Compressive Sensing Process for Proficient Data Transmission in Wireless Sensor Networks

1 J.Dinesh & 2 M. Venkatesh Naik

1PG Student, Dept. Of Cse, Chiranjeevi Reddy Institute Of Engineering & Technology, Affiliated To Jntua, Andhra Pradesh, India
2Assistant Professor, Dept. Of Cse, Chiranjeevi Reddy Institute Of Engineering & Technology, Affiliated To Jntua, Andhra Pradesh, India

Abstract—
We focus on wireless sensor networks (WSNs) that perform data collection with the objective of obtaining the whole data set at the sink (as opposed to a function of the data set). In this case, energy efficient data collection requires the use of data aggregation. Whereas many data aggregation schemes have been investigated, they either compromise the fidelity of the recovered data or require complicated in-network compressions. In this paper, we propose a novel data aggregation scheme that exploits compressed sensing (CS) to achieve both recovery fidelity and energy efficiency in WSNs with arbitrary topology. We make use of diffusion wavelets to find a sparse basis that characterizes the spatial (and temporal) correlations well on arbitrary WSNs, which enables straightforward CS-based data aggregation as well as high fidelity data recovery at the sink. Based on this scheme, we investigate the minimum energy compressed data aggregation problem. We first prove its NP-completeness, and then propose a mixed integer programming formulation along with a greedy heuristic to solve it. We evaluate our scheme by extensive simulations on both real datasets and synthetic datasets. We demonstrate that our compressed data aggregation scheme is capable of delivering data to the sink with high fidelity while achieving significant energy saving.

Index Terms—Wireless sensor networks (WSNs); data collection; data aggregation; compressed sensing (CS); diffusion wavelets; energy efficiency

I. INTRODUCTION
Energy efficiency of data collection is one of the dominating issues of wireless sensor networks (WSNs). It has been tackled from various aspects since the outset of WSNs. This includes, among others, energy conserving sleep scheduling, topology control, mobile data collectors and data aggregation. While the first three approaches (and many others) focus on the energy efficiency of protocols, data aggregation directly aims at significantly reducing the amount of data to be transported, and it hence complements other approaches. Although data aggregation techniques have been heavily investigated, there are still imperfections to be improved on. Existing WSN applications fall into two classes depending on the information that is needed at the sink. The first class corresponds to cases where only simple function values of the data are needed at the sink (e.g., MAX/AVG/MED); we call this class “functional”. In that case, the aggregation functions only extract certain statistical quantities from the collected data and the original data are thus not fully recoverable. The second class of applications refers to cases where the sink wants to obtain the full data set. We call this class “recoverable”. The recoverability can be achieved by applying distributed source coding technique, such as Slepian-Wolf coding, to perform non-collaborative data compression at the sources. However, it is not exactly practical due to the lack of prior knowledge on the spatial data correlation structure. Collaborative in-network compression makes it possible to discover the data correlation structure through information exchange, but it either requires a simple correlation structure, or often results in high communication load that may potentially offset the benefit of recoverable aggregation technique. As a newly developed signal processing technique, compressed sensing (CS) promises to deliver a full recovery of signals with
high probability from far fewer samples than their original dimension, as long as the signals are sparse or compressible in some domain [1]. Although this technique may suggest a way of reducing the volume of the data to transmit over WSNs without the need for adapting to the data correlation structure [2], the complexity of the interaction between data routing and CS-based aggregation has postponed the development on this front until very recently [2]. Moreover, as these recent studies rely on conventional signal/image processing techniques (e.g., DCT, wavelets) to sparsify the collected data, they require regular WSN deployments (e.g., grids), which make them less practical. In this paper, we propose a new data aggregation technique derived from CS, and we aim at minimizing the total energy consumption when collecting data from nodes of an arbitrarily deployed WSN for an application requiring full data recovery. We make use of diffusion wavelets; this facilitates a high fidelity recovery of data sensed by nodes in an arbitrary network by allowing both spatial and temporal correlations to be naturally taken into account. To the best of our knowledge, we are the first to address the problem of CS based data aggregation and recovery in arbitrary networks. More specifically, we are making the following contributions:

• We propose a scheme that exploits compressed sensing for data aggregation.
• We employ diffusion wavelets to design the sparse basis for data recovery. This technique delivers high fidelity recovery for data collected in arbitrary WSNs by naturally taking into account spatial (and temporal) correlations.
• We show that, by choosing a proper sparse basis, we can partition a WSN into sub networks and apply compressed data aggregation to these sub networks individually. Compared with non-partitioning case, this significantly improves energy efficiency without sacrificing the fidelity of data recovery.
• We study, for our CS-based aggregation scheme, the minimum energy data gathering problem. We first provide the characterization of the optimal solutions and prove the hardness of the problem. Then both optimal and heuristic approaches are presented to solve the problem.
• We demonstrate the superiority of our compressed data aggregation scheme through experiments on extensive datasets: compared with non-aggregation, our proposal achieves significant energy saving while delivering data to the end with high fidelity.

II. RELATED WORK

Based on the compressive sensing theory, sensor nodes can compress sensory data without additional prior knowledge and recover these data with a relatively small number of measurements. These features make it appropriate for data aggregation in WSNs and may result in lower transmission cost, shorter acquisition time and less power consumption. These approaches can be classified into three categories: CS technique improvements, CS-based aggregation path optimizations, and applications in WSNs. A. CS Technique Improvements In [1], a universal compressive wireless sensing (CWS) scheme is proposed for single hop data gathering. The sensor nodes deliver the linear projections of readings to the fusion center through synchronized analog transmissions. Due to the difficulty to apply analog synchronization in large-scale sensor network, CWS is not applicable for multi-hop data gathering. Reference [1] presents the first complete CS-based data gathering scheme in large-scale WSNs. In [1], every sensor node compresses incoming data by summing M coded vector without prior knowledge. Therefore, the traffic load on the aggregation path remains the same regardless of the number of nodes on the path. Most CS-based data gathering schemes assume that the data transmissions are reliable, but in actual environment the transmissions are usually unreliable. Reference [1] presents an oversampled CS source coding scheme to handle transmission errors. Reference [1] proposes an adaptive data gathering scheme to deal with abnormal readings. In [1], the number of measurements can be adjusted according to the variation of the sensed data. However, neither discusses how to reduce the global communications. B. CS-Based Aggregation Path Optimizations To reduce unnecessary transmissions at the early stage by directly applying compressed sensing, Reference [1] proposes a hybrid-CS scheme to improve the performance, in which CS-based collection scheme is used only at nodes whose traffic is larger than a threshold. Reference [1] aims at minimizing the energy consumption of the hybrid-CS data gathering scheme. It proves that the optimization problem is NP-complete and gives a
greedy algorithm, called MECDA, to grow routing paths from the sink with polynomial time complexity by examining all candidate neighbouring nodes one by one at each round. However, when the network topology changes, this solution requires heavy computation to reconstruct the aggregation tree. Reference [1] focuses on finding a routing path that can maximize the lifetime of WSNs and minimize the delay of data gathering. Nevertheless, it also needs global network information and complex computation to create the tree structure. C. CS Applications in WSNs A growing number of researches aim to apply CS in other applications in WSNs. To investigate a CS-based approach for target detection and positioning in WSNs. Reference [2]and [3] propose a CS-based data persistence scheme in which sensed data are stored in the network until a mobile sink visits or a data collector query for the readings.

The fundamental assumption of in-network data compression is that sensor nodes have spatial correlations in their readings. According to where the spatial correlation is utilized, we can classify existing in-network data compression techniques into two categories.

Conventional Compression Conventional compression techniques utilize the correlation during the encoding process and require explicit data communication among sensors. Cristescu et al. [1] propose a joint entropy coding approach, where nodes use relayed data as side information to encode their readings. Again take the multi-hop route in Fig. 2 as an example. First, node s1 encodes its reading d1 into message p1 using H(d1) bits, where H(d1) is the entropy of d1. Then, when s2 receives p1, it encodes its reading d2 into message p2 using H(d2|d1) bits, where H(d2|d1) is the conditional entropy. Since d1 and d2 are correlated, H(d2|d1) is smaller than H(d2). Therefore, jointly encoded messages cost less bits than independently encoded messages. The above approach utilizes data correlation only unidirectionally. If data are allowed to be communicated back and forth during encoding, nodes may cooperatively perform transform to better utilize the correlation. Ciancio et al. [1] and A’cimovi´c et al. [2] propose to compress piecewise smooth data through distributed wavelet transform. In doing so, even nodes first broadcast their readings. Upon receiving the readings from both sides, odd nodes compute the high pass coefficients h(·). Then, odd nodes transmit h(·) back and even nodes compute the low pass coefficients l(·).Quantization of a group of readings to one representative value is another form of conventional compression. The clustered aggregation (CAG) technique [2] forms clusters based on sensing values. By grouping sensors with similar readings, CAG only transmits one reading per group to achieve a predefined error threshold. Gupta et al. [5] exploit a similar idea. In each round of data gathering, it only involves a subset of nodes, which is sufficient to reconstruct data for the entire network. There are two main problems with conventional compression techniques. First, the compression performance relies heavily on how the routes are organized. In order to achieve the highest compression ratio, compression and routing algorithms need to be jointly optimized. This has been proved to be an NP-hard problem [1]. Second, the efficiency of an in-network data compression scheme is not solely determined by the compression ratio, but also depends on the computational and communication overheads. However, joint entropy coding techniques perform complex computation in sensors, while transform based techniques require a large amount of data exchanges. Data aggregation is one of the major research topics for WSNs, due to its promising effect in reducing data traffic. Given the page limitation, we only discuss, among the vast literature, a few contributions that are closely related to our proposal. Applying combinatorial optimizations to data aggregation was introduced in [1], assuming an aggregation function concave in the input. Whereas [1] aims at deriving an algorithm with a provable bound (albeit arbitrarily large) for all aggregation functions, we are considering a specific aggregation function that is inspired by our CDA, and we propose fast near-optimal solution techniques for practical use. Involving the correlation structure of data, other types of optimal data aggregation trees are derived in [1]. However, as we explained in Sec. I, such approaches are too correlation structure dependent, hence not as flexible as our CS-based aggregation. Compressed sensing is a recent development in signal processing field, following several important contributions from Candes, Donoho, and Tao. It has been applied to WSN for single hop data gathering [2], but only a few proposals apply CS to multihop networking. In [2], a throughput scaling law is derived for the plain CS aggregation. However, as we pointed out in Sec. V-B,
plain CS aggregation is not an energy efficient solution. In addition, our results in [3] also demonstrate the disadvantage of plain CS aggregation in terms of improving throughput. [3] investigates the routing cost for CS aggregation in multi-hop WSNs where the sensing matrix is defined according to the routing paths. [2] draws the observation that the sensing matrix has to take the characteristics of the sparse domain into account. The routing-dependent or domain dependent design, unfortunately, contradicts the spirit of CS that sensing matrices can be random and easily generated. None of these proposals can recover CS aggregated data from arbitrarily deployed WSNs, which is one of the key issues that we address in this paper and [3]. Though we require reliable transmissions to be handled at a lower layer, an alternative solution is over-sampled CS source coding. Due to the inherent randomization of CS, packet loss is somewhat equivalent to reducing k at the coding end. Therefore, data recovery at the sink can still be performed (though bearing a larger error) in the face of packet loss.

III. COMPRESSIVE DATA GATHERING

The objective of compressive data gathering is twofold: compress sensor readings to reduce global data traffic and distribute energy consumption evenly to prolong network lifetime. Similar to distributed source coding, the data correlation pattern shall be utilized on the decoder end. Besides, compression and routing are decoupled and therefore can be separately optimized.

3.1 Data gathering

The intuition behind CDG is that higher efficiency can be achieved if correlated sensor readings are transmitted jointly rather than separately. We have given a simple example in Section I, showing how sensor readings are combined while being relayed along a chain-type topology to the sink. In practice, sensors usually spreads in a two-dimensional area, and the ensemble of routing paths presents a tree structure. a typical routing tree in which the sink has four children. Each of them leads a sub tree delimited by the dotted lines. Data gathering and reconstruction of CDG are performed on the sub tree basis. In order to combine sensor readings while relaying them, every node needs to know its local routing structure. That is, whether or not a given node is a leaf node in the routing tree or how many children the node has if it is an inner node. To facilitate efficient aggregation, we have made a small modification to standard ad-hoc routing protocol: when a node chooses a parent node, it sends a ”subscribe notification” to that node; when a node changes parent, it sends an ”unsubscribe notification” to the old parent.

IV. CONCLUSION

Leveraging on the recent development of compressed sensing, we have proposed a compressed data aggregation (CDA) scheme for WSN data collection in this paper. Our major contributions are twofold:

1) We have designed a proper sparse basis based on diffusion wavelets to achieve high fidelity recovery for data aggregated from arbitrarily deployed WSNs. We have developed this idea to allow for arbitrary network partitions and to integrate temporal correlations along with the spatial ones, which can significantly reduce energy consumption while maintaining the fidelity of data recovery.

2) We have investigated the minimum energy CDA problem by characterizing its optimal configurations, analysing its complexity, as well as providing both an exact solution (for small networks) and approximate solutions (for large networks). For performance evaluation, we have carried out extensive experiments on both synthetic datasets and real datasets. The results, on one hand, demonstrate that high fidelity data recovery can be achieved by properly designing the sparse basis; and on the other hand, validate the significant energy efficiency in data collection.

We have described in this paper a novel scheme for energy efficient data gathering in large scale wireless sensor networks based on compressive sampling theory. We believe this is the first complete design to convert the traditional compress-then-transmit process into a compressive gathering (compress-with-transmission) process to address the two major technical challenges that today’s large scale sensor networks are facing. In the development of the proposed scheme, we have carried out the analysis of capacity for wireless sensor network when compressive data gathering is adopted. We have shown that CDG can achieve a capacity gain of
N/M over baseline transmission. We have also designed ns-2 simulations to validate the proposed scheme when contention-based MAC is used. Furthermore, numerical studies based on real sensor data not only verified data sparsity in practical data acquisition, but also demonstrated the efficiency and robustness of the sensor data reconstruction with and without abnormal readings. It should be noted that successful application of CDG depends on the properties of sensor field. If sensor readings are not sparse in any known domain and in any proper order, CDG cannot achieve capacity gain because an important prerequisite of compressive sampling theory is missing. At the other extreme, when sensing data are sparse in the original domain, i.e. only a small fraction of sensors acquire nonzero readings, it would be more efficient to directly transmit these non-zero readings through multi-hop forwarding. CDG is not suitable for small scale sensor networks when signal sparsity may not be prominent enough and the potential capacity gain may be too small. CDG is also more effective for networks with stable routing structure. This is because frequent node failure or dynamic route change will lead to high control overhead that potentially cancel out the gain from data compression. We are currently investigating the extension of CDG to more challenging networking scenarios and the exploitation of fault tolerance of the compressive sampling principles to achieve more robust performance in sensor data gathering.

REFERENCES


