IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

Enhanced Data Gathering with Compressive Sensing in WSN using Dynamic Merge Compressive Algorithm

Sumaiah Barveen S[1], C Grace Padma MCA, M.Phil, MBA[2]

M.Phil Scholar, Dept of Computer Science, Rathnavel Subramanian College of Arts and Science, Coimbatore, India [1]

Associate Professor & HOD, Department of Computer Applications, Rathnavel Subramaniam College of Arts and Science, Coimbatore, India[2]

Abstract: Data gathering is one of most important functions provided by WSNs, where sensor readings have to be collected from sensor nodes to one or few data collection sinks node. Due to the fact that there may exist high correlations among these sensor readings, it is inefficient to directly deliver raw data to the destination(s).In this study the application of CS with random walks for data gathering in WSNs. The proposed system adopt the standard. Previous works focus on finding the movement patterns of each single object or all objects. This paper proposes an efficient distributed mining algorithm to jointly identify a group of moving objects and discover their movement patterns in wireless sensor networks. Afterward, a compression algorithm, called (Enhanced 2 phase and 2D) E2P2D is proposed, which utilizes the discovered group movement patterns shared by the transmitting node and the receiving node to compress data and thereby reduces the amount of delivered data. The enhance compression algorithm includes a sequence merge and an entropy reduction phases. In the sequence merge phase, a Merge algorithm is proposed to merge and compress the location data of a group of moving objects. The experimental results show that the proposed compression algorithm leverages the group movement patterns to reduce the amount of delivered data effectively and efficiently.

Keyword: Data gathering, Data Collection, Distributed Mining Algorithm, Compression Algorithm, E2P2D.

I. INTRODUCTION

This Wireless sensor networks have recently come into prominence because they hold the potential to revolutionize many segments of our economy and life, from environmental monitoring and conservation, to manufacturing and business asset management, to automation in the transportation and health care industries. The design, implementation, and operation of a sensor network requires the confluence of many disciplines, including signal processing, networking and protocols, embedded systems, information management and distributed algorithms. Such networks are often deployed in resource-constrained environments, for instance with battery operated nodes running un-tethered.

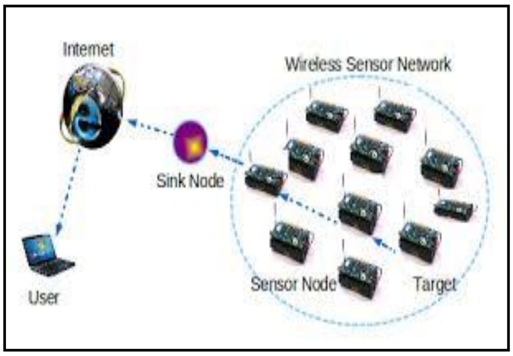

Fig 1.1 Wireless Sensor Network

In this paper, study the application of CS with random walks for data gathering in WSNs. We adopt the standard random walk algorithm to collect random measurements along multiple random paths. However, such an approach will lead to the non-uniform selection of measurements, which is different from uniform sampling in the traditional CS theory. It is still unknown whether such an approach can be used to recover sparse signals in a WSN scenario. To the best of our knowledge, the problem of data gathering with CS based on random walk algorithm has not been significantly investigated.

However, they focus on designing rate less code to exploit the correlation of the signals and use belief propagation decoding algorithm. In analysis the problem of compressive sensing for sparse signals over graph for network tomography applications. In their work, the authors model the network as a random graph G (n, p) in which an edge exists between any two nodes with probability p in the network within nodes. However, such a model is not appropriate to wireless sensor networks since connectivity in such networks is dependent on the distance between two nodes. Moreover, their work focuses on the case where the mixing time has an upper bound as the number of nodes n grows, which is different from our case.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

In this paper, we make the first step to understand why a CS based approach combined with random walks might work in a WSN scenario. Most importantly, we provide mathematical foundations for CS to use random walks in a more general wireless network from the perspectives of graph theory and CS theory. The goal of this paper are summarized as follows:

- Motivated by data gathering applications, we propose a simple measurement collection algorithm with random walks for compressive sensing.
- We provide mathematical foundations to use such an approach for understanding the intrinsic principles between CS theory and graph theory.
- We also demonstrate the effectiveness of the proposed scheme through simulations. The simulation results show that the proposed scheme can significantly reduce the communication cost for data gathering applications

II. RELATED WORKS

Wei Wang [1] describe the key idea is that the sparsity of the random projections greatly reduces the communication cost of pre-processing the data. The algorithm proposed in this study [1] allows the collector to choose the number of sensors to query according to the desired approximation error. The reconstruction quality depends only on the number of sensors queried, enabling robust refinable approximation. Suppose a wireless sensor network measures data which is compressible in an appropriate transform domain [2,3], so that n data values can be well-approximated using only k << n transform coefficients.

Giorgio Quer [2] Compressive Sensing (CS) shows high promise for fully distributed compression in wireless sensor networks (WSNs). In theory, CS allows the approximation of the readings from a sensor field with excellent accuracy, while collecting only a small fraction of them at a data gathering point. However, the conditions under which CS performs well are not necessarily met in practice. CS requires a suitable transformation that makes the signal sparse in its domain. Also, the transformation of the data given by the routing protocol and network topology and the sparse representation of the signal have to be incoherent, which is not straightforward to achieve in real networks.

In this work address the data gathering problem in WSNs, where routing is used in conjunction with CS to transport random projections of the data. They analyse synthetic and real data sets and compare the results against those of random sampling. In doing so, they consider a number of popular transformations and they find that, with real data sets, none of them are able to sparsify the data while being at the same time incoherent with respect to the routing matrix. The obtained performance is thus not as good as expected and finding a suitable transformation with good sparsification and incoherence properties remains an open problem for data gathering in static WSNs.

Liu Xiang [3] investigate the application of CS to data collection in wireless sensor networks, and they aim at minimizing the network energy consumption through joint routing and compressed aggregation. They first characterize the optimal solution to this optimization problem, then they prove its NP-completeness. They further propose a mixed integer programming formulation along with a greedy heuristic, from which both the optimal (for small scale problems) and the near-optimal (for large scale problems) aggregation trees are obtained. 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, which include, among others, energy conserving sleep scheduling [13], topology control, mobile data collectors and data aggregation [14]. Whereas the first three approaches (and many others) focus on the efficiency of networking techniques that transport the sensory data, data aggregation directly aims at significantly reducing the amount of data to be transported, and it hence complements other approaches and is deemed as the most crucial mechanism to achieve energy efficient data collection for WSNs.

Fatemeh Fazel [4] describe a power-efficient underwater sensor network scheme employing compressed sensing and random channel access. The proposed scheme is suitable for applications where a large number of sensor nodes are deployed uniformly over a certain area to measure a physical phenomenon. The underlying assumption is that most physical phenomena have sparse representations in the frequency domain. The network is assumed to have a Fusion Centre (FC) that collects the observations of sensor nodes and reconstructs the measured field based on the obtained measurements. The proposed method is completely decentralized, i.e., sensor nodes act independently without the need for coordination with each other or with the FC. During each frame, a Bernoulli random generator at each node determines whether the node participates in sampling or stays inactive during that sampling period. If selected, it measures the physical quantity of interest, e.g. temperature. A second random generator with a uniform distribution then picks a (random) delay for the node to send its data to the FC.

Weiyu Xu [5] motivated by network inference and tomography applications, they study the problem of compressive sensing for sparse signal vectors over graphs. In particular, they are interested in recovering sparse vectors representing the properties of the edges from a graph. Unlike existing compressive sensing results, the collective additive measurements they are allowed to take must follow connected paths over the underlying graph. For a sufficiently connected graph with n nodes, it is shown that, using O(k log(n)) path measurements, they are able to recover any k-sparse link vector (with no more than k nonzero elements), even though the measurements have to follow the graph path constraints. They further show that the computationally efficient ℓ1 minimization can provide theoretical guarantees for inferring such k-sparse vectors with O(k log(n)) path measurements from the graph. In operations of communication networks, they are often interested in inferring and monitoring the network performance characteristics, such as delay and packet loss rate, associated with each link. However, making direct measurements and monitoring for each link can be costly and operationally difficult, often requiring the participation from routers or potentially unreliable middle network nodes. Sometimes the responses from the middle network nodes are unavailable due to physical or protocol constraints. This raises the question of whether it is possible to quickly infer and monitor the network link characteristics from indirect end-to-end (aggregate) measurements. The problem falls in the area of network tomography, which is useful for network traffic engineering and fault diagnosis.

Compressive sensing is a new paradigm in signal processing theory, which challenges to sample and recover parsimonious signals efficiently. It has seen quick acceptance in such applications as seismology, error correction and medical imaging since the breakthrough works [20], although its role in networking is still limited. Its basic idea is that if an object being measured is well-approximated by a lower dimensional object in an appropriate space, one can exploit this property to achieve perfect recovery of the object.

III. DATA COMPRESSIVE PARADIGM

A. Network Construction

In this module, an object is defined as a target, such as an animal or a bird, which is recognizable and traceable by the tracking network. To represent the location of an object, geometric models and symbolic models are widely used. A geometric location denotes precise two-dimension coordinates; while a symbolic location represents an area, such as the sensing area of a sensor node or a cluster of sensor nodes, defined by the application.

B. Object Tracking

Object tracking is defined as a task of detecting a moving object's location and reporting the location data to the sink periodically at a time interval. Hence, an observation on an object is defined by the obtained location data. A picture box control is used as the 2D graph area in which circles are drawn such that one circle represents one sensor node and redrawn such that they are moving randomly in all directions with given random speed. The X and Y coordinates are also maintained at regular intervals and updated to the cluster head nodes. It is assumed that sensor nodes wake up periodically to detect objects. Using the timer controls, it is designed such that sensor nodes functions.

When a sensor node wakes up on its duty cycle and detects an object of interest, it transmits the location data of the object to its CH. Here, the location data include a timestamp, the ID of an object, and its location. Instead of forwarding the data upward immediately, the CH compresses the data accumulated for a batch period and sends it to the CH of the upper layer. The process is repeated until the sink receives the location data. To learn the significant movement patterns, Probabilistic Suffix Tree (PST) is adopted for it has the lowest storage requirement. PST is also useful and efficient in predicting the next item of a sequence. For a given sequence s and a PST T, the predict next algorithm is proposed which outputs the most probable next item, denoted by predict next(T,s).

C. EGMP Mine Algorithm

The EGMP Mine algorithm extracts the movement patterns from the location sequences by learning a PST for each object. In this module, a new similarity measure sim_p to compare the similarity of two objects is proposed. For each of their significant movement patterns, the new similarity measure considers not merely two probability distributions but also two weight factors, i.e., the significance of the pattern regarding to each PST. The similarity score simp of oi and oj based on their respective PSTs, Ti and Tj. The similarity score is equally divided into a threshold value so that the objects can be grouped in a single cluster if the similarity score falls within that range.

D. Compression Algorithm with Group Movement Pattern

To reduce the amount of delivered data, the 2P2D algorithm is proposed which leverages the group movement patterns derived to compress the location sequences of moving objects elaborately. The algorithm includes the sequence merge phase and the entropy reduction phase to compress location sequences vertically and horizontally. In the sequence merge phase, the Merge algorithm is proposed to compress the location sequences of a group of moving objects.

Since objects with similar movement patterns are identified as a group, their location sequences are similar. The Merge algorithm avoids redundant sending of their locations, and thus, reduces the overall sequence length. It combines the sequences of a group of moving objects by 1) trimming multiple identical symbols at the same time interval into a single symbol or 2) choosing a qualified symbol to represent them when a tolerance of loss of accuracy is specified by the application.

Therefore, the algorithm trims and prunes more items when the group size is larger and the group relationships are more distinct. Besides, in the case that only the location center of a group of objects is of interest, the approach can find the aggregated value in the phase, instead of transmitting all location sequences back to the sink for post-processing. To compress the location sequences for a group of moving objects, the proposed system processes the Merge algorithm.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

F. Advantages of the Proposed System

The paper proposes a novel compression algorithm to compress the location data of a group of moving objects with or without loss of information. It avoids transmitting unnecessary and redundant data by transmitting only the local grouping results to a base station (the sink), instead of all of the moving objects' location data. Using two phases - Enhanced 2D algorithm, the proposed system utilizes the discovered group movement patterns shared by the transmitting node and the receiving node to compress data. The proposed compression algorithm effectively reduces the amount of delivered data and enhances compressibility.

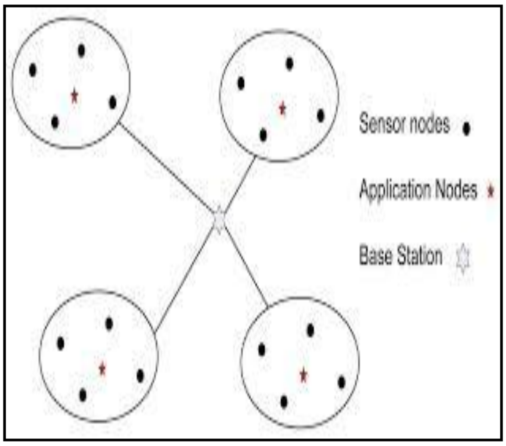

Fig 4.2 Cluster Formation

IV. APPROXIMATE OBJECT DATA COMPRESSION

Approximate data collection is a wise choice for long-term data collection in WSNs with constrained bandwidth. In many practical application scenarios with densely deployed sensor nodes, the gathered sensor data usually have inherent spatial-temporal correlations. For example, Fig. 1.2 shows the temperature readings of five nearby sensor nodes deployed in a garden more than 10 hours at night. The temperature readings recorded by the five nodes keep decreasing in the first 4 hours and then become stable in the next 6 hours, which exhibit apparent spatial and temporal correlations among themselves.

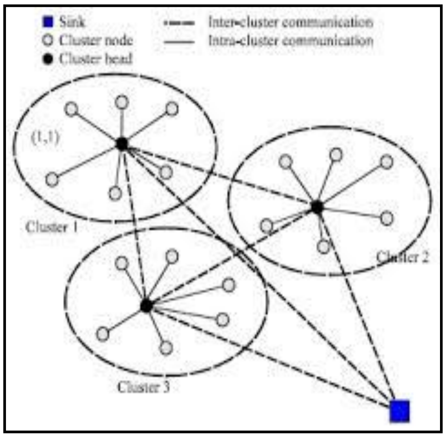

Fig 4.1 Cluster Compression

By exploring such correlations, the sensor data can be collected in a compressive manner within pre specified, application-dependent error bounds. The data traffic can be reduced at the expense of data accuracy. The granularity provided by such approximate data collection is more than sufficient, especially considering the low measuring accuracy of sensors equipped on the sensor nodes. Study on approximate data collection is thus motivated by the need of long-term operation of large-scale WSNs, e.g., the Green Orbs paper.

There are several factors to be considered in the design of an approach for approximate data collection. First, the data collection approach should be scalable.

In many real applications, sensor networks consist of hundreds or even thousands of sensor nodes. For example, Green Orbs has deployed 330 nodes and expects to deploy 1;000þ sensor nodes in a network.

In practice, in large WSNs, the information exchange between the sink and the related sensor nodes may consume considerable bandwidth and the acquisition of complete sensor data set of a WSN is too costly to be practical. Second, in approximate data collection, the spatial-temporal correlation model used for data suppression should be light-weight and efficient so as to meet the constraints on sensor node's memory and computation capacity. For densely deployed WSNs, many models can be used to describe temporal and/or spatial correlation of sensor data. But it is often nontrivial to build a light-weight correlation model to suppress spatial-temporal redundancy simultaneously.

Most of the existing models are too expensive, i.e., consuming a large amount of computing capacity or storage capacity, to be run on the existing sensor nodes. Some of them are too simple to contain enough information and ignores the trend of sensor readings, or only consider either temporal correlation or spatial correlation separately.

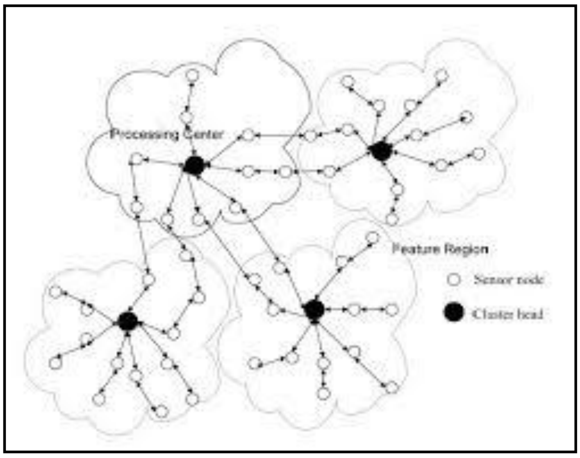

Fig 4.3 Data Compression

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

This paper approach shows that simplicity and efficiency can be achieved by exploiting implicit sensor node cooperation and elaborately distributing data processing tasks to sensor nodes. Third, the data collection scheme should be self-adaptive to environmental changes. Note that physical environmental changes are usually complex and hard to be modelled comprehensively with a simple estimation model. For long-term data collection, the approximate data collection scheme should be capable of automatically adjusting its parameters according to the environmental changes so as to guarantee its correctness.

The sink can estimate the sensor readings according to the model parameters updated by the cluster heads. This distributed data process scheme makes ADC can be easily applied to WSNs with different system scales. As the sensor network scale increases, ADC only needs to increase the number of clusters.

A.OBJECT TRACKING

In object tracking applications, many natural phenomena show that objects often exhibit some degree of regularity in their movements. For example, the famous annual wild-ebeest migration demonstrates that the movements of creatures are temporally and spatially correlated. Biologists also have found that many creatures, such as elephants, zebra, whales, and birds, form large social groups when migrating to find food, or for breeding or wintering. These characteristics indicate that the trajectory data of multiple objects may be correlated for biological applications.

Discovering the group movement patterns is more difficult than finding the patterns of a single object or all objects, because it is required to jointly identify a group of objects and discover their aggregated group movement patterns. The constrained resource of WSNs should also be considered in approaching the moving object clustering problem. However, few of existing approaches consider these issues simultaneously. On the one hand, the temporal-and-spatial correlations in the movements of moving objects are modelled as sequential patterns in data mining to discover the frequent movement patterns. The existing system has following disadvantages,

- Sequential patterns consider the characteristics of all objects.
- Sequential patterns lack information about a frequent pattern's significance regarding individual trajectories
- Sequential patterns carry no time information between consecutive items.
- Make them unsuitable for location prediction and similarity comparison.

B.COMPRESSION ALGORITHM

A proposed system clustering algorithm is proposed to find the group relationships for query and data aggregation efficiency. First, since the clustering algorithm itself is a centralized algorithm, the project further considers systematically combining multiple local clustering results into a consensus to improve the clustering quality and for use in the update-based tracking network. Second, when a delay is tolerant in the tracking application, a new data management approach is required to offer transmission efficiency, which also motivates this study. The project defines the problem of compressing the location data of a group of moving objects as the group data compression problem.

The paper formulates a moving object clustering problem that jointly identifies a group of objects and discovers their movement patterns. The application-level semantics are useful for various applications, such as data storage and transmission, task scheduling, and network construction. The thesis proposes an efficient distributed mining algorithm to minimize the number of groups such that members in each of the discovered groups are highly related by their movement patterns. Specifically, the E2P2D algorithm comprises a sequence merge and an entropy reduction phases. In the sequence merge phase, a Dynamic Merge Compressive algorithm is proposed to merge and compress the location data of a group of objects.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The following **Table 5.1** describes experimental result for existing system performance rate analysis. The table contains number of cluster, cluster size and number of compression data and average aggregated data details are shown

Table 5.1 Compression Size: Performance Rate

S No	Number Of Cluster	A	B	C	D	E	F
1	2 Cluster	73	56	72	74	72	65
2	3 Cluster	62	73	73	75	69	74
3	4 Cluster	69	70	76	65	74	69
4	5 Cluster	65	77	75	68	73	72
5	6 Cluster	69	75	71	64	65	68
6	7 Cluster	70	76	70	62	67	72
7	8 Cluster	72	69	68	75	74	69
8	9 Cluster	78	78	65	59	72	74
	No. of Aggregated data	558	574	570	542	566	563
	Average %	69.75	71.75	71.25	67.75	70.75	70.375

Table 5.2 Cluster: Compression Size

Object Cluster	Compression Size	AVG % Compression Size
Cluster A	580	72.5
Cluster B	597	74.62
Cluster C	578	72.25
Cluster D	557	69.62
Cluster E	579	72.37
Cluster F	569	71.12

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

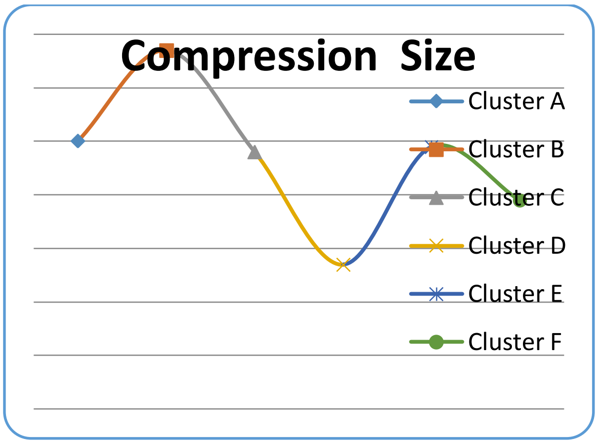

Fig 5.1 Compression Size

VI. CONCLUSION

Through this paper, the data management process becomes easy. All the day-to-day activities are assigned to them through browser interface. The administrator can view the contents in the server, select a file and download to the system wherever the administrator is working. Likewise, the files can be uploaded from the web page also very easy manner.

The new system eliminates the difficulties in the existing system. It is developed in a user-friendly manner. The system is very fast and any transaction can be viewed or retaken at any level. Error messages are given at each level of input of individual stages. Many of the time-tested practices and technologies for managing trust relationships in traditional enterprise IT environments can be extended to work effectively in both private and public clouds. The paper provides a best assistance in object tracking and merging two path sequences. The application become useful if the below enhancements are made in future. If the paper is designed as web service, it can be integrated in many network applications. The paper is developed such that above said enhancements can be integrated with current modules.

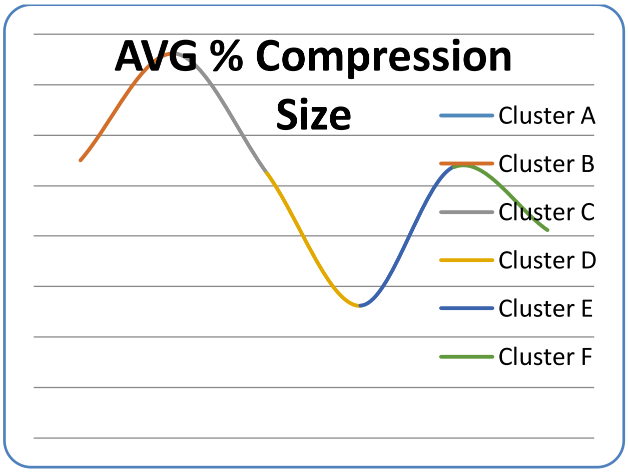

Fig 5.2Average Compression Size

REFERENCES

[1] W. Wang, M. Garofalakis, and K. Ramachandran, “Distributed Sparse Random Projections for Refinable Approximation,”Proc. ACM/IEEE 6th Int. Conf. Inf. Process. Sensor Network. (IPSN ’07), Apr. 2007.

[2] M. Vetterli and J. Kovaˇ cevi´ c. Wavelets and Subband Coding. Prentice Hall, Englewood Cliffs, NJ, 1995.

[3] S. Mallat. A Wavelet Tour of Signal Processing. Academic Press, San Diego, CA, 1999.

[4] E. Candes and T. Tao. Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies. IEEE Transactions on Information Theory, 52(12), pp. 5406-5425, December 2006.

[5] D. Donoho. Compressed Sensing. IEEE Transactions on Information Theory, 52(4), pp. 1289-1306, April 2006.

[6] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M.Zorzi, “On the Interplay between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks,”Proc.Inf. Theory Appl. Workshop (ITA ’09), Feb. 2009.

[7] D. Donoho, “Compressed sensing,” IEEE Trans. on Information Theory, vol. 52, no. 4, pp. 4036–4048, 2006.

[8] E. Cand`es and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?” IEEE Trans. On Information Theory, vol. 52, no. 12, pp. 5406–5425, 2006.

[9] E. Cand`es, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. on Information Theory, vol. 52, no. 2, pp.489–509, 2006.

[10] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, “Compressive Sensing for Networked Data: a Different Approach to Decentralized Compression,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 92–101, 2008.

[11] J. Liu, M. Adler, D. Towsley, and C. Zhang, “On optimal communication cost for gathering correlated data through wireless sensor networks,” in MOBICOM, 2006.

[12] L. Xiang, J. Luo, and A. Vasilakos, “Compressed Data Aggregation for Energy Efficient Wireless Sensor Networks,” Proc. 8th Ann. IEEE Comm. Soc. Conf. Sensor, Mesh and Ad Hoc Comm. And Networks (SECON ’11), June 2011.

[13] R. Subramanian and F. Fekri, “Sleep Scheduling and Lifetime Maximization in Sensor Networks: Fundamental Limits and Optimal Solutions,” in Proc. of 5th ACM IPSN, 2006.

[14] S. He, J. Chen, D. Yau, and Y. Sun, “Cross-layer Optimization of Correlated Data Gathering in Wireless Sensor Networks,” in Prof. of the 7th IEEE SECON, 2010.

[15] F. Fazel, M. Fazel, and M. Stojanovic, “Random Access Compressed Sensing for Energy-Efficient Underwater Sensor Networks,”IEEE J. Sel. Areas Comm., vol. 29, no. 8, pp. 1660-1670, Sept. 2011.

 DOI 10.17148/IJARCCE.2016.59102