]^T$ is the vector with readings from different sensors observing the source X_l.

In data acquisition systems, the value $\phi_{j,i}^{(l)} x_l(i)$ that contributes to the j-th measurement and is sent from the sensor i to its neighbor, is contaminated with noise (such as quantization). This noise is usually modelled as white additive Gaussian noise (AWGN) $Z_{\text{meas},j}^{(l)} \sim \mathcal{N}(0, \sigma_{\text{meas},j}^{(l)})$, where $\sigma_{\text{meas},j}^{(l)}$ is the standard deviation. In addition, the value $\phi_{j,i}^{(l)} x_l(i)$ is also corrupted by additive white Gaussian transmission noise $Z_{\text{trans},j}^{(l)} \sim \mathcal{N}(0, \sigma_{\text{trans},j}^{(l)})$ with standard deviation $\sigma_{\text{trans},j}^{(l)}$. Hence, the j-th measurement received at the sink node is written as

$$y_l(j) = \sum_{i=1}^{N} \phi_{j,i}^{(l)} x_l(i) + \sum_{i=1}^{N} z_{\text{meas},j}^{(l)}(i) + \sum_{i=1}^{N} z_{\text{trans},j}^{(l)}(i).$$

We define $z_{\text{MS}}^{(l)}(j) = \sum_{i=1}^{N} z_{\text{meas},j}^{(l)}(i)$ and $z_{\text{TR}}^{(l)}(j) = \sum_{i=1}^{N} z_{\text{trans},j}^{(l)}(i)$ the total measurement and transmission noise value that corresponds to the j-th measurement of source X_l, where $z_{\text{MS}}^{(l)}(j)$ and $z_{\text{TR}}^{(l)}(j)$ are drawn from $\mathcal{N}(0, \sigma_{\text{MS},j}^{(l)})$ and $\mathcal{N}(0, \sigma_{\text{TR},j}^{(l)})$, respectively. The measurements in (4) are then written as

$$\underline{y_l} = \boldsymbol{\Phi}^{(l)} \cdot \underline{x_l} + \underline{z_l}, \tag{5}$$

where $\underline{z_l} = [z_l(1) \ldots z_l(j) \ldots z_l(M_l)]^T$ is the aggregate noise vector, each component of which is assumed to be drawn *i.i.d.* from the normal distribution $\mathcal{N}(0, \sigma_z^{(l)})$ with standard deviation $\sigma_z^{(l)} = \sqrt{\left(\sigma_{\text{MS}}^{(l)}\right)^2 + \left(\sigma_{\text{TR}}^{(l)}\right)^2}$.

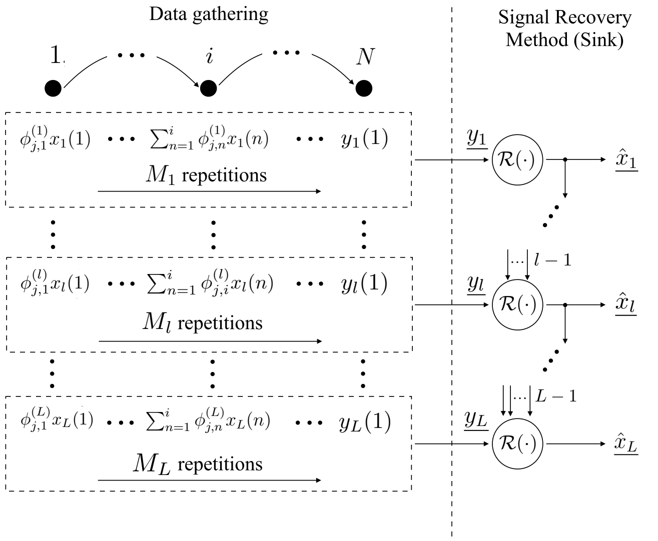


Fig. 2. Proposed system for gathering and successively reconstructing correlated sensor data. For each source X_l, the multi-hop transmission among the sensors takes place for M_l repetitions till the measurements vector $\underline{y_l}$ is fully gathered at the sink node.

2) Data Recovery: Upon receiving the set of measurements from all sources, the sink node proceeds to the data recovery stage, which deals with reconstructing the sensor readings $\left\{\underline{x_l}\right\}_{l=1}^{L}$ based on the measurements $\left\{\underline{y_l}\right\}_{l=1}^{L}$ and the matrices $\{\boldsymbol{\Phi}^{(l)}\}_{l=1}^{L}$. Applying classical CS algorithms, such as OMP [27], CoSaMP [28], AMP [29] or BP-CS [30], to independently recover each signal vector $\underline{x_l}$, based only on the measurements $\underline{y_l}$ and the matrix $\boldsymbol{\Phi}^{(l)}$, would only exploit the *intra*-source signal correlation (*baseline scenario*). To exploit both the *inter*- and *intra*-source dependencies, in this work, we propose a novel design that recovers the ensemble of signals $\{\underline{x_l}\}_{l=1}^{L}$ using the ensemble of measurements vectors $\{\underline{y_l}\}_{l=1}^{L}$ and the matrices $\{\boldsymbol{\Phi}^{(l)}\}_{l=1}^{L}$. As depicted in Fig. 2, the proposed scheme applies successive reconstruction of the sensor readings $\left\{\underline{x_l}\right\}_{l=1}^{L}$. In particular, the joint decoder at the sink node is separated into L recovery stages. At the l-th stage, the reconstruction of $\underline{x_l}$ is done via the recovery algorithm $\mathcal{R}\left(\underline{y_l}, \boldsymbol{\Phi}^{(l)}, \underline{\hat{x}_1}, \ldots, \underline{\hat{x}_{l-1}}\right)$ that uses the corresponding gathered measurements $\underline{y_l}$, the matrix $\boldsymbol{\Phi}^{(l)}$ as well as all the previously reconstructed vectors $\underline{\hat{x}_1}, \ldots, \underline{\hat{x}_{l-1}}$ as multiple side information (i.e., multi-hypothesis scenario).

The recovery algorithm $\mathcal{R}(\cdot)$ is based on the $\ell_1 - \ell_1$ algorithm proposed in [15], where we modified the optimization problem so as to leverage multiple side information signals. To this end, the problem can take the following form

$$\underline{\hat{s}_l} = \arg\min_{\underline{s_l}} \left(\|\underline{s_l}\|_1 + \sum_{k=1}^{l-1} \omega_k \|\underline{s_l} - \underline{\hat{s}_k}\|_1 \right) \text{ s.t. } \underline{y_l} = \mathbf{A}^{(l)} \underline{s_l},$$

where $\mathbf{A}^{(l)} = \boldsymbol{\Phi}^{(l)} \boldsymbol{\Psi}$ is the measurement matrix of X_l, $\underline{s_l} = \boldsymbol{\Psi}^{-1} \underline{x_l}, \forall l \in \{1, 2, \ldots, L\}$ are the compressible representations of the signals $\{\underline{x_l}\}_{l=1}^{L}$ with $\underline{\hat{s}_l}$ denoting the corresponding reconstructed vector, and $\{\omega_k > 0\}_{k=1}^{l-1}$ are the weights establishing a tradeoff between signal sparsity and fidelity to the information signal $\underline{\hat{s}_k}$. This paper considers a simple form of (**??**) where $\omega_k = 1, \forall k \in \{1, \dots, l-1\}$.

We assume that all sensors use the same measurement matrix $\mathbf{\Phi}^{(l)}$. To avoid transmitting $\mathbf{\Phi}^{(l)}$ from the sensors to the sink node, we adhere to the method in [14]. Namely, the sink broadcasts a random seed to the entire network and each sensor generates its own seed and its unique identification. The coefficients used by the sensors are identically reproduced at the sink using the same pseudo-random generator (assuming that the sink knows the identifications of all sensors).

3) Tree-based WSN structure: In practical settings, designing an extremely long multi-hop routing path via many sensors is difficult. This is due to variations in the density of sensor devices in urban and rural areas. For example, in an urban environment the sensor density is significantly higher, mainly because: (a) non line-of-sight transmissions between sensors in a city area call for less inter-sensor distance, and (b) urban regions contain numerous pollution sources and, hence, more refined pollution monitoring is required. To address this issue, we consider WSNs organized in a tree-based network structure [31]. In a typical tree-based design, the sink node comprises a number of children nodes, called *root* nodes, where each of them aggregates the sensor values of its assigned subtree. Within each subtree, the proposed data gathering scheme is applied: each parent node waits until receiving the values from its children nodes, it adds its own value and then transmits the weighted sum to the next parent node. The procedure repeats until the root nodes receive the weighted sums from their children nodes. The root nodes are capable of transmitting information to the sink node via Internet or satellite.

V. EXPERIMENTS

To evaluate the performance of the proposed system, we used 6×10^5 actual sensor readings of three pollutants, namely, CO, NO_2 and SO_2, from the United States Environmental Protection Agency database [17], measured during 2015. To provide a realistic example, we considered a tree-based network architecture with $N = 1000$ sensors, which comprises 15 subtrees defined by the sensor density in the geographic area[1]. The transmission of the sensor values is assumed to be conducted via the Long Range Wide Area Network (LoRaWAN) protocol [32], where the inter-sensor distance in a subtree does not exceed 2 km for urban areas and 22 km for rural areas. LoRaWAN is the most recent low-power wireless networking protocol, specifically designed for IoT architectures, which allows for extremely low-rate data transmission to long ranges. The low data rates (down to few bytes per second) and the LoRaWAN modulation lead to very low receiver sensitivity, which means extremely large link budgets (up to 148 dB). In addition, regarding the transmission part, we assume that the values sent from a sensor node to its neighbor are discretized using an analog-to-digital converter, where the bit-depth is 16 bits.

We investigate the robustness of the proposed design against noise modeled as AWGN with a common standard deviation ${\sigma_z}^2$. We varied the noise level by assuming that $\sigma_z \in \{0, 2, 5\}$ and calculated the aggregate normalized MSE, $\sum_{l=1}^{L} \frac{\|x_l - \hat{x}_l\|}{\|x_l\|}$, as a function of the measurements M_l, which are assumed to be common for all sources. The different considered schemes are: *(i)* the baseline scenario where each source is independently reconstructed using belief-propagation-based CS [30], *(ii)* the proposed method using CS with side information, and *(iii)* the DCS technique [10]. In the classical DCS scenario each signal of interest is constructed by many readings of the same sensor. In order to have a fair comparison with our design, we have modified this framework by assuming that each signal of interest contains readings from different sensors observing the same source.

The results for the noiseless case, which are depicted in Fig. 3(a), demonstrate that the proposed algorithm outperforms the baseline scenario achieving an MSE reduction of up to 28.2%. Moreover, when the number of measurements is less than 550, the proposed system outperforms DCS, resulting in a reduction of up to 18.3% in the MSE of the reconstructed data. Nevertheless, the MSE performance of both techniques is very similar when $M_l \geq 550$.

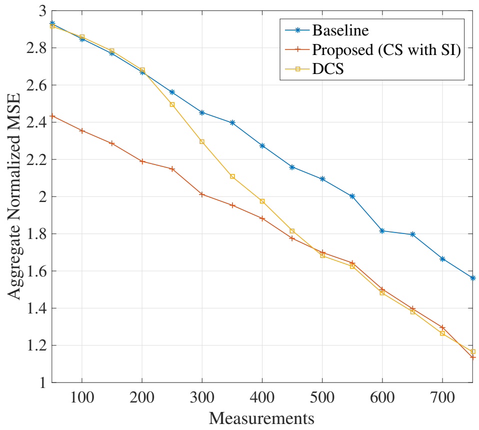
The results for the noisy case are depicted in Fig. 3(b), where we consider a moderate ($\sigma_z = 2$) and strong ($\sigma_z = 5$) noise scenario. When $\sigma_z = 2$, the MSE reductions against the baseline scenario and the DCS setup reach up to 27.4% and 21.2%, respectively. When $\sigma_z = 5$, the corresponding improvements mount to 27.9% and 20.4%. It should be mentioned that, when $\sigma_z = 5$, the performance of the proposed design is significantly higher that DCS until the number of measurements reach $M = 550$. Above $M = 550$ measurements, DCS provides better results. It can been observed that the proposed scheme provides robustness against noise, especially when the number of measurements is small. In particular, the MSE increases on average by 3.1% ($\sigma_z = 2$) and 8.4% ($\sigma_z = 5$) compared to the noiseless case.

VI. CONCLUSION

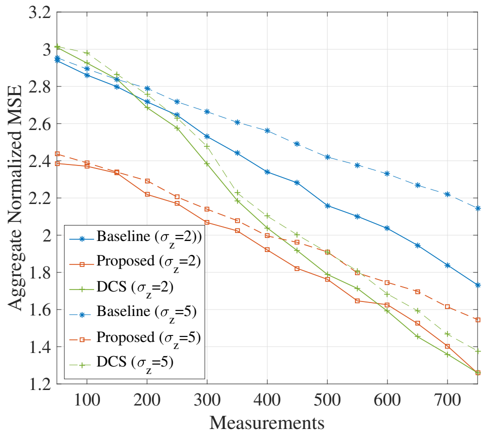
We proposed a novel data aggregation framework that is well-suited for large-scale IoT-based applications, such as air-pollution monitoring. The proposed framework efficiently exploits the intra- and inter-source correlations among multiple correlated sources. As shown by experimentation using data from the EPA dataset, the proposed scheme provides significant improvements in MSE reduction against prior art [10], [12], which are consistent even when the noise level increases.

[1]Each subtree corresponds to one of the following states: CA, NV, AZ, NC, SC, VA, WV, KY, TN, MA, RI, CT, NY, NJ, MD.

[2]In this experiment we have assumed that the standard deviation of the noise is the same for all sources. Hence, we have omitted the superscript (l).



(a) $\sigma_z = 0$ (Noiseless case)



(b) $\sigma_z \in \{2, 5\}$

Fig. 3. Performance evaluation for different noise levels.

REFERENCES

[1] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[2] M. Vecchio, R. Giaffreda, and F. Marcelloni, "Adaptive lossless entropy compressors for tiny IoT devices," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1088–1100, 2014.

[3] D. I. Sacaleanu, R. Stoian, D. M. Ofrim, and N. Deligiannis, "Compression scheme for increasing the lifetime of wireless intelligent sensor networks," in *Proc. Signal Process. Conf.* IEEE, 2012, pp. 709–713.

[4] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Annu. Joint Conf. IEEE Comput. and Commun.*, vol. 3. IEEE, 2003, pp. 1848–1857.

[5] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation," *IEEE Trans. Parallel and Distributed Syst.*, vol. 18, no. 7, pp. 1010–1023, 2007.

[6] V. Stanković, L. Stanković, S. Wang, and S. Cheng, "Distributed compression for condition monitoring of wind farms," *IEEE Trans. Sustainable Energy*, vol. 4, no. 1, pp. 174–181, 2013.

[7] N. Deligiannis, E. Zimos, D. Ofrim, 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.

[8] N. Deligiannis, E. Zimos, D. M. Ofrim, Y. Andreopoulos, and A. Munteanu, "Distributed joint source-channel coding with raptor codes for correlated data gathering in wireless sensor networks," in *Int. Conf. Body Area Networks*. ICST, 2014, pp. 279–285.

[9] F. Chen, M. Rutkowski, C. Fenner, R. C. Huck, S. Wang, and S. Cheng, "Compression of distributed correlated temperature data in sensor networks," in *Data Compression Conf.* IEEE, 2013, pp. 479–479.

[10] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *45th Annu. Allerton Conf. Commun., Control, and Computing*, 2005.

[11] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

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

[13] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "Data acquisition through joint compressive sensing and principal component analysis," in *IEEE Global Telecommun. Conf.* IEEE, 2009, pp. 1–6.

[14] 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.

[15] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *[Online]. Available: http://arxiv.org/abs/1408.5250*, 2014.

[16] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conf. Signal and Inform. Process.*, 2014, pp. 512–516.

[17] [Online]. Available: http://www3.epa.gov/airdata/.

[18] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[19] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[20] 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.

[21] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[22] E. Zimos, J. F. C. Mota, M. R. D. Rodrigues, and N. Deligiannis, "Bayesian compressed sensing with heterogeneous side information," in *IEEE Data Compression Conference*, 2016.

[23] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, 2010.

[24] J. F. C. Mota, L. Weizman, N. Deligiannis, Y. Eldar, and M. R. Rodrigues, "Reference-based compressed sensing: A sample complexity approach," in *IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016.

[25] J. F. C. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. Rodrigues, "Dynamic sparse state estimation using $\ell_1 - \ell_1$ minimization: Adaptive-rate measurement bounds, algorithms and applications," in *IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015.

[26] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," in *IEEE Int. Conf. Acoust., Speech Signal Process.* IEEE, 2014, pp. 3330–3334.

[27] 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.

[28] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Compututational Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[29] 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.

[30] 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.

[31] D. B. West *et al.*, *Introduction to graph theory*. Prentice hall Upper Saddle River, 2001, vol. 2.

[32] [Online]. Available: https://www.lora-alliance.org/.