Efficient Data Gathering with Compressive Sensing in Wireless Sensor Networks

Nikhil S. Hage¹, Mrs.S.Kayalvizhi²
¹ M-Tech (Embedded system technology).
²Assistant professor, Department of Electronics and Communication Engineering, SRM UNIVERSITY, CHENNAI-603203,INDIA

Abstract - In this paper, we study the problem of data gathering with compressive sensing (CS) in wireless sensor networks (WSNs). Unlike the conventional approaches, which require uniform sampling in the traditional CS theory, the wireless sensor networks are very useful in such area where the human being is unable to go and monitor. In such areas the continuous monitoring is require without failure of network or nodes but wireless sensors network have some energy constrain and cost constrain. We propose a random walk algorithm for data gathering in WSNs and for future purpose we can compare the result of random walk approach to chain based algorithm for data gathering which is traditional algorithm regularly used for network formation in WSNs. Random walk algorithm approach will absorb the constrain like path constrain for network efficiently and give non-uniform measurement. In this paper, from the perspectives of Compressive sensing theory and graph theory, we provide mathematical foundations to allow random measurements to be collected in a random walk based manner. We obtain random matrix from expander graph which will constructed by node measuremnet and for reconstructing we use l1 minimization theorem. Comparing two approaches with respect to their probability of data gathering, we also carry out simulation of both the scheme. Simulation result shows that our proposed scheme random walk approach can significantly reduce communication cost and reduce noise.

Key Words: Random walk, Compressive sensing, l1-minimization, Gaussian and binomial function, Data gathering.

1. INTRODUCTION

As we know, in the past few years, wireless sensor networks (WSNs) have been deployed in a wide range of application scenarios, such as battle field surveillance, environment monitoring, and security systems.

We consider data collection problem in large-scale WSNs. Sensor are deployed to monitor physical phenomena such as temperature, humidity and light over geometric area. Data gathering is one of most important functions provided by WSNs, where sensor will sense the data collect all data and transfer this data to one another i.e. from nodes to sink or sink to nodes. 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. Many techniques that attempt to reduce the traffic load have been developed, such as distributed compression algorithms [2], [3] and distributed source coding (DSC) approaches [3], [4].

However, the classical compression techniques for WSNs, [1], [2], [4], [5], [9], [18]. Typically associated with routing algorithms, impose high computation and communication overhead on sensor nodes. We have various kind of technique to study and apply to networks i.e. WSNs networks to collect data Compressive data gathering using sub-Gaussian random matrix which use term as opportunistic pipelining to data gather [8].forming chain-based networks to data gathering explain in chain-based protocol under compressive sensing framework [9]. In that paper the some protocol for energy efficient data gathering are explain such as LEACH [9] which is cluster-based method to collect data and another one is PEGASIS [9] (Power-Efficient Gathering In Sensor Information System) which is chain-based method to collect data in networks. Various other papers will discuss about application on data gathering with or without compressive sensing method.

In particular the compressive sensing method is now most emerging technique in the field of wireless network which will give more advantages than any other sampling theorem spatially Shannon sampling theorem or Nyquist theorem. as we know conventional approaches to sampling signals or image follow shannon’s theorem: the sampling rate must be at least the twice the maximum frequency present in the signal so called Nyquist rate. Main fact is that, this principle will only consider nearly all the signal acquisition protocol used in consumer audio and video electronics, medical imaging devices, radio receivers, and so on. For some signal such as images that are not naturally
Compressive sensing (CS) is a signal processing technique that allows the recovery of a signal from a limited number of measurements. This is particularly useful in scenarios where the signal is known to be sparse or compressible, meaning it can be represented by a few significant components. The main advantage of CS is that it can reduce the amount of data that needs to be collected and transmitted, which is crucial in environments where communication resources are limited, such as wireless sensor networks (WSNs).

### 2 Preliminaries and Basic Concepts

#### 2.1 Compressive Sensing Basics

Compressive sensing provides a new technique for signal acquisition and processing. As per the theory of CS, a sparse or compressible signal can be reconstructed with high probability from a small number of measurements, which is far smaller than the length of the original signal. Let's consider an n-dimensional signal vector $x = (x_1, \ldots, x_n)^T$. We say that the vector $x$ is perfectly $k$-sparse if it has at most $k$ ($k \ll n$) nonzero entries. For convenience, we emphasize and study $k$-sparse signals including perfect sparse and approximately sparse signals. Further, suppose that $x$ can be represented as $x = \sum_{i=1}^{n} \theta_i \Psi_i$ in some domain.
2.2 The Sensing Problem

In this section, we discuss sensing mechanisms of information about a signal f(t) and how it is obtained by

Linear functional recording the values

\[ y_k = (f, \phi_k), \quad k = 1, \ldots, m. \]  \hspace{1cm} (4)

We will consider correlations of the object we wish to acquire with the waveforms \( \phi_k(t) \). This is a standard setup. If the sensing waveforms are Dirac delta functions then \( y \) is a vector of sampled values of \( f \) in the time or space domain. If the sensing waveforms are indicator functions of pixels, then \( y \) is the image data typically collected by sensors in a digital camera. If the sensing waveforms are sinusoids, then \( y \) is a vector of Fourier coefficients. This is the sensing modality used in magnetic resonance imaging (MRI). Although one could develop a CS theory of continuous time/space signals, we restrict our attention to discrete signals \( f \in \mathbb{R}^M \). The reason for considering discrete signal is essentially twofold: first, this is conceptually simpler and second, the available discrete CS theory is far more developed we are then interested in under sampled situations in which the number \( m \) of available measurements is much smaller than the dimension \( n \) of the signal \( f \). Such problems are extremely common for a variety of reasons.

Due to these circumstances there are some basic questions we should consider. Is accurate reconstruction possible from \( m \ll n \) measurements only? Is it possible to design \( m \ll n \) sensing waveforms to capture almost all the information about \( f \)? And how can one approximate \( f \) from this information? Letting \( A \) denote \( m \times n \) sensing matrix with the vectors \( \phi^*1, \ldots, \phi^*m \) as rows (\( a^* \) is the complex transpose of \( a \)), the process of recovering \( f \) from \( y = Af \in \mathbb{R}^m \) is ill-posed in general when \( m \ll n \) there are infinitely many candidate signals \( f \) for which \( Af = y \). This will become easy by taking or imagine the realistic model of objects \( f \) which naturally exist.

2.3 Under sampling and signal recovery

This section will give information about recovery part of the CS theory. Ideally, we would like to measure all the coefficients of \( f \), but we only get to observe a subset of these and collect the data \([1]\)

\[ y_k = (f, \phi_k), \quad k \in M, \]  \hspace{1cm} (5)

Where \( M \subset \{1, \ldots, n\} \) is a subset of cardinality \( m \ll n \). With this information, we decide to recover the signal by \( l_1 \)-norm minimization: the proposed reconstruction \( f^* \) is given by \( f^* = \Psi x^* \), where \( x^* \) is the solution to the convex optimization program

\[ \min_{x \in \mathbb{R}^n} \|x\|_1 \text{ Subject to } y_k = (\phi_k, \Psi x), \quad k \in M. \]  \hspace{1cm} (6)
That is, among all objects \( f = \Psi x \) consistent with the data, we pick that whose coefficient sequence has minimal \( l_1 \) norm.

**THEOREM 1 [22]**

Fix \( f \in \mathbb{R}^n \) and suppose that the coefficient sequence \( x \) of \( f \) in the basis \( \Psi \) is \( S \)-sparse. Select \( m \) measurements in the \( \Phi \) domain uniformly at random. Then if

\[
m \geq C \cdot \mu^2(\Phi,\Psi) \cdot S \cdot \log n
\]

For some positive constant \( C \), the solution to (6) is exact with overwhelming probability. (It is shown that the probability of success exceeds \( 1 - \delta \) if \( m \geq C \cdot \mu^2(\Phi,\Psi) \cdot S \cdot \log(n/\delta) \). In addition, the result is only guaranteed for nearly all sign sequences \( x \) with a fixed support. [5]

![Fig1](a) A sparse real valued signal (b) its reconstruction from 60 Fourier coefficient by \( l_1 \) minimization. (c) The min. energy reconstruction obtained by \( l_2 \) norm.

### 3. RANDOM SENSING

Returning to the RIP, here we find sensing matrices with the property that column vectors taken from arbitrary subsets are nearly orthogonal. It’s better to have larger subset. This is where randomness re-enters the picture.

From available material we consider the following sensing matrices: i) form \( A \) by sampling \( n \) column column vectors uniformly at random on the unit sphere of \( \mathbb{R}^m \); ii) form \( A \) by sampling i.i.d. entries from the normal distribution with mean 0 and variance \( 1/m \); iii) form \( A \) by sampling a random projection \( P \) as in “Incoherent Sampling” and normalize: \( A = \sqrt{n/m}P \); and iv) form \( A \) by sampling i.i.d. entries from a symmetric Bernoulli distribution \( P(A_{i,j}= \pm 1/\sqrt{m}) = 1/2 \) or other sub-Gaussian distribution. With overwhelming probability, all these matrices obey the RIP (i.e. the condition of our theorem) provided that

\[
m \geq C \cdot S \log(n/S)
\]

It has also been proven that some sparse 0-1 matrices, such as the adjacency matrices of expander graphs, can satisfy the RIP-1 property. Using sparse matrices, the \( l_1 \)-minimization algorithm can also be used to recover \( k \)-sparse signals. Another in which each entry \( A_{ij} \) is i.i.d drawn according to

\[
A_{ij} = \sqrt{s} \begin{cases} +1 & \text{w.p.} \frac{1}{2s} \\ \frac{1}{\sqrt{s}} & \text{w.p.} \frac{1}{2s} \\ -1 & \text{w.p.} \frac{1}{2s} \\ 0 & \text{w.p.} \frac{1}{s} \end{cases}
\]

It has been shown that we can construct matrix from expander graph which will show the connection of all nodes connected during formation of network by random walk. This matrix will give the exact connection between nodes and graph will give detailed connection for example as shown below:
This matrix will give the connection such as first node is not connected and denoted by 0 respectively. Second node is 1 so this form connection with another node and so on. Further nodes are shown connected in diagram let \( G = (V, E) \) be an undirected graph with \( |V| = n \) and \( A \) be the Boolean matrix with \( m \times n \). If we perform Independent random walk let say \( m \) walks then each row of \( A \) can be seen as the characteristic vector of a subset vertices in \( V \) visited by the walk as shown in following figure. The nodes on left hand side of following figure corresponds to the signal coefficient set \( V \) and on right hand side corresponds to measurements set \( M \) as \( |M| = m \). For our project we take one example of random measurement matrix and based on the obtained matrix we form the bipertite graph only for the understanding purpose of connection in random walk and how it will form the wireless network within nodes and sink.

### 3.1 Theory on Random walk

According to the author of random walk and electric networks[13] give the information about random walk basics. Let \( G = (V, E) \) be a connected graph with \( n \) nodes and \( m \) edges. If we perform random walk on \( G \) and if we start at \( V_0 \) and up to the node \( V_t \) with \( t \)-steps then probability of connecting neighbour node is \( 1/d(V_t) \) and for other node in network the overall probability is also same as distribution

\[
P(t) = \text{Prob}(v_t = i);
\]

So
\[
p_{ij} = \begin{cases} 
1/d(i) & \text{if } ij \in E, \\
0 & \text{otherwise}.
\end{cases}
\]

#### A) Random walks on Finite networks

In one dimension network the probability will be \( \frac{1}{2} \) the example below illustrate the finite network.

![Fig3 (a): Random walk in one dimension](image)

In 2 - dimension network the probability will be \( \frac{1}{4} \) and example are shown below and the walker moves from \( x = (a; b) \) to each of the four neighboring points \( (a + 1; b), (a - 1; b), (a; b + 1), (a; b - 1) \) with probability \( 1/4 \).

![Fig3 (b): Random walk in two dimensions](image)

#### B) Gaussian distribution and Binomial distribution

[23] The graph of a Gaussian is a characteristic symmetric “bell curve” shape. The parameter ‘\( a \)’ is the height of the curve’s peak; ‘\( b \)’ is the position of the center of the peak and controls width of the “bell”. Gaussian function are widely used in statistics where they describe the normal distribution, in signal processing where they serve to define Gaussian filter and give probability function of signals. In image processing where two-dimensional
Gaussian distribution used for Gaussian blurs. Gaussian distribution will give the continuous distribution as output. Gaussian function arises by composing the exponential function with a concave quadratic function. This function will is in the Form of

\[ F(x) = a e^{-(x-b)^2/2\sigma^2} \]  

(8)

We are using this Gaussian distribution for finding exact probability from output. Similarly, we are using binomial distribution to convert that continuous distribution in discrete format. In probability theory and statistics, the binomial distribution with parameter \( n \) and \( p \) is the discrete probability distribution of the number of successes in a sequence of \( n \) independent yes or No experiments, each of which yield success with probability \( p \). The binomial distribution is frequently used model the number of successes in a sample of size \( n \) drawn replacement from a population of size \( N \). The Binomial Cumulative Distribution function which we used in finding probability is expressed as

\[ F(k;n,p) = \Pr(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \]  

(9)

The following diagram will show Gaussian and binomial distribution function:

![Gaussian and Binomial Distribution Diagram](image)

Fig4 (a): Gaussian distribution with variance function

Fig4 (b): Binomial distribution with probability 'p' and number of nodes 'n'.

4. PROBLEM STATEMENT

For formation of network we consider the \( m \) sensor nodes for initialize \( m \) independent random Walk. As we proceed as per the walk, each step one node will choose one of its neighbours and perform linear combination with the previous measurement of the neighbour. At the end of random walk, \( m \) random projection are generated and all will be send to sink through shortest path routine Strategy. So at sink we will get the constructed signal of all measurement and at sink we used recovery algorithm to construct original signal of measurement. Step for describe the operation of random walk is follows:

Step (1): at the beginning, \( i \) nodes let's say \( v_i \in V \) is uniform and randomly selected.

Step (2): after selected and collecting measurement form beginning node it will choose node \( v_j \) uniformly at random walk and adds previous and current measurement collected as

\[ x_j(1) = x_i(0) + x_j(0) \]

At the same time the node \( v_i \) decrements the length \( t \).

Step (3): Repeat the above steps up to \( t = 0 \) i.e. length become 0 and at the end the final constructed signal equation will be:

\[ x_p(t') = x_k(t' - 1) + x_p(t' - 1). \]

For \( p \) last node and \( k \) previous node in random walk. This way we can form the overall network and as explained above, that final signal will collect by sink and we can recover the original signal. For performing the random
walk with compressive sensing we used following methodology:

Taking matrix form of reading
Form Bipertile graph
Finding random walk probability
Finding exact probability

The above methodology will rough sketch for project. Here we are finding efficiency of our algorithm in wireless sensor network for finding network and dependent application with respect to other algorithm like chain Based algorithm or different routine strategies like:
1. Tree routing strategy
2. Cluster based strategy
3. Distributed strategy and
4. Random strategy

5. ANALYSIS

In this section we will prove that our strategy i.e. random walk strategy will give considerably good probability than other strategy and for future aspect we can compare well know and regularly used chain based strategy. The vertex visited by random walk will prove with the used of lemma 1:

Lemma 1. Let $B_t(v)$ be the even that the random walk $W$ starting at $u$ visits $v$ by time $t (t >= T)$. Then the probability $Pr (B_t(v))$ will be between $\frac{(1-\mu)^t}{2(n+2)^Cn}$ and $\frac{(1+\mu)^t}{cn}$ where $\eta = \frac{1}{(1-\mu)^k}$ and $c$ is a constant

The above lemma shows that the upper bound and the lower bound of the probability that a random walk visits a certain vertex are on the same order. the proof of above lemma is provided by supplementary file on online material and further information is in material random walk and electric network[13] and the introduction about graph will provided by random walk on graph[14].

6. NUMERICAL SIMULATION

In this section, we will study performance of our proposed algorithm through simulation. We consider the $n$ nodes which are connected randomly. The prescribed Distance for two nodes will be $\sqrt{\frac{3logn}{n}}$. For simulation we used MATLAB for simulation. We taken 100 nodes with in square area and perform random walk with sparse signal. For finding exact probability as explain above we used Gaussian and binomial distribution at simulation stage. With prescribe algorithm and flowchart we get output.

In simulation, we make bar graph to showed simplified solution and it’s easy to understand. For algorithm, we specify the length of spec steps and time steps that are ‘delta’ and ‘tau’ specify in program and as per user we can change the value of both parameter and see the different probability. For our simulation we stick to 0.1 and 0.01 values we give us probability value up to 0.1 and for different values of nodes and our parameter, we can observe different probability graph. For future purpose, we can compare this different probability result to distinguish different algorithm.Fig6 will give us the graph with exact probability. We used binomial function to find the exact probability. We can find mean value distance between two bars, maximum distance between two nodes and RMS distance between two nodes. Simulation will consider for the sparse and uniformly random numbers of nodes and form network with our scheme Fig7 shown original signal which is sense by the different nodes. Here we take sample signal from available source and using compressive sensing we mixed the original signal with one randomly generated matrix. We have our original signal in the form of $m \times n$ matrix form and at the sink node we use recovery method to recover the original signal by
simulation we got exact Original signal so we use compressive sensing for our scheme very efficiently.

7. CONCLUSIONS

In this paper, we studied the problems in wireless sensor network when we actually build networks. Actual problem with data gathering in networks and we discuss the compressive sensing technique. random walk approach which is more applicable for most of the WSN applications like data gathering, dissemination and recovery. We studied the details of random walk and approach to apply in data gathering with compressive sensing. Simulation result will show that the sparse signal can be reconstructed with probability which will change according to random sparse signal and this will go up to maximum of 0.1. The exact probability with mean distance 0.02 will give the brief probability in case of n=100. Hence this will give us that all signals can be reconstructed with random walk strategy whose average probability will go up to 0.1. This will show that the random walk algorithm will give high probability of data gather. This will also reduce the cost of the network as probability increases. As we get the average probability of random walk approach in WSN network as 0.1 i.e. 90% this will show that this approach have more chance to establish the network efficiently and maintain continuously. The SNR ratio will up to 25% shows efficient network in presence of noise and finally the compressive sensing used with recovery theorem to obtain original signal.

REFERENCES


[4] Xuangou Wu, Yan Xiong, Mingxi Li and Wenchao Huang “Distributed Spatial-Temporal Compressive Data Gathering for Large-scale WSNs” School of Computer Science and Technology, University of Science and Technology of China, Hefei, China State Key Laboratories of Transducer Technology, Shanghai, Email: wxgou@mail.ustc.edu.cn, yxiong@ustc.edu.cn, keeper@gmail.com, wchuang@ustc.edu.cn, IEEE, 2013

Compressive Sensing

Emmanuel J. Candès and Michael B. Wakin

Xiao-Yang Liu, Yamin Zhu, Linghe Kong, Cong Liu, Yu Guy, Athanasios V Vasilakoss, "CDC:Compressive Data Collection for Wireless Sensor Networks" Shanghai, JiaoTong University, China Singapore University of Tech. and Design, Singapore University of Texas at Dallas, USA University of Western Macedonia, Greece

Xiaohan Yu and Seung Jun Baek "Compressive data aggregation in wireless sensor networks using sub-Gaussian random matrices" College of Information and Communication, Korea University Anamdong, Seongbukgu,Seoul,Korea,136713 Email :{riccardo.masiero,giorgio.quer,rossi,zorzi}@dei.unipd.it,IEEE,2009


Xiaohan Yu and Seung Jun Baek "Compressive data aggregation in wireless sensor networks using sub-Gaussian random matrices" College of Information and Communication, Korea University Anamdong, Seongbukgu,Seoul,Korea,136713 Email :{riccardo.masiero,giorgio.quer,rossi,zorzi}@dei.unipd.it,IEEE,2009


[7] Xiao-Yang Liu, Yamin Zhu, Linghe Kong, Cong Liu, Yu Guy, Athanasios V Vasilakoss, “CDC:Compressive Data Collection for Wireless Sensor Networks” Shanghai, JiaoTong University, China Singapore University of Tech. and Design, Singapore University of Texas at Dallas, USA University of Western Macedonia, Greece


[9] Dongfeng Xie, Qianwei Zhou, Jianpo Liu, Baoping Li and Xiaobing Yuan "A chain-based data gathering protocol under compressive sensing framework for wireless sensor networks" Wireless Sensor Network Laboratory Shanghai Institute of Microsystem and Information Technology shangai, China dfhsie @126.com Dongfeng Xie, Qianwei Zhou University of Chinese Academy of Sciences Beijing,China, Email:zhuangxyan@gmail.com,International Conference on Computational and Information Sciences, 2013

[10] Zhuang Xiaoyan "Wireless Sensor Networks based on Compressed Sensing" School of Automation Engineering, University of Electronic Science and Technology of China chengdu, China, Email: zhuangxyan@gmail.com,IEEE, 2010


[12] "Compressive Sensing", rice university dsp.rice.edu/cs


[15] David L. Donoho, Michael Elad "Optimally Sparse Representation in General (non-Orthogonal) Dictionaries via I1 Minimization" Department of Statistics,Room 128, Sequoia Hall, Stanford University, Stanford, CA 94305, USA, Email: donoho@stat.stanford.edu and Department of Computer Science (SCCM), Stanford University, Stanford, CA 94305-9025, USA

[16] Richard Hartley and Frederik Schaffalitzky “L∞ minimization in geometric reconstruction problems” National ICT Australia The Australian National University Oxford University IEEE,2004


[23] "Gaussian distribution and binomial distribution" Wikipedia.