

DATA AGGREGATION AND SECURED ROUTING IN WIRELESS SENSOR NETWORKS

Thesis

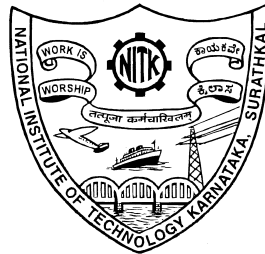
Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

DEEPA PUNEETH

(Reg. No. EC14F03)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING**

**NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA
SURATHKAL, MANGALORE - 575025**

January 2021

DECLARATION

I hereby *declare* that the Research Thesis entitled **DATA AGGREGATION AND SECURED ROUTING IN WIRELESS SENSOR NETWORKS** which is being submitted to the *National Institute of Technology Karnataka, Surathkal* in partial fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy in department of Electronics and Communication Engineering* is a *bona fide report of the research work carried out by me*. The material contained in this thesis has not been submitted to any University or Institution for the award of any degree.



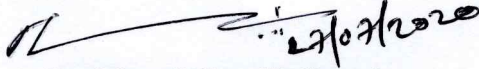
DEEPA PUNEETH
Register No.: EC14F03
Department of Electronics and
Communication Engineering

Place: NITK - Surathkal

Date: 27-07-2020

CERTIFICATE

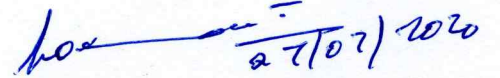
This is to *certify* that the Research Thesis entitled **DATA AGGREGATION AND SECURED ROUTING IN WIRELESS SENSOR NETWORKS** for dissertations submitted to the **National Institute of Technology Karnataka, Surathkal** by **DEEPA PUNEETH** (Register Number: EC14F03) as the record of the research work carried out by her, is *accepted* as the *Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **DOCTOR OF PHILOSOPHY**.



Dr. Muralidhar Kulkarni
Research Guide
Professor
Dept. E and C Engg.
NITK Surathkal - 575025

Dr. Muralidhar Kulkarni
Professor
Dept. of Electronics & Communication Engineering
National Institute of Technology Karnataka, Surathkal
Srinivasnagar, Mangalore-575 025, KARNATAKA, INDIA

(Signature with Date and Seal)



Dr. T. Laxminidhi
H.O.D
Professor
Dept. E and C Engg.
NITK Surathkal - 575025

प्राध्यापक एवं विभागाध्यक्ष / PROF & HEAD
ई एवं सी विभाग / E & C Department
एन आई टी के, सुस्तकल/NITK, Surathkal
मंगलूर / MANGALORE - 575 025

(Signature with Date and Seal)

ACKNOWLEDGEMENTS

Over the course of my study at NITK, I met many people who have motivated, guided and helped me to learn many things. I would not be able to complete this work without them. During this opportunity, I would like to express my deepest gratitude and appreciation to them.

I am highly indebted to my research guide, **Prof. Muralidhar Kulakarni**, Dept. of E&C, for his guidance, constant supervision and for motivating me to complete my research work. He helped me throughout by providing necessary information and pushing me forward to finish the research work.

I am thankful to **Prof. M. S. Bhat** and **Prof. U. Shripathi Acharya**, the former Heads of the Dept. of E&C and to **Prof. T. Laxminidhi**, the Head of the Dept. of E&C for providing me this opportunity to continue my studies and for providing all the facilities required for the research work.

I am thankful to my RPAC members, **Dr. Ashvini Chaturvedi** and **Prof. G. Ram Mohana Reddy**, for their suggestions and questions during my seminars. I appreciate all my friends for their help and making my stay at NITK memorable.

I would also like to thank the entire teaching and non teaching staff of Dept. E&C, NITK, for all their help.

Last but not least, I am extremely grateful to my family who have been always supportive and encouraging me to pursue my studies.

DEEPA PUNEETH

January, 2021

ABSTRACT

In large scale Wireless Sensor Networks (WSNs) the amount of data generated is enormous. Data gathering in an energy efficient way is one of the important phenomenon. The nodes in WSNs are randomly deployed, the data emerging from these nodes are highly correlated either spatially or temporally. The data has to be processed efficiently before it reaches the Base Station (BS) by using an efficient routing algorithm as well as data aggregation methods. The data aggregation scheme should employ simple encoding since the sensor nodes are battery operated. The proposed method discusses about a data aggregation scheme using Compressive Sensing (CS) technique which makes use of correlation among the sensor nodes. Using CS we can preserve the information contained in a signal through linear projections and recover the signal using reconstruction algorithm.

Ensuring energy efficiency, data reliability and security is important in WSNs. A combination of variants of the cryptographic secret sharing technique and the disjoint multipath routing scheme is an effective strategy to address these requirements. While Shamir's Secret Sharing (SSS) provides the desired reliability and information theoretic security, but it is not energy efficient. Alternatively, Shamir's Ramp Secret Sharing (SRSS) provides energy efficiency and data reliability, but it is only computationally secure. We argue that both these approaches can suffer from a Compromised Node (CN) attack when a minimum number of nodes are compromised, and propose a new scheme, which is energy efficient, provides data reliability, and is secure against a CN attack. Proposed method, which we call Split Hop AES (SHAES) is highly energy efficient, is independent of the underlying routing scheme and provides Semantic Security, which helps in resisting CN attacks.

TABLE OF CONTENTS

| | |
|---------------------------------------------------------------|------------|
| ACKNOWLEDGEMENTS | i |
| ABSTRACT | ii |
| TABLE OF CONTENTS | iii |
| LIST OF FIGURES | v |
| LIST OF TABLES | ix |
| NOTATIONS | x |
| ABBREVIATIONS | xi |
| 1 INTRODUCTION | 1 |
| 1.1 Introduction to Wireless Sensor Networks (WSNs) | 1 |
| 1.2 Data Aggregation in WSNs | 2 |
| 1.3 Compressed Sensing fog data aggregation | 3 |
| 1.3.1 Temporal correlation | 3 |
| 1.3.2 Spatial correlation | 3 |
| 1.4 Security in WSNs | 4 |
| 1.5 Problem formulation | 5 |
| 1.6 Objectives | 6 |
| 1.6.1 Objective 1 | 6 |
| 1.6.2 Objective 2 | 7 |
| 1.6.3 Objective 3 | 7 |
| 1.7 Organization of the Thesis | 7 |
| 2 LITERATURE REVIEW | 9 |
| 2.1 Data Aggregation using CS | 9 |

| | | |
|----------|--------------------------------------------------------------------|-----------|
| 2.1.1 | Related work on CS in WSNs | 9 |
| 2.2 | Data aggregation using DCS | 13 |
| 2.3 | Energy-efficient and reliable data collection in WSNs | 14 |
| 3 | DATA AGGREGATION USING COMPRESSIVE SENSING IN WSNs | 17 |
| 3.1 | Introduction | 17 |
| 3.2 | Compressive Sensing (CS) | 18 |
| 3.2.1 | Signal acquisition and method of reconstruction in CS based system | 18 |
| 3.2.2 | Reconstruction model | 19 |
| 3.2.3 | Compressive sensing in Wireless Sensor Networks | 22 |
| 3.3 | Routing Protocols | 24 |
| 3.3.1 | Impact of using adjustable transmit power levels | 24 |
| 3.3.2 | Cluster based Node-disjoint Multi path routing | 26 |
| 3.4 | Sparse vector reconstruction approaches | 27 |
| 3.4.1 | Convex Relaxation | 28 |
| 3.4.2 | Greedy iterative pursuits | 28 |
| 3.5 | Data aggregation using CS for improved network lifetime | 31 |
| 3.5.1 | System Model | 32 |
| 3.5.2 | Results and Analysis | 33 |
| 3.5.3 | Experimental results of CS recovery based on Greedy Algorithms. . | 36 |
| 3.6 | Concluding Remarks | 40 |
| 4 | DISTRIBUTED COMPRESSIVE SENSING FOR WSNs | 45 |
| 4.1 | Introduction | 45 |
| 4.2 | Intra and Inter correlation effects | 46 |
| 4.3 | Distributed Compressive Sensing. | 46 |
| 4.3.1 | Models based on joint sparsity | 47 |
| 4.3.2 | Recovery of Jointly sparse signals | 48 |
| 4.3.3 | Separate recovery using OMP | 49 |
| 4.3.4 | Recovery using Simultaneous- OMP | 49 |
| 4.4 | Results and analysis | 51 |
| 4.4.1 | DCS for multi-channel EEG | 60 |

| | | |
|----------|----------------------------------------------------------------------|-----------|
| 4.5 | Conclusion | 61 |
| 5 | ENERGY EFFICIENT, SECURE AND RELIABLE DATA COLLECTION IN WSNs | 63 |
| 5.1 | Introduction | 63 |
| 5.2 | Secret Sharing Schemes - Overview | 65 |
| 5.3 | Proposed work | 66 |
| 5.3.1 | Near-Sink CN Attack | 66 |
| 5.4 | SHAES scheme | 69 |
| 5.5 | Security analysis of the SHAES: | 72 |
| 5.6 | SRSS and SHAES Combination | 73 |
| 5.6.1 | SRSS+SHAES Security Analysis | 74 |
| 5.6.2 | Energy Efficiency and Reliability Analysis of SRSS+SHAES | 77 |
| 5.7 | conclusions | 83 |
| 6 | CONCLUSIONS AND FUTURE SCOPE | 85 |
| 6.1 | Conclusions | 85 |
| 6.2 | Future Scope | 87 |
| | REFERENCES | 89 |
| | Publications based on the thesis | 97 |
| | BIO-DATA | 99 |

LIST OF FIGURES

| | | |
|------|-----------------------------------------------------------------------------------------------------|----|
| 1.1 | Applications of WSNs | 1 |
| 3.1 | Signal/Image acquisition in CS. | 18 |
| 3.2 | Humidity signal in acquired and transformed domain. | 19 |
| 3.3 | Subspaces containing sparse vector in R_3 | 21 |
| 3.4 | l_2 minimization | 21 |
| 3.5 | l_1 minimization | 21 |
| 3.6 | Network scheme with Compressive sensing Hu and Hao (2012) | 23 |
| 3.7 | Network life improvement in LEACH using CS | 34 |
| 3.8 | Network life improvement in M-TRAC using CS | 34 |
| 3.9 | Summary of dead nodes in LEACH and M-TRAC | 35 |
| 3.10 | Comparison of Reconstruction error using l_1 Regularization and OMP | 36 |
| 3.11 | Reconstruction error for temperature and humidity values using l_1 -norm regularization | 36 |
| 3.12 | Reconstruction of temperature and humidity values using l_1 -norm regularization | 37 |
| 3.13 | CS recovery using OMP. | 38 |
| 3.14 | Relative error using OMP. | 38 |
| 3.15 | CS recovery using RMP. | 39 |
| 3.16 | Relative error using RMP. | 39 |
| 3.17 | CS recovery using StOMP. | 40 |
| 3.18 | Relative error using StOMP. | 40 |
| 3.19 | CS recovery using SP. | 41 |
| 3.20 | Relative error using SP. | 41 |
| 3.21 | CS recovery using greedy methods. | 42 |
| 3.22 | Relative error using greedy methods. | 42 |
| 3.23 | CS recovery using greedy methods for temperature data. | 43 |
| 3.24 | CS recovery using greedy methods for humidity data.. . . . | 43 |

| | | |
|------|-------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1 | Separate recovery using OMP, convergence in case of L=8. | 50 |
| 4.2 | Separate recovery using OMP, convergence in case of L=16. | 50 |
| 4.3 | Joint recovery using YALL1 L=8, by considering different values of, K_c and K_l | 52 |
| 4.4 | Joint recovery using Somp L=8, by considering different values of, K_c and K_l | 52 |
| 4.5 | Joint recovery using YALL1 ,for varying values of L and sparsity | 53 |
| 4.6 | Joint recovery using SOMP,for varying values of L and sparsity | 53 |
| 4.7 | Joint recovery using OMP, for varying values of L and sparsity | 54 |
| 4.8 | Joint recovery using YALL1 ,for L=1,4,8,32,64 | 54 |
| 4.9 | Joint recovery using Somp ,for L=1,4,8,32,64 | 55 |
| 4.10 | Joint recovery using YALL1 and separate recovery using OMP ,for L=2,8,32 | 55 |
| 4.11 | Joint recovery using SOMP and separate recovery using OMP ,for L=2,8,32 | 56 |
| 4.12 | Joint recovery using SOMP and separate recovery using OMP, for L=1, 2, 4, 8, 16, 32 | 56 |
| 4.13 | Joint recovery using SOMP by considering real data-I,M=75 | 57 |
| 4.14 | Joint recovery using SOMP by considering real data-I,M=300 | 57 |
| 4.15 | Joint recovery using SOMP by considering real data-II,M=75 | 58 |
| 4.16 | Joint recovery using SOMP by considering real data-II,M=950 | 58 |
| 4.17 | Joint recovery using SOMP by considering Temperature data (outdoor) N = 1024 | 59 |
| 4.18 | Joint recovery using SOMP by considering Humidity data (outdoor) N=1024 | 60 |
| 4.19 | Joint recovery using SOMP by considering real data-III,M=175 | 61 |
| 5.1 | Overview of SRSS and SH-AES combination process | 65 |
| 5.2 | Centralized sink deployment | 67 |
| 5.3 | Corner sink deployment | 68 |
| 5.4 | Near-Sink CN attack on Centralized Sink deployment under different power levels | 69 |
| 5.5 | Near-Sink CN attack on Corner Sink deployment under different power levels | 70 |
| 5.6 | SHAES graphical representation | 71 |
| 5.7 | Graphical representation of achieving the Secret sharing and SHAES combination | 74 |
| 5.8 | Graphical representation of retrieving the original message from the Secret sharing and SHAES combination at the BS | 75 |

| | | |
|------|----------------------------------------------------------------------------------------------------------------------|----|
| 5.9 | Reliability analysis of SRSS and SSS schemes | 77 |
| 5.10 | Computational overhead analysis in terms of number of multiplication operations for different approaches | 78 |
| 5.11 | Computational overhead analysis in terms of number of multiplication operations using only SRSS | 79 |
| 5.12 | Polynomial evaluations for different data sizes using various SRSS parameters | 79 |
| 5.13 | Computational overhead analysis in terms of number of multiplication operations using SRSS+AES | 80 |
| 5.14 | Computational overhead analysis in terms of number of multiplication operations using SRSS+SHAES | 81 |
| 5.15 | Computational overhead analysis in terms of number of multiplication operations using optimized SRSS+SHAES | 82 |
| 5.16 | Communication overhead analysis in terms of data size expansion using SSS and SRSS schemes | 82 |

LIST OF TABLES

| | | |
|-----|---------------------------------------------------------------------------------------------|----|
| 2.1 | A summary of existing works | 15 |
| 3.1 | Chipcon cc2420 transceiver supported power levels and its power consump- tions | 25 |
| 3.2 | Simulation Parameters | 33 |
| 5.1 | Consolidated cryptanalysis attack complexity on 4 round AES | 73 |
| 5.2 | Consolidated analysis of different approaches | 83 |

NOTATIONS

| SYMBOL | MEANING |
|-------------------|--------------------------------------------------------------------------------------|
| \dot{D} | Data size |
| a_0 | Secret data used in the SSS evaluation |
| a_i | Random data used in the SSS evaluation. |
| C | Cipher text space |
| D | Decryption rule |
| E | Encryption Rule |
| $E_{TX}(N, d)$ | Energy consumed for transmitting 'N'bit message |
| $E_{RX}(N)$ | Energy consumed for receiving 'N'bit message |
| $\Phi_{inter}(m)$ | Inter node correlation |
| $\Phi_{intra}(m)$ | Intra node correlation |
| K | Key space |
| k | key choice |
| m_i | <i>i</i> th message |
| Φ | Measurement/sensing matrix |
| n | Number of shares |
| P | Plain text Space |
| p | Plain text choice |
| r | Random Choice |
| S | Shares |
| $S(m_i)$ | Set of all possible shares generated using share generation function for m_i |
| t | Required number of shares needed for Reconstruction of secret data using SSS |
| t_0 | Secret data used in the SRSS evaluation |
| t_i | Random data used in the SRSS evaluation |
| t_2 | Required number of shares needed for reconstruction of secret data using SRSS scheme |
| $x [N]$ | Sparse input signal of length N |
| \hat{x} | Estimated sparse signal |
| y | Cipher text choice |
| $y[M]$ | compressed vector of length M |
| Z_C | Common sparse component |
| Z_I | Common innovation component |

ABBREVIATIONS

| Abbreviation | Expansion |
|--------------|------------------------------------------------|
| AES | Advanced Encryption Standard |
| AK | AddroundKey |
| BP | Basis Pursuit |
| BS | Base Station |
| CH | Cluster Head |
| CN | Compromised Node |
| CP | Chosen Plaintext |
| CS | Compressed Sensing |
| DCS | Distributed Compressed Sensing |
| DCT | Discrete Cosine Transform |
| EEG | Electro Enceleno Graph |
| GF | Galois Field |
| JSM | Joint Sparsity Model |
| KP | Known Plaintext |
| MC | MixColumns |
| NIST | National Institute of Standards and Technology |
| OMP | Orthogonal Matching Pursuit |
| RMP | Residual Minimization Pursuit |
| RIP | Restricted Isometric Property |
| SB | SubByte |
| SHAES | Split Hop Advanced Encryption Standard |
| SOMP | Simultaneous OMP |
| SP | Subspace Pursuit |
| SR | ShiftRows |
| SSS | Shamirs threshold Secret Sharing |
| SRSS | Shamir Ramp Secret Sharing |
| St-OMP | Stagewise Orthogonal Matching Pursuit |
| WSN | Wireless Sensor Network |

CHAPTER 1

INTRODUCTION

1.1 Introduction to Wireless Sensor Networks (WSNs)

The decision making process of human being is carried out based on the observation by collecting the surrounding information. A Wireless Sensor Network (WSN) mimics this human intelligence but in a larger scale, which can be utilized for different applications. Figure.1.1 shows several applications of WSN.

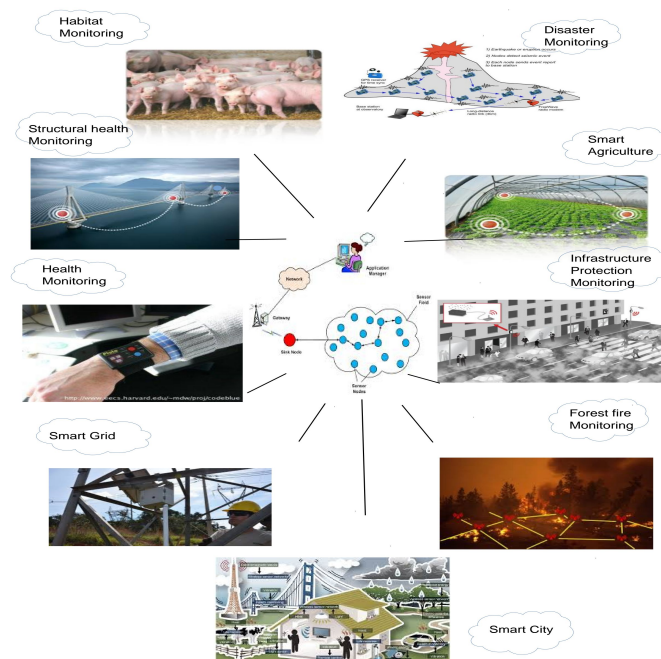


Figure 1.1: Applications of WSNs

WSNs consist of miniaturized nodes termed as sensor nodes which have limited power, as they are battery operated. The tiny nodes are composed of a sensor and mote. The physical parameters such as temperature, humidity, light etc are picked up by the sensor. Mote is accountable for communication as well as computation.

The ultimate goal of WSNs is to gather information on environment factors or object conditions on behalf of human being. To achieve this goal, wireless sensor devices have been

designed and characterized as follows : *First*, they should be realized at low cost. *Second*, they should have self-organizational capability. *Third*, they should be implemented as small as possible. *Last*, but most important, they should have long lifetime [Kim et al. \(2015\)](#).

This chapter has been organized in the following sections. Section 1.2 provides a general introduction of data aggregation in WSN's. Section 1.3 gives an insight into Compressed Sensing (CS) for data aggregation in WSNs. Section 1.4 gives details about how the Distributed Compressive Sensing (DCS) helps to reduce data transmission. Section 1.5 elaborates the need for security in WSN's. Finally in 1.7, a brief outline of the contribution and the organization of this thesis is given.

1.2 Data Aggregation in WSNs

The significant purpose of data aggregation schemes in WSNs, is to collectively accumulate the data in an energy efficient way. The collected data has to be processed before arriving at the Base Station (BS). At the BS, necessary action is taken depending on the information collected by these tiny sensors deployed in the region of interest. For certain applications the data outcome is really massive as well as redundant. The redundancy is introduced because of the correlation, which exists among the collected data. The end objective at the BS is to take necessary action, from the sensors with minimal amount of collected data, which in return enhances the life span of the network. Collection, processing and delivery of the sensor data are to be managed in an efficient way. When the purpose to achieve is, an energy-efficient data collection without compromising the fidelity of the recovered data then data, aggregation methods play a very important role.

Sensing, data processing and transmission are the three different phases of a sensor node which consume energy. But data processing usually consumes lesser power than the communication module. Data delivery from the nodes to the BS happens in a multi-hop manner. The consumption of energy, of a particular node may lead to network disconnections leading to the data unreachable to the BS [Jung et al. \(2011\)](#); [Marcelloni and Vecchio \(2008\)](#). Thus, there is a need to aggregate the data before transmission for the betterment of the entire network.

Few data aggregation schemes involve only simple function values of the sensed data. In certain applications, we do not expect full recovery of the sensed data. In such cases the aggregate module only extracts certain quantities from the collected data [Madden et al. \(2002\)](#). By

using Slepian-wolf coding [Slepian and Wolf \(1973\)](#), the original data can be exactly recovered; nevertheless prior knowledge of the spatial data correlation is required. Several distributed coding algorithms have been developed which involve inter-sensor communication overhead that affects the power consumption of the involved nodes. In distributed coding, only limited progress has been achieved, the direct implementation would need huge look-up tables, and approaches combining pre or post processing of the data to remove spatial correlations combined with Slepian wolf coding [Slepian and Wolf \(1973\)](#) appear to have limited applicability [Duarte et al. \(2006\)](#).

1.3 Compressed Sensing fog data aggregation

Data aggregation scheme for a WSN, must minimize the information complexity without the loss of fidelity of the acquired data. The outcome of such aggregation scheme reflects in the energy consumption of the sensor nodes which in turn enhances the life span of the entire network. The objective can be achieved by taking correlation properties into account either spatial, temporal or both.

1.3.1 Temporal correlation

In WSNs, few applications require periodic data update, for example, event tracking needs the data to be monitored as well as periodic transmission. The periodic data may be correlated depending on the correlation present between the periodic data. The amount of correlation between successive sensor measurements defines the degree of temporal correlation. The redundancy of the network is determined by the degree of correlation among the periodic measurements.

1.3.2 Spatial correlation

Certain WSN applications require the sensor nodes to be deployed densely and spatially. In that case few sensors in the deployed region record similar phenomenon, in that case the data from these sensors are spatially correlated. As these sensors are randomly deployed there may be few sensors which fall in the same vicinity. The overlap of the coverage area of these sensors, introduces redundancy into the network which further degrades the lifetime of the entire network.

WSN equipped for an environmental, habitat monitoring collects information like temperature, humidity, pressure, light intensity etc, which may be correlated either spatially or temporally. While designing a WSN, one has to give importance to criteria such as energy consumption, lifetime, delay etc. As these sensor nodes are battery powered, energy consumption is a primitive factor while designing the network. The process of efficient collection and transmission of the sensed data over the network for final decision at the sink, along with expanding the life span of the network can be done through Compressive Sensing. The same can be achieved by exploring the correlations among the sensed data.

1.4 Security in WSNs

The sensor networks deployed for mission-critical tasks, such as military applications must employ certain security needs during its design. In wireless system, it is easy to extract information, which results in a leakage of information as well as chance of eavesdropping and packet inflow by an adversary. Thus security to WSN to ensure secrecy and privacy of the data has to be properly addressed. One of the major internal attacks is, node compromise, by which an adversary can inject an internal attack. A Compromised Node (CN) attack is an attack in which an adversary compromises a certain subset of nodes to passively intercept data packets traversing the compromised nodes [Lou and Kwon \(2006a\)](#); [Liu *et al.* \(2012a\)](#). Attacks by a Compromised node happens in the following ways:.

- It can steal information from the encrypted/non encrypted data which is forwarded by the compromised node.
- It can convey false message to the network.
- It can give false information about a normal node as a Compromised node.
- It can breach routing by initiating routing attacks such as manipulating the routing table, selective forwarding etc.
- It may collude with other compromised node and interrupts its normal networking function.

In addition to security issues, Reliability also becomes important, particularly in multi-path routing that is often achieved with the help of data redundancy. Typically, reliability is achieved in multi-path routing by creating multiple copies of the same data and routing them

via different paths. When security is combined with reliability in multi-path routing, creating copies of data increases the chances of an adversary accessing the data, unless some security mechanism is employed (like encryption). A common method for combining reliability and security in multi-path routing is to split the data based on secret sharing schemes then send the shares on different paths to reach the BS. In order to achieve greater security, previous works have used the approach of dispersing the shares randomly and then sending the data towards the BS. Original data is reconstructed only when the required number of shares reaches the BS. Higher the dispersion of shares, higher the associated communications and thus, higher the communication energy drain. Even after investing more communication energy in dispersion, degree of security achieved may be small because all shares must converge at the BS. Therefore, the dispersion of shares may not completely solve the security problem when multi-hop communication routing is used and when the shares converge to single BS. Therefore, we opted not to disperse the data for security purposes and thereby reduced the communication energy drain. Instead of dispersion, the approach followed in this work invests a small amount of computation energy to achieve better security over the entire network including the area near the BS.

1.5 Problem formulation

In WSNs the sensor nodes are randomly deployed, thus the chances of existence of redundancy among the nodes are very high. Due to this redundancy there is unnecessary exhaustion of energy of the sensor nodes. One of the major issues in WSN is the energy efficiency, as these tiny sensor nodes are battery operated in which the battery can be neither replaced nor charged. Thus there should be a mechanism which evaluates the correlation, and finally reduces the redundant data traversing in the network which finally improves the network lifetime. The data aggregation method must involve simple procedure at the encoding side so that the sensor nodes must not expel much of its energy. A method in which the complexity must be moved to the BS which does not pose energy saving constraint.

Aging of WSN can be minimized by, compressing the data of individual sensor node, which cumulatively serves the purpose. Due to random deployment of sensor nodes there exists intra and inter correlations among the sensor nodes. Data aggregation using Compressive Sensing can be used to explore the intra correlation. Further the data can still be aggregated by exploring

intra as well as inter correlations, which can further enhance the lifetime of the network.

The sensors are usually deployed in remote and hostile environment, which are susceptible to internal/external attacks. The aggregated data, traverses in the network through multi-hop manner and finally reaches the BS, where finally the data is reconstructed. But if the node or group of nodes gets compromised the data can not be reconstructed accurately. As compared to the traditional networks WSNs faces more security issues as they are deployed in hostile environment.

In order to achieve secure reliable communication in WSN, the data is split into multiple shares using secret sharing schemes [Lou and Kwon \(2006a\)](#). Finally the shares need to converge at the BS, thus the nodes around the BS will be pruned to be the prime location for data compromise. The adversary can compromise a certain number of nodes for the passive interception of data i.e CN attack.

1.6 Objectives

We propose following research problems towards the research topic Efficient Data Aggregation and Secured Routing in WSN. The first two objectives deal with data aggregation and the third one with secured routing.

1.6.1 Objective 1

A data aggregation scheme is proposed, which deals with data compression and reconstruction based on Compressive Sensing (CS), which uses correlation between and within nodes.

The performance metrics namely the lifetime of the network, the throughput and the reconstruction error will be evaluated by considering Low Energy Adaptive Clustering Hierarchy (LEACH) and Multi Threshold adaptive Range Clustering (MTRAC) algorithms. The compressed data at the BS is recovered using greed based method. Evaluation of the available greedy methods for data recovery and a comparison of the greedy recovery methods considering synthetic and real data are carried out.

1.6.2 Objective 2

Another data aggregation scheme is proposed exploring the intra and inter correlations, through the concept of Joint Sparse Models (JSM) to reduce the amount of redundant data into the WSN.

A comparison of joint recovery procedure with separate recovery using Simultaneous Orthogonal Matching Pursuit (S-OMP) and Orthogonal Matching Pursuit (OMP) algorithms has been attempted along with validation of the results with indoor and outdoor data sets.

1.6.3 Objective 3

A new scheme that is energy efficient, reliable, and secure against CN attacks is proposed by combining Shamir's Ramp Secret Sharing (SRSS) and a round-reduced AES cipher, which we call split hop AES (SHAES).

Ensuring energy efficiency, data reliability, and security are important issues in WSNs. A combination of variants from the cryptographic secret sharing technique and the disjoint multipath routing scheme is an effective strategy to address these requirements.

1.7 Organization of the Thesis

In chapter 2, a detailed literature survey of the various methods employing CS for data aggregation in WSNs is provided. It elucidates an insight into the various protocols and algorithms proposed related to the proposed objective. It also includes a detailed survey related to security issues in WSNs, in which a detailed survey of the possible threats and solutions to them have been summarized.

Using CS, we can preserve the information contained in a signal through linear projections and recover the signal using reconstruction algorithm. Introduction to CS along with the mathematical background has been presented in Chapter 3. Routing plays an important role in achieving energy efficiency in a network. We employ CS for clustered single hop network by considering LEACH and M-TRAC protocols, we evaluate the lifetime of the network with/without CS. For different compression ratios the error has been calculated by considering the real data set. Further we investigate the easy implementation of greedy based methods for sparse signal recovery. Comparison of the greedy based methods with synthetic and real

data has been presented. The detailed study of CS and its implementation for WSN has been presented in Chapter 3.

Chapter 4 discusses the Joint sparse models and Joint recovery techniques in WSNs. Along with simulations, validation of the joint recovery procedures has been presented. Explanation regarding how the reduction of the compressed vector length can be achieved by exploring Temporal and Spatial correlations is presented. The simulations have been carried out using greedy based methods. The results of joint recovery and separate recovery have been presented which shows if there exists correlation among the sensors DCS is a favorable solution.

In Chapter 5, we discuss about Shamir's Secret Sharing (SSS) and Shamir's Ramp Secret Sharing (SRSS) its advantages and drawbacks and propose a method to combat Compromised Node (CN) attack.

Finally, Chapter 6 provides concluding remarks of the thesis and future scope.

CHAPTER 2

LITERATURE REVIEW

2.1 Data Aggregation using CS

Data gathering in an energy efficient way is one of the important requirements in WSNs. With incorporation of CS, as a data aggregation scheme the information contained in the signal is safely maintained through its projections which can be reconstructed later. Inclusion of CS for an energy efficient routing technique further enhances the lifetime of the network. The dimensionality reduction at the transmitter is carried out using a measurement matrix. At the receiver data is recovered using reliable CS recovery methods. Depending upon the need of the application, suitable recovery methods can be applied. The reconstruction errors vary depending upon the reconstruction methods. Greedy based recovery methods are more popular, because of its low complexity and low implementation cost. The success rate of any reconstruction method depends on the sparsity of the data. Identification of correct basis is an essential requirement as the signal may not be sparse in the acquired domain. The following section lists several WSN architectures using CS, as a data aggregation method.

2.1.1 Related work on CS in WSNs

[Zheng et al. \(2015\)](#) in his paper, Data Gathering with Compressive Sensing in Wireless Sensor Networks: A Random Walk Based Approach, proposed a random walk algorithm for data gathering in WSNs. The paper provides mathematical foundations to allow random measurements to be collected in a random walk based manner. Simulation results show that the proposed scheme can significantly reduce the communication cost compared to existing schemes using dense random projections and sparse random projections.

[Zhu et al. \(2015\)](#) proposed an energy efficient Data Gathering Scheme (DGS) for unreliable WSN using CS, which addresses packet loss problem in CS based aggregation. DGS-CS consists of 3 procedures. i) Procedure of the Source Node (PSN) ii) Procedure of the Intermediate Node (PIN) iii) Procedure of the Sink Node (POS) Performance comparison of DGS-CS

and TRS through numerical analysis has been carried out. Energy consumption of TRS and DGS-CS have been derived by defining Energy Saving efficiency (ESE), Performance analysis of Traditional Routing Scheme (TRS) and DGS-CS is carried out by considering WSN application such as temperature or humidity capture (data packets have high degree of correlation).

[Qin and Yin \(2015\)](#) proposed a robust sparsity estimation method in CS. In this paper a greedy algorithm that uses relative threshold to estimate the sparsity is proposed. Results show that estimation methods do not require the reconstruction of the whole signal and they do not rely on the power of the signal. Comparison of the estimation with traditional methods with different measurements and SNR scenarios is presented. When the measurements are large enough both methods present almost same performance. Nevertheless, proposed method shows 20% improvement when the measurements gets reduced than traditional method.

[Xing et al. \(2015\)](#) in this paper, Energy-Balanced Data Gathering and Aggregating in WSNs: A Compressed Sensing Scheme proposes Energy-balanced data Gathering and Aggregating (EDGA) scheme that integrates a clustering hierarchical structure with the CS. Design of a data reconstruction algorithm is based on orthogonal matching pursuit theory. Results shows that for the sparse network settings, proposed EDGA scheme achieves an improvement by 15.9% and 30.6% in terms of the network lifetime compared with Multi Channel singular Spectrum Analysis (MSSA) and Energy Efficient Information Collection (EEIC) algorithms, In dense and mediated network settings, EDGA scheme achieves the enhancement in network lifetime by an amount of 25.1% and 76.6%, 21.1%, and 54.2% . Claims scheme achieved better reconstruction accuracy with less than 20% error in face of 90% data missing probability.

[Liu et al. \(2015\)](#) proposed The Method of data aggregation for WSNs based on LEACH-CS. Here, the Formation LEACH protocol is used for the cluster formation. Gaussian random matrix is used to perform linear compression of the sensor data at each cluster head. Results show that the energy consumption at each node is one of critical issues. Reconstruction of the data at the BS is done considering, compression ratio of 0.4, 0.6 and ,0.8. Errors of reconstructed data using Total Variation (TV) method are 0.0007, 0.0001, 0.00009 respectively. TV matching method is claimed to be better compared to Orthogonal Matching Pursuit (OMP) and optimal OMP.

[Rossi et al. \(2015\)](#) in their paper, Evaluating the Gap Between Compressive Sensing and Distributed Source Coding in WSN presented a comparative performance analysis of CS and

DSC in terms of reconstruction error vs energy requirements. Authors have compared Temporal Correlation (TC) based algorithms (LTC, DCT) against Source Coding based (DSC) and CS-based techniques. If the correlation statistics are unknown, CS is deemed a valid compression approach as it often outperforms competing algorithms and, in the worst cases it performs in between temporal and spatial correlation-based compression.

[Caione *et al.* \(2014\)](#) in their paper, Compressive Sensing Optimization for Signal Ensembles in WSNs presented an investigation of the two frameworks on sparsity and compressibility of multidimensional signals and signal ensembles, Distributed compressed sensing (DCS) and Kronecker compressive sensing (KCS). Authors have exploited the inter-signals correlations present in WSNs data-sets to achieve a better compression factor. A better reconstruction quality under energy constraints is achieved. However a trade-off between recovery quality and energy spent in compression is needed. DCS and KCS schemes have been compared. From the results the recovery complexity for KCS can be seen, which infers that DCS is preferred for signal recovery in WSNs when CS is used for medium-sized networks.

[Xu *et al.* \(2013\)](#) proposed a power-efficient Hierarchical data aggregation scheme using CS in WSNs. Integration of a multi-resolution hierarchical structure with CS is presented to optimize the amount of data transmitted. Authors proposed a multiple compression threshold, which adapts based on the cluster sizes at different levels. Paper presents simulations of the SNR graph for the proposed HDACS for networks of sizes 300,400,500,600,700. It is seen that the performance of HDACS method is independent of network size. Cost comparison chart of HDACS with Plain CS (PCS) and Hybrid CS (HCS) is presented. HDACS method gives the best energy efficiency especially for nodes working as cluster heads.

[Xiang *et al.* \(2013\)](#) in their paper Compressed Data Aggregation: Energy-Efficient and High-Fidelity Data Collection presented a data aggregation using CS that achieves both recovery fidelity and energy efficiency in WSNs. Diffusion wavelets are used to find a sparse basis. A minimum-energy data gathering problem has been proposed (MECDA) Investigation of minimum-energy CDA problem providing both an exact solution (for small networks) and approximate solutions (for large networks) is presented. Authors have designed a proper sparse basis based on diffusion wavelets to achieve high-fidelity recovery for data aggregated from arbitrarily deployed WSNs. Comparison plot of CDA with plain CS is presented in the paper. Comparison of CDA against non-aggregation and bench marking MECDA_GREEDY

for large-scale networks and comparison of non-aggregation and CDA with/without network partition is done to validate the better energy efficiency of CDA with network partition.

[Fragkiadakis *et al.* \(2013\)](#) proposed a joint compressed-sensing and matrix-completion for efficient data collection in WSNs. Minimizing the data to be transmitted to the sink is done by applying the compressed sensing principles. Missing information due to packet loss is efficiently recovered using the matrix completion theory. Using CS sensors compress the temperature measurements using one out of three possible compression ratios (25%, 50%, and 75%). The compressed measurements are transmitted to the sink using a suitable protocol over UDP. The transmitted packet rate is varied so as to create an average packet loss in WSN that varies from 10% to 80%, with a step of 10%. Each experiment is repeated for 50 times.

[Quer *et al.* \(2012\)](#) in their paper Sensing, Compression, and Recovery for WSNs (SCoRe1): Sparse Signal Modeling and Monitoring Framework proposed a sparsity model that allows the use of CS for the online recovery of large data sets in real WSN scenarios, exploiting Principal Component Analysis (PCA) to capture the spatial and temporal characteristics of real signals. Bayesian analysis is utilized to approximate the statistical distribution of the principal components and to show that the Laplacian distribution provides an accurate representation of the statistics of real data. SCoRe1 accommodates diverse interpolation techniques, either deterministic or probabilistic, and embeds a control mechanism to automatically adapt the recovery behavior to time varying signal statistics, while bounding the reconstruction error. Authors claim that the proposed method is also robust to unpredictable changes in the signal statistics.

[Wang *et al.* \(2010\)](#) proposed a CS based random routing for multi-hop WSNs. The paper focuses on a novel random routing scheme for efficient data gathering. The basic approach of CS, recent technical advancements and their applications are presented. The paper proposes a random routing method executed with CS. The paper presents a comparison Random routing (RR)-CS with i) Sparse random sampling with CS SRS-CS ii) Dense sampling with CS (DS-CS). The following are the inferences in the paper a) SRS-CS is not suitable for data gathering with deficient sampling b) (RR-CS) is efficient for data gathering, performance is much better than that of SRS-CS, and is very close to that of DS-CS, when the number of measurements is relatively large. The energy consumption of SRS-CS and that of RR-CS is nearly the same. But DS-CS is much higher than those of the other two schemes.

Cao *et al.* (2008) in their paper Data Aggregation and Recovery in Wireless Sensor Networks Using Compressed Sensing addresses QoS issues by considering packet loss and energy dissipation. A CS-oriented data aggregation technique for the multi-hop topology has been presented. Results of recovered lost data using CS (OMP) are presented. Paper gives simulation results based on Recovery via the RIP properties.

2.2 Data aggregation using DCS

The inter-signal and intra-signal correlations are explored in DCS through the concept of joint sparse models. Under DCS there are few Joint sparse models which fits into particular situations pertain to WSN. Reconstruction of these models are done using Joint recovery procedures. In the literature we can find several recovery procedures. In DCS the data is reduced based on the intra and inter correlations present among the data from the sensor nodes. Compared to plain CS, in DCS it needs lesser measurements to reconstruct the data at the receiver.

Related work on data aggregation using DCS

A rich literature is available for collection of data ensemble in WSNs, most of the methods exploit the correlation (intra/inter) among the collected data. The basic idea behind DCS theory can be found in Baron *et al.* (2009), in which primary focus is on, compressing the vector length, which further effects the communication cost of the signal to be transmitted. It also explains the different joint sparsity models, relating the models with practical scenario and modeling the framework through suitable joint sparsity model. With graphical model and proof of theorems it also analyses the theoretical bound on measurement rates which is essential to guarantee the perfect recovery of the signal through the compressed sparse signal.

Liu *et al.* (2018) proposes common-innovation subspace pursuit (CISP), to estimate the common and innovation support sets separately, in order to minimize the reconstruction error and computing time. Sundman *et al.* (2011) explored joint sparsity using joint OMP and joint SP, but not experimented by considering real data. The work described in paper by Duarte *et al.* (2006) explains DCS for WSN which can be widely applicable in sensor network environment. By considering different sensor network datasets, implemented joint sparsity model to recover the sensor signals. Even though the signals are not perfectly sparse the Joint Sparse Model (JSM) provides a better approximation to explore the intra/inter correlations which exists in the

collected data from sensors.

[Alippi *et al.* \(2013\)](#) described an aggregation method for WSN, making use of temporal and spatial data dependencies present among the sensor nodes. In this paper these dependencies are used to reconstruct the missing data, from the sensors. Authors in [Wimalajeewa and Varshney \(2017\)](#) proposed a DCS method which is based on the covariance information of the uncompressed samples but did not experiment with real data.

Implementation of a multi-channel EEG monitoring, based on CS is explained in the paper by [Djelouat *et al.* \(2017\)](#). Recovery of the multi-channel signals through greedy based system is presented. [Caione *et al.* \(2013\)](#) explained two framework DCS and Kronecker Compressive Sensing (KCS) to reduce the amount of data, and to improve the network lifetime.

2.3 Energy-efficient and reliable data collection in WSNs

In the literature, there are various contributions towards security in WSNs. [Liu *et al.* \(2013\)](#) proposes an authenticated group key agreement (AGKA) protocol, and demonstrates how it can tackle node replication and Sybil attacks.

[Zhou \(2013\)](#) discusses efficient and secure routing protocol based on encryption and authentication. This method involves encryption of all communicated packets. A few similar security management methods can be found in [Liu *et al.* \(2013\)](#), which is based on trust-based management [Pan *et al.* \(2013\)](#), signcryption [Gu *et al.* \(2013\)](#), key pre-distribution scheme, and so on.

In [Liu *et al.* \(2009\)](#), reliability is achieved through re transmissions. [Dong *et al.* \(2016\)](#) proposed a routing scheme called Reliability and Multipath Encounter Routing (RMER), to achieve reliability and energy efficiency.

The first major contributions to the secure reliable data collection of sensor networks started with H-SPREAD proposed by [Lou and Kwon \(2006a\)](#), which used Samir's Secret Sharing (SSS) scheme to generate multiple shares of the data. Additionally, a hybrid multipath scheme was used to route the shares. However, the achieved security was low, because fixed multipaths were used to send the data. In addition, the presence of an adversary near the BS was not considered in their approach.

The work by [Shu *et al.* \(2010a\)](#) addresses the shortcomings of H-SPREAD by using the randomized and highly dispersive nature of routing. This approach increased security compared to H-SPREAD with the help of random dispersion.

Network lifetime and security were jointly considered in [Liu *et al.* \(2012a\)](#) with a combination of randomized and deterministic multipath routing; however, the approach was very specific to one particular type of deployment strategy, with respect to the deployment of nodes (circular) and the BS (center of the network area).

Table 2.1: A summary of existing works

| Works | Core objective | Assumption of secure area around BS | Type of secret sharing used | Encryption used | Security achieved |
|--------------------------------------|-------------------------------------------------------------|-------------------------------------|-----------------------------|-----------------|------------------------------------------------------------------------|
| Anfeng Liu et.al (2012) | Secure and energy efficient reliable data collection | Yes | SSS | No | Medium (CN attack is possible and random dispersion of shares) |
| Tao Shu et.al (2010) | Secure Data Collection | Yes | SSS | No | Medium (CN attack is possible and random dispersion of shares) |
| Wenjing Lou and Younggoo Kwon (2006) | Secure and reliable data collection | explicitly not mentioned | SSS | No | Low (CN attack is possible and no random dispersion of shares) |
| Ching-Fang Hsu et.al (2011) | Secure group communications | explicitly not mentioned | SRSS | No | Low (CN attack is possible with lesser number of shares (SRSS)) |
| Our approach | Secure and energy efficient reliable data collection | No | SRSS | Yes | High (achieves semantic security and CN attack is not possible) |

The objective of the paper by [Challal *et al.* \(2011\)](#) was to achieve fault tolerance with the help of the secure and efficient disjoint multipath routing strategy. The authors used data duplication and the Information Dispersal Algorithm (IDA) to create multiple data for routing. Similarly, this work assumed the perimeter area around the BS to always be secured. None of

the previous approaches considered the possibility of adversaries being near the BS, which is the prime location for obtaining maximum information from the complete network area. If an adversary compromises enough nodes to obtain the threshold shares, then security is lost. The work by [Claveirole *et al.* \(2008\)](#) addresses securing the data from aggregator node compromise, making use of secret sharing and multi path routing.

A brief consolidated comparison of the previous works that relate to our proposed core objective is presented in [Table 2.1](#).

In [Perrig *et al.* \(2002\)](#), the challenges in the security design of sensor networks are explored. The notion of semantic security in the area of sensor networks was used in their Secure Network Encryption Protocol (SNEP) design. Additionally, the authors emphasized that their first choice was the use of the AES block cipher algorithm; however, due to constraints in sensor node memory at that time, they opted for the RC5 algorithm. The need for semantic security in sensor networks to avoid information leakage via eavesdropping has been reported in [Shaheen *et al.* \(2007\)](#). The work in [Hsu *et al.* \(2011a\)](#) considers an ideal linear multi secret sharing (SRSS) scheme in order to provide secure and energy efficient group communications in wireless mesh networks. This approach enhanced energy efficiency, but could not overcome the CN attack problem.

CHAPTER 3

DATA AGGREGATION USING COMPRESSIVE SENSING IN WSNs

3.1 Introduction

The main objective of sensors is to sense the physical changes in the area of interest, which may include environmental monitoring, security applications, health monitoring etc. These sensors finally have to transmit the observations to the control unit usually the Base Station (BS), in order to perform necessary action. As these sensors are randomly deployed, few nearby sensors may pick up same observations which are redundant. These redundancies cause burden to the entire network which in turn lead to reduced network lifetime. The energy resources and communication range of a sensor node, are limited. Thus expensive data transmission costs in general can be reduced by using a suitable data compression technique.

Data aggregation methods minimize the redundant data transmission in order to reduce the bandwidth usage of the network. There exists a good amount of correlation between the sensor nodes. CS is an emerging field in WSNs in which simultaneous sensing and compression offers a promising result particularly in large scale WSN. It greatly reduces sampling and computational costs.

CS works on the theory that the signal of interest can be preserved by using dimensionality reduction technique. CS has already proved to be an effective solution in the field of image compression as well as signal processing. Because of its attractive features, CS is also becoming popular in the field of wireless communication and sensor networks. CS plays a significant role in wireless channel estimation, signal detection data gathering and so on. An efficient WSN must concentrate on enhancing the lifetime of the network by reducing the data transmission [Xu et al. \(2013\)](#).

3.2 Compressive Sensing (CS)

Methodology followed in data processing involves, sensing or measuring the data in its full length and then compress the data prior to storage/transmission. The conventional data acquisition process is termed as full sensing plus compression. CS is a method in which the length of the acquired data is minimized during sensing itself, so that extra compression step can be eliminated. In a stage wise method when we get a relaxation in particular step the burden is still present but is shifted. Similarly in CS since the acquisition step is made easy, but the burden is moved to BS. Nevertheless the recovery procedures become complex. In order to accomplish CS, the signal needs to be sparse. A signal is termed as sparse if most of the entries are zeroes. Thus if it is possible to transform the signal into sparse then it is easy to employ CS for such signals.

3.2.1 Signal acquisition and method of reconstruction in CS based system

$$\begin{array}{c}
 \left[\begin{array}{c} y \\ \vdots \end{array} \right] \\
 \text{data} \\
 M \times 1 \\
 \\
 K < M \ll N
 \end{array}
 =
 \begin{array}{c}
 \left[\begin{array}{c} \Phi \\ \vdots \end{array} \right] \\
 M \times N \text{ Projection Matrix}
 \end{array}
 \begin{array}{c}
 \left[\begin{array}{c} x \\ \vdots \end{array} \right] \\
 \text{Sparse signal} \\
 (K \text{ is non zero entity}) \\
 N \times 1
 \end{array}$$

Figure 3.1: Signal/Image acquisition in CS.

Let $x = (x_1, x_2, x_3, \dots, x_N)^T$ is the signal of dimension N , and x is k sparse when it consists of only k nonzero values in the acquired signal either in the domain in which it is acquired or in the transform domain. Consider the Fig.3.2, which depicts the humidity values captured by the sensor. The signal in time domain can not be compressed, as it is not sparse. In order to apply CS, the signal has to be transformed into a sparse signal. We perform the DCT of the signal, then we can have larger and smaller component of the signal which can be considered as a sparse signal as in Fig.3.2.

We can define x is k sparse in ψ if there are an orthonormal basis denoted by $(\psi_1, \psi_2, \psi_3, \dots, \psi_N)$ as in equation 3.1.

$$x = \sum_{i=1}^n k_i \Psi_i \quad \text{or} \quad x = \Psi k \quad (3.1)$$

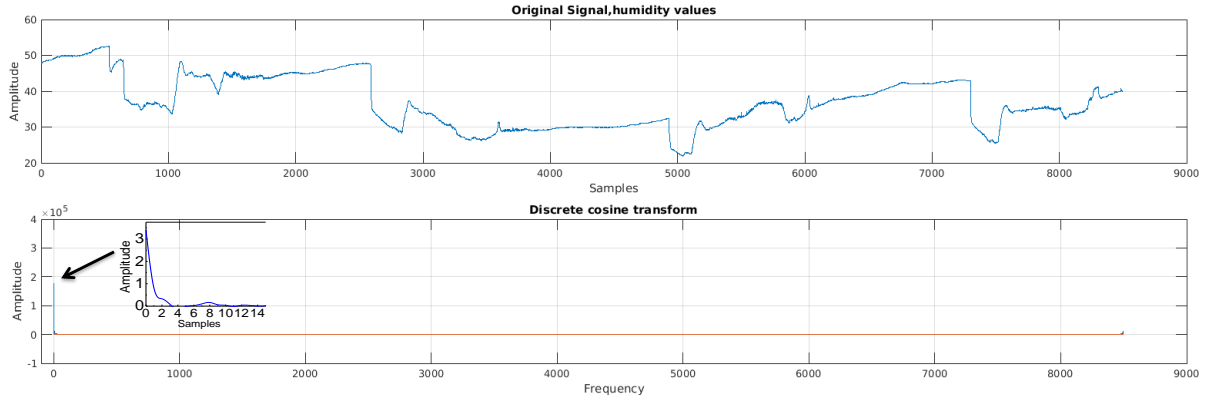


Figure 3.2: Humidity signal in acquired and transformed domain.

CS theory affirms that a k sparse signal x of N dimension can be confined into y using M ($M \ll N$) linear projections with the help of a $M \times N$ matrix as in equation 3.2

$$y = \Phi x = \Phi \Psi k = \Theta k \quad (3.2)$$

Where Φ is the projection or measurement matrix.

In order to preserve the information contained in x , the $M \times N$ matrix Φ , must maintain the inherent properties of a k sparse signal during the transformation of $x \in \mathbb{R}^N$ to $y \in \mathbb{R}^M$.

3.2.2 Reconstruction model

A nonlinear algorithm is used in CS at receiver end to reconstruct original signal. This nonlinear reconstruction algorithm requires knowledge of a representation basis (original or transform) in which signal is sparse. There it needs a stable *measurement matrix* Φ and a *Reconstruction Algorithm* to recover x from only m measurements y . Convex optimization and Greedy pursuits are two primary reconstruction algorithms in CS. Reconstruction algorithms in CS, try to solve $y = \Phi x$, and exploit the fact that solution is sparse, usually by minimizing l_0 , l_1 or l_2 norm over solution space. Restricted Isometric Property (RIP) Donoho (2006) guarantees that we can fully describe the signal in compressed form by the M measurements where $M < N$, but does not reveal anything about retrieving the original signal x . For Under-determined system with $M < N$, how many directions ' x ' can move in to preserve $Ax = 0$?. The space of such direction is known as the *null space*. If we move a point ' u ' in any such direction, we leave the value of $Au = A(u + x) = Au + Ax = Au$ unchanged.

The signal reconstruction algorithm must take the m measurements in the vector y , the ran-

dom measurement matrix Φ and the basis Ψ and reconstruct the length- N signal S , or equivalently its sparse coefficient vector S . Therefore we can say that this space can be spanned by $N - M$ linearly independent directions. For K -sparse signals, since $M < N$ there are infinitely many x' that satisfy $Ax' = y$. This is because if $Ax = y$ then $A(x+u)=y$ for any vector u in the null-space of A . Therefore, main goal is to find the signal's sparse coefficient vector in the translated null space (dimension $(N - M)$) [Baraniuk \(2007\)](#)

l_0 norm reconstruction

l_0 norm counts the number of non-zero entries in x .

$$\hat{x} = \operatorname{argmin} \|x'\|_0 \text{ such that } Ax' = y.$$

l_0 minimization is computationally intractable (in fact, it is an NP-hard problem in general), this is because l_0 minimization is not a convex optimization problem.

l_2 norm reconstruction

l_2 norm measures the signal energy. Algorithm tries to find the vector in the translated null-space with the smallest l_2 norm(energy) by solving

$$\hat{x} = \operatorname{argmin} \|x'\|_2 \text{ such that } Ax' = y.$$

l_2 optimization has the closed-form solution $\hat{x} = A^T(AA^T)^{-1}y$. But l_2 minimization will never find a sparse solution instead it returns a non-sparse \hat{x} with many non-zero elements.

Thus when solving an under-determined problem $Ax = y$, l_2 minimization is easy to compute, but often wrong. When x is sparse, l_0 minimization is often correct, but very difficult to compute.

l_1 norm reconstruction

Optimization based on l_1 norm can exactly recover sparse signals and closely approximate compressible signals.

$$\hat{x} = \operatorname{argmin} \|x'\|_1 \text{ such that } Ax' = y.$$

l_1 minimization is a convex optimization problem and can be solved fairly quickly by linear programming methods.

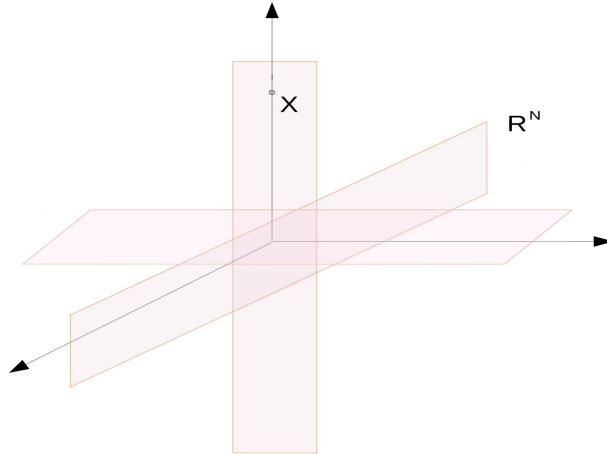


Figure 3.3: Subspaces containing sparse vector in R_3

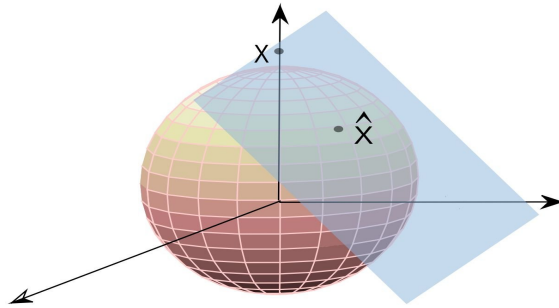


Figure 3.4: l_2 minimization

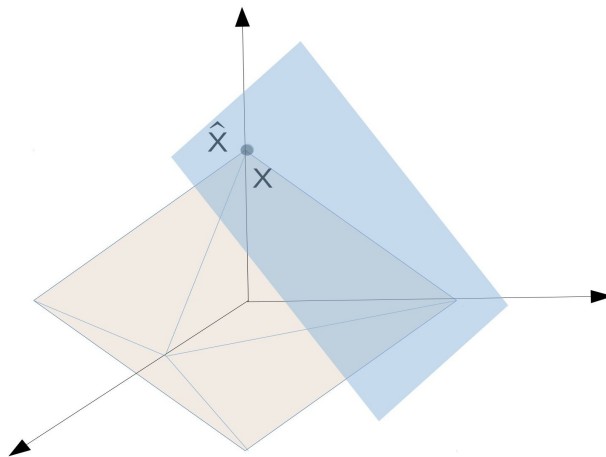


Figure 3.5: l_1 minimization

The set of all K -sparse vectors x in R^N is a highly non-linear space consisting of all K -dimensional hyperplanes that are aligned with the co-ordinate axes as in Figure 3.3.

The l_2 minimizer \hat{x} is the point on the translated null-space, which is closest to the origin. We can solve this by blowing up a hyper-sphere (l_2 ball) until it contacts the plane (light blue plane) as shown in Figure 3.4. Due to the random orientation of the plane, the closest point

\hat{x} will lie away from the co-ordinate axes and that is why the solution is not sparse as well as will not be close to the actual answer x . Surprisingly when we consider l_1 ball has points aligned with the co-ordinate axes. When we blow the l_1 ball it will first contact the translated null space(light blue plane) as shown in Figure 3.5, at a point near the co-ordinate axes, which is clearly where the sparse vector x is located.

The recovery of data samples is guaranteed based on the measurement matrix. The matrix must possess certain property. In that case even from an under determined system, it is possible to obtain good estimation of the data at the receiving end. The same is verified by the RIP of the matrix which is used for the dimensionality reduction. Restricted isometric constant measurement matrix Φ , for a k sparse vector smallest $\Delta_k > 0$ such that equation 3.3 holds good

$$(1 - \Delta_k)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \Delta_k)\|x\|_2^2 \quad (3.3)$$

Measurement matrix Φ must satisfy the RIP. Successful recovery of the data depends on RIP. If $\Delta_k = 0$ for all $k \leq N$, it implies that Φ is orthonormal. But when the signal measurement M is greater than original acquired signal dimension N , the value of Δ_k will be no longer zero. If the value is nearly zero means that the measurement matrix is nearly orthonormal, which increases the estimation accuracy of x .

The transformation of the signal from the signal space to measurement space is done with the help of projection matrix. As we know $M < N$, thus measurement space is usually lesser than signal space. Thus RIP property guarantees, that the Euclidean space is not altered by this transformation.

3.2.3 Compressive sensing in Wireless Sensor Networks

In a networked data gathering method, the collected sensor data are processed through certain algorithms before transmitting to the BS. This is done in order to filter out the redundant data and transmit only the necessary amount of data. CS uses dimensionality reduction to transmit the data. It reduces the complexity at the acquisition end by the complex recovery procedures Liu *et al.* (2015). As compared with conventional compression, CS provides a means to acquire the compressed samples directly rather than processing at the intermediate stages. It also offers several methods to estimate the original signal from the compressed samples. Now

suppose that the joint compression is not aimed at and each sensor compresses the signal on its own. Firstly, the compression achieved by this approach is not optimal as inter-sensor correlation is not exploited at all. The total volume of the independently compressed data is much larger than that of jointly compressed data. This may produce a large traffic volume in the WSN, and a large amount of transmission power is wasted from the sensor nodes that transmit essentially the same information to the BS. Thus, this is an inefficient strategy as well.

In comparison with traditional compression schemes, CS based method directly acquires the compressed samples. The requirement to achieve the same is that, the signal of interest must be sparse. It can exhibit sparsity in the acquired or transfer domain. By exploring intra-sensor correlation CS provides a direct means for signal compression.

In order to compress the high-dimensional signal x into a low-dimensional signal y , algorithm makes use of a $M \times N$ projection matrix $A_j, j \in 1, 2, \dots, J$, where j is the sensor index, as depicted in Figure 3.6. The dimensionality reduction is achieved by using a projection matrix which is usually termed as a measurement matrix.

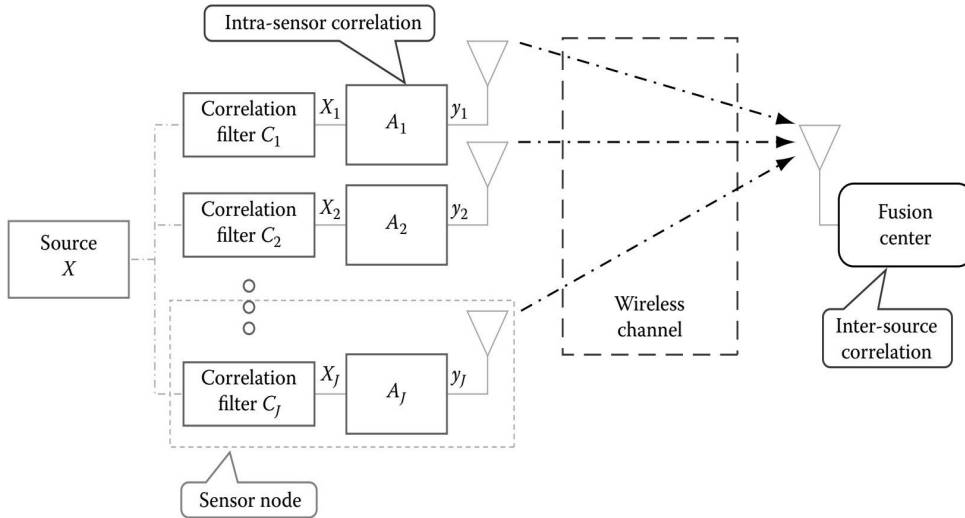


Figure 3.6: Network scheme with Compressive sensing [Hu and Hao \(2012\)](#)

In an aggregation scheme based on CS, each signal picked up by the sensors are compressed using a projection matrix and then transmitted to the BS. Inter-correlation can be achieved at the BS using joint recovery. In DCS, both intra and inter correlations are explored which will be discussed in the next chapter. Thus there is no need to collaborate at a single point before transmission to explore inter correlation. Relaxation of having intermediate stages offers us to

reduce the time delay as well. Thus each sensor generates lesser amount of traffic with reduced time delay which inturn results in the betterment of network lifetime.

3.3 Routing Protocols

Reliability can be achieved through multi-path routing. The data can be routed through more than one path which improves the accuracy of data reception at the BS.

3.3.1 Impact of using adjustable transmit power levels

The transmission power levels directly influence, the transmission range, communication energy drain, and connectivity in a sensor network. The transmission range in multi hop communication can also be considered as shorter hops (smaller transmission range between the hops) and longer hops (larger transmission range between the hops). There have been attempts in the past to decide which is better in multi hop communication from different perspectives but no consensus has been reached [Haenggi and Puccinelli \(2005\)](#). As a result we have used a combination of shorter and longer hops of different transmission ranges in multi-hop communication and verified its effects on network lifetime (considering only the communication energy drain). The importance of nodes near the BS with respect to network lifetime based on variable transmission power levels is also presented in these examples.

Most of the time fixed smaller transmission ranges cause a bottlenecking problem with less nodes near the BS to access the BS for multi hop communication (unless more nodes are deployed near the BS that can reach the BS with that particular transmission range). Most often the nodes near the BS die faster than other nodes because they must carry all the data traffic from the rest of the nodes to reach the BS, thus leading to early network partition. From the perspective of network lifetime, the biggest advantage of nodes capable of adjusting to different transmission power levels is that a large number of nodes can reach the BS and can help to achieve a better overall network lifetime.

The importance of nodes near the BS with respect to network lifetime based on variable transmission power levels is presented in these examples.

Example 1: Assume there are 100 nodes randomly distributed in an area of 100×100 me-

Table 3.1: Chipcon cc2420 transceiver supported power levels and its power consumptions

| Serial Number | Available Transmit output power in dbm | Power consumed in mW | Transmission ranges in mtr |
|---------------|----------------------------------------|----------------------|----------------------------|
| TX-Power 1 | -25 | 15.3 | 4.216 |
| TX-Power 2 | -15 | 17.82 | 11.00 |
| TX-Power 3 | -10 | 20.16 | 17.78 |
| TX-Power 4 | -7 | 22.5 | 23.71 |
| TX-Power 5 | -5 | 25.02 | 28.72 |
| TX-Power 6 | -3 | 27.36 | 34.80 |
| TX-Power 7 | -1 | 29.7 | 42.16 |
| TX-Power 8 | 0 | 31.32 | 50.00 |

ters. The requirement is to have all the nodes with a node degree (number of nodes that can be communicated with each node) of at least 6. Let us assume that this can be achieved by having a transmission power level of -7 dbm corresponding to a range of 23.71m and a power drain of 22.5mW from table 3.1. Using a fixed transmission range or precisely a fixed power level, each node spends 22.5mW of power for every communication. With the selection of a low power level such as -10 dbm having lower power loss can result in a large percentage of nodes to have the connectivity or node degree of at least 6 and the remaining smaller percentage of nodes have a node degree less than 6 but greater than zero. Then a large section of nodes consume less power and a smaller section of nodes consume more power to have a node degree of at least 6. This can minimize the overall communication power drain to a large extent and help in achieving better network lifetime.

Example 2: Consider another scenario in which a fixed transmission power of -7 dbm is used corresponding to 22.5mW of power drain by all the nodes. Suppose that the nodes are relatively far from the BS need to communicate to the BS, then it is achieved through multi hop communication in sensor networks. If it takes 6 hops to reach the BS, the total end to end power drain would be $6 \times 22.5\text{mW} = 135\text{mW}$. If a higher transmission power level is used, for example -3 dbm causing a power drain of 27.36mW and if its H_{Ind} gets reduced to 4 hops due to an enhancement in the transmission range, then the total end to end power drain comes to $4 \times 27.36\text{mW} = 109.44\text{mW}$. It is clear from the calculations that a considerable amount of power saving can be achieved using a higher transmission power to collect data at the BS through multi hop communication.

Example 3: Most of the time fixed smaller transmission ranges cause a bottleneck problem with less nodes near the BS to access the BS for multi hop communication (unless more nodes are deployed near the BS that can reach the BS with that particular transmission range). Most often the nodes near the BS die faster than other nodes because they must carry all the data traffic from the rest of the nodes to reach the BS, thus leading to early network partition. From the perspective of network lifetime the biggest advantage of nodes capable of adjusting to different transmission power levels is that a large number of nodes can reach the BS and can help to achieve a better overall network lifetime.

3.3.2 Cluster based Node-disjoint Multi path routing

In multi hop communication, the importance of nodes near the BS often determines the factors affecting it. The main factors would be the deployment strategies of sensor nodes and the BS or the BSs. [Heinzelman et al. \(2002\)](#). There is a significant amount of literature available on the deployment strategies in sensor networks and forms another area of research direction itself [Xing et al. \(2015\)](#). Routing in sensor networks can be broadly classified into two main categories: BS selecting and initiating the routing and the other is distributed [Nishant et al. \(2012\)](#). We opted for the BS selecting the routes and initiating the multi-path. This was due to possibility of many applications in the field of sensor networks such as secure monitoring of protected areas or disputed areas or even surveillance applications where the deployment of sensor nodes is predetermined. There could also be many applications which require nodes to be location aware, be equipped with GPS units and convey their location information to the BS. BS selects and conveys the information regarding routing process to all nodes.

The BS selects proper intra transmission and inter transmission power levels associated with each multi path. The transmit power which is able to provide the minimum node degree is decided as intra transmit power level, and above intra transmit are the inter transmit power levels. Selection of CHs is based on the highest degree of a node among the nodes available in the intra cluster area. During the process of CH selection there may be a few nodes which do not plunge under the territory of any CH, Those nodes need to have higher power level in order to get connected with the nearest CH which is termed as H_{node} . Selected CH has to bear the data traffic of the network, thus there is a need to account the number of hops related with

individual selected CH as defined in [Nishant et al. \(2012\)](#) and given equation 3.4.

$$H_{Ind} = \frac{\sum_{I=1}^{N_{CH}} H_I}{N_{CH}} \quad (3.4)$$

Where H_{Ind} is the counter for Average number of Hops of the network.
 N_{CH} is the total number of nodes selected as CHs in that particular combination of power levels
 H_I accounts for number of hops required by individual CH to arrive at the BS.
Average end to end communication energy drain is as given equation 3.5

$$E_{drain} = H_{Ind} \times ICH_{pd} \times DT_{dur} \quad (3.5)$$

Where E(drain) in milli joules accounts for communication energy drain.

ICH_{pd} is the inter CH communication drain in milli watts.

DT_{dur} is the data packet transmission duration in seconds.

H_{Ind} is the counter for Average number of Hops of the network.

The choice of proper inter transmit power levels with elected intra transmit power levels for different path in node disjoint multi-path routing is achieved by minimizing the energy drain of the network.

$$Minimize(E_{drain}) \quad (3.6)$$

Among the available inter transmission power levels, the power level that satisfies the above objective function is selected as the inter transmission power level and its corresponding path will be one of the node-disjoint path in the multipath routing.

3.4 Sparse vector reconstruction approaches

The solution for sparse recovery problem and its related applications can be solved using CS. From the literature there are two approaches towards sparse vector recovery. The first is the Convex relaxation approach which can be implemented using linear programming [Candes et al. \(2006\)](#), where as other one is based on greedy methods in which the problem is solved with its current form by using an approximation method . Both the approaches have their advantage and disadvantages.

3.4.1 Convex Relaxation

The reconstruction algorithm tries to estimate the signal x at the BS, with the help of measurement matrix A and a measurement vector y . The usual solution, is to solve the problem using Least squares problem. Since the measurement matrix is a column-rank deficient, thus there will be infinitely many solution. As we know the signal is sparse, it can be solved easily using l_0 norm. But this is computationally complex, thus an approximation is made for CS recovery problems. Here comes Convex relaxation procedure, in which problem formulation is relaxed so that the problem can be solved optimally. Thus non-convex l_0 norm can be replaced by convex l_1 norm, which is termed as Basis Pursuit (BP).

3.4.2 Greedy iterative pursuits

This type of reconstruction solves the recovery process by evaluating the solution, in steps by an iterative strategy. In each iterative step the current evaluation for the solution vector x is refined. Halting criteria of the algorithm varies for different matching pursuits, which may involve the number of iterations or if the residual has a smaller magnitude etc. Most commonly used greedy algorithms include Orthogonal Matching Pursuit (OMP) [Tropp and Gilbert \(2007\)](#), one among the earliest methods of sparse signal recovery method. For several applications OMP may not offer ample performance, which led to the development of improved pursuit methods which work better and yield substantially optimal results. Several greedy based algorithms have been listed in the literature to address a few variations of OMP i.e regularized OMP [Needell and Vershynin \(2009\)](#), Residual Minimization Pursuit [Song et al. \(2013\)](#) and Stagewise OMP [Donoho et al. \(2012\)](#), CoSaMP [Needell and Tropp \(2009\)](#), Subspace Pursuit [Dai and Milenkovic \(2009\)](#). The attractive feature of this greedy pursuits is the low implementation costs and the speed of recovery, but the implementation cost turns to be costly when the sparsity of the signal is low.

Orthogonal Matching Pursuit (OMP)

OMP [Tropp and Gilbert \(2007\)](#) is an improved method over Matching Pursuit (MP), In MP we need to approximate, the best matching projection of the data from the dictionary or to be precise from the sensing matrix in order to approximate the sparse solution. The process of orthogonalization is an extra addition in OMP, in each iteration of the algorithm it ensures the

orthogonal direction of projection. A detailed explanation is as follows.

Residual Minimization Pursuit (RMP)

In RMP, the algorithm considers on minimization of the residual rather than considering the correlation maximization as it is done in OMP. OMP algorithm first finds the columns of the measurement matrix which is highly correlated with residual signal to estimate one active element at each iteration. In paper by [Song et al. \(2013\)](#) target localization problem is considered analogous to sparse recovery problem. In order to achieve the same they used RMP algorithm, which provides a suitable platform for target localization alternate to OMP based method. When the sensing matrix has orthonormal rows, RMP converges to OMP.

| Orthogonal Matching Pursuit | |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input: | compressed vector y [M], Measurement/sensing matrix Φ sparse input signal x [N] $M < N$ |
| Init: | residual $r_0 = y$, $k=1$, $\hat{x}=0$, $\hat{\Phi}_0 = \emptyset$ index set $= \Delta_s = \emptyset$. $\emptyset =$ empty set, $\delta =$ small constant. Repeat until stopping criteria holds, while $\ r\ < \delta$ |
| (1) | Identify the column vector Φ_c which is highly correlated with the residual. $\delta_k = \text{argmax}_{i=1,2,\dots,N} \langle r_{k-1}, \Phi_c \rangle $ |
| (2) | Update the index set by augmenting it with chosen column. along with the matrix $\Delta_k = \Delta_{k-1} \cup \delta_k$, $\Phi_k = [\Phi_{k-1}, \Phi_{\Delta_k}]$ |
| (3) | Estimation of the sparse signal by computing least square problem $x_k = \text{argmin} \ y - \Phi_k x\ _2$ |
| (4) | Update the residue $y_k = \Phi_k x_k$, $r_k = y - y_k$ |
| (5) | Advance the counter, if $k < s$ go to step 2 |
| | End |
| Output: | \hat{x} , $\hat{\Phi}$ |

Stagewise Orthogonal Matching Pursuit (St-OMP)

The above mentioned sparse recovery algorithms, OMP and RMP take 's' iterations (sparsity of the signal). StOMP [Donoho et al. \(2012\)](#) runs faster than OMP and RMP. In OMP based method, only one coefficient enters at each iteration, but in StOMP many coefficients can enter. Rather than selecting the largest component of the vector, selection is done based on a threshold and several coefficients are selected if it is above the specified threshold. Further, residual is calculated based on least square problem. The algorithm runs only for fixed number of stages. StOMP is faster than OMP, as it can take many coefficients rather than one as in OMP. StOMP converges to OMP, when the threshold is set such a way that only one term enters in each stage. The algorithm is as follows.

| St-OMP | |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input: | compressed vector y [M], Measurement/sensing matrix Φ , sparse input signal x [N] $M < N$ |
| Init: | residual $r_0 = y$, $s=1$ (stage counter), $\hat{x}=0$, $\hat{\Phi}_0 = \emptyset$, index set= $\Delta_s = \emptyset$. Repeat until stopping criteria holds |
| (1) | Compute the inner product. $C_s = \operatorname{argmax}_{i=1,2,\dots,N} \langle r_{s-1}, \Phi_{ci} \rangle $ |
| (2) | The algorithm proceeds to find the significant nonzero by performing hard threshold. Which results in a set J_s which has larger coefficients. t_s is threshold value and σ_s is noise level. $J_s = \{j: C_s(j) > t_s \sigma_s\}$ $\Delta_k = \Delta_{k-1} \cup J_s, \Phi_k = [\Phi_{k-1}, \Phi_{\Delta_k}]$ |
| (3) | Estimation of the sparse signal by computing least square problem $\hat{x}_k = \operatorname{argmin} \ y - \Phi_k x\ _2$ |
| (4) | Update the residue $y_k = \Phi_k x$, $r_k = y - y_k$ |
| | End |
| Output: | $\hat{x}, \hat{\Phi}$ |

Subspace Pursuit (SP)

The algorithm SP [Dai and Milenkovic \(2009\)](#) is similar to StOMP, in the sense that many coefficients can enter the algorithm, rather than one as in OMP. But the strategy of finding the k columns is different than StOMP. Initially it chooses k columns from the sensing matrix, then it keeps refining the subset of k columns at every iteration. In the current iteration of estimating x (sparse signal), if y does not exist in the estimated subspace, then the estimate list of k columns is updated by holding the reliable candidates and releasing the unreliable ones. The process continues by adding the same number of candidates. In contrast to the other revised versions of OMP, in SP there is a simple procedure of re-evaluation of the reliability of the candidates at every iteration step. The algorithm is as below.

| Subspace pursuit | |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input: | Measurement vector y , Measurement/sensing matrix Φ , sparse vector x |
| Init: | residual $r_0 = y$, $\hat{x}=0$, $\hat{\Theta}_0 = \emptyset$, index set= $\Delta_s = \emptyset$. Repeat until stopping criteria holds, $\ r^k\ _2 \leq \ r^{k-1}\ _2$ |
| (1) | Compute the inner product. $C_s = \operatorname{argmax}_{i=1,2,\dots,N} \langle r_{s-1} \Theta_{c_j} \rangle $ |
| (2) | Update the index set by augmenting it with chosen columns. along with the matrix $J_s = 's'$ indices which corresponds the largest value of C_s , $\Delta_k = \Delta_{k-1} \cup J_s$ $\Phi_k = [\Phi_{k-1}, \Phi_{\Delta_k}]$ |
| (3) | Estimation of the sparse signal $\hat{x}_k = \operatorname{argmin} \ y - \Phi_k x\ _2$ |
| (4) | Update the residue $y_k = \Phi_k x$, $r_k = y - y_k$ |
| | End |
| Output: | \hat{x} , $\hat{\Phi}$ |

3.5 Data aggregation using CS for improved network life-time

In this section, we will detail, how the life time of a WSN can be improved by using CS. The simulations are performed in MATLAB. Recovery based on Basis Pursuit are done using

l_1 magic [Candès et al. \(2006\)](#) as well as greedy method.

3.5.1 System Model

In large scale WSNs the amount of data generated is enormous. The data has to be processed efficiently before it reaches the BS by using an efficient routing algorithm as well as data aggregation methods. The nodes in WSNs are randomly deployed, the data emerging from these nodes are highly correlated either spatially or temporally. The data aggregation scheme should employ simple encoding since the sensor nodes are battery operated. The proposed method discusses about a data aggregation scheme using CS technique which makes use of correlation among the sensor nodes. Our primary focus is to increase the lifetime of the overall network. The underlying protocols used are Low-Energy Adaptive Clustering Hierarchy (LEACH) and Multi-Threshold Adaptive Range Clustering (M-TRAC). We have computed several network parameters for different network configuration. The reconstruction algorithm is sufficiently robust against noise. The reconstruction of the data is done using greedy method and l_1 norm regularization. The implementation of the algorithm is done using the real data- set from Intel Lab [Koushanfar et al. \(2006\)](#). Simulation results validate that the data aggregation scheme guarantees data accuracy and doubles the network lifetime.

In order to estimate the transmission energy cost, we have incorporated a standard transmission model [Heinzelman et al. \(2002\)](#). In this model, the energy per bit for transmission over a wireless link is a function of the distance between a transmitter and a receiver. Let $E_{TX}(N, d)$ and $E_{RX}(N)$ be the energy consumed for transmitting or receiving a ' N ' bit message over a distance ' d ', are given in 3.7 and 3.8 respectively.

$$E_{TX}(N, d) = E_{T-elec} \times N + \epsilon_{amp} \times N \times d^2 \quad (3.7)$$

$$E_{RX}(N) = E_{R-elec} \times N \quad (3.8)$$

E_{T-elec} , E_{R-elec} are the energy consumption for transmitting and receiving one bit message, and ϵ_{amp} is the transmission amplifier. Initial simulations were conducted by considering the parameters given in the [Table 3.2](#)

Table 3.2: Simulation Parameters

| Parameters | Typical Values |
|---------------------------|----------------|
| Network area | 100 m × 100 m |
| Position of the sink node | (50,100) |
| Initial energy of node | 0.5 J |
| E_{T-elec} | 50 nJ/bit |
| E_{R-elec} | 50 nJ/bit |
| ϵ_{amp} | 100pJ/bit |
| Size of the data packet | 128bytes |

We have used Intel Lab data [Koushanfar *et al.* \(2006\)](#) in order to validate the results. The underlying protocols used are LEACH [Akkaya and Younis \(2005\)](#) and M-TRAC [Shivaprakasha *et al.* \(2013\)](#) which uses variable transmission ranges. The sensed data is compressed using a random Gaussian matrix as the measurement matrix. Then the compressed data is transmitted to the BS. With the addition of data aggregation using compressive sensing scheme, network lifetime has been improved.

We have used different compression ratios to compress the data and the corresponding error in the recovered data has been obtained using the following equation 3.9.

$$\epsilon = \frac{\|x(n) - \hat{x}(n)\|_2}{\|x(n)\|_2} \quad (3.9)$$

Where $x(n)$ is raw data, $\hat{x}(n)$ is the recovered data.

$$\|x(n)\|_2 = \sqrt{\sum_{k=1}^{k=n} x(k)^2}$$

3.5.2 Results and Analysis

In this evaluation, the performance indices like network lifetime and reconstruction error are used for analysis of the network. The lifetime of the network using LEACH with and without CS technique is compared and the plot is as shown in Figure 3.7. The plot verifies that the lifetime of the network is significantly increased with the use of CS technique. The underlying protocol used for comparison is LEACH. To validate the compression algorithm, we have considered CS for LEACH, MTRAC by considering 100 nodes random deployment in 100 m × 100 m area. In every simulation round, the data from the nodes are transmitted to the CH, and the CH transmits the compressed data to the BS. Then the data is reconstructed

using l_1 norm regularization and greedy methods. From Figure 3.8 we can infer that using CS, network lifetime is improved.

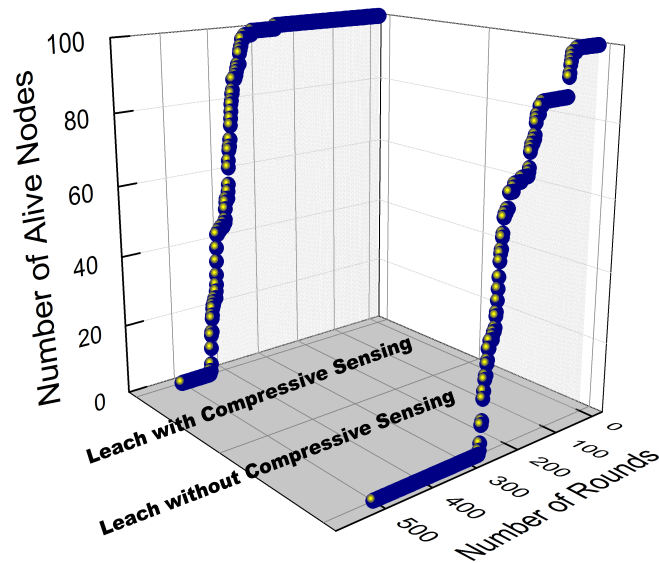


Figure 3.7: Network life improvement in LEACH using CS

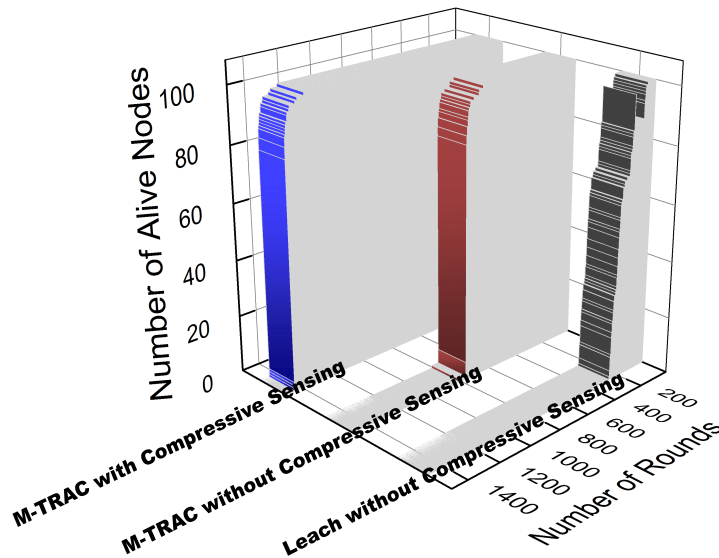


Figure 3.8: Network life improvement in M-TRAC using CS

Figure 3.9 gives the summary of all dead nodes. LEACH and M-TRAC performance improves with the incorporation of CS. The data from sensor nodes have been reduced significantly which results in keeping the nodes alive for longer duration.

Inclusion of CS as a data aggregation scheme, reduces the amount of redundant data along with network lifetime improvement. Reduction of the redundant data results in increased

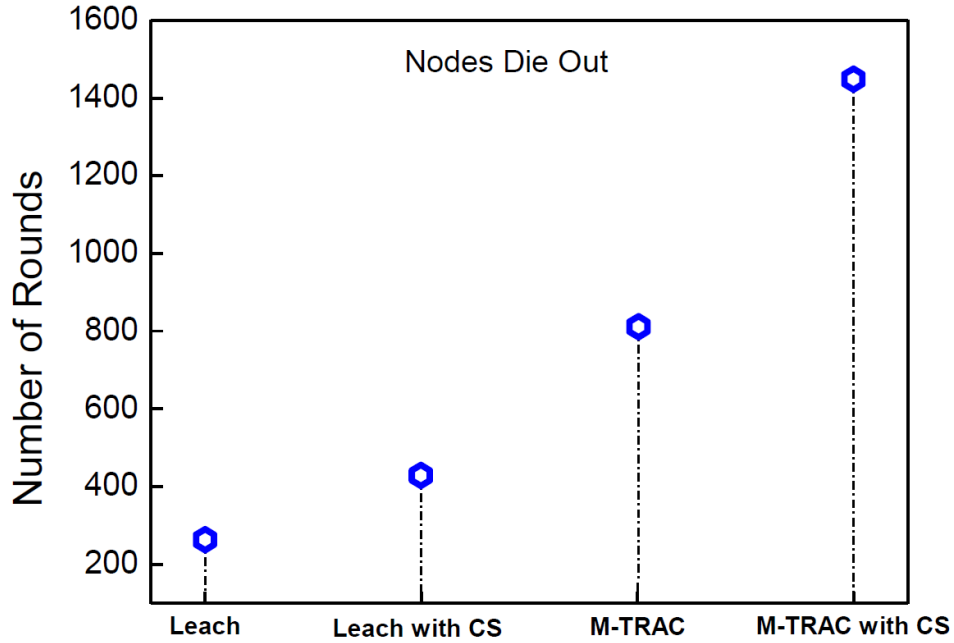


Figure 3.9: Summary of dead nodes in LEACH and M-TRAC

throughput and increased lifetime of the network. Data has been compressed using several compression ratios and reconstructed at the BS. Corresponding reconstruction error can be calculated using Equation(3.9) which is shown in Figure 3.11. Simulations have been carried out using temperature and humidity parameters of the Intel dataset. Intra-node correlations have been considered. The values for 3 days have been considered, collected around 3000 samples. The reconstruction has been done by considering different compression ratios. The real data are not usually sparse, but compressible. In this simulation, we have transformed the data into an appropriate sparsifying basis, then the data have been compressed.

Figure 3.12 shows the original and reconstructed values for different compression ratios for temperate and humidity values respectively. Depending on the correlation among the nodes and the amount of sparse data, there will be variation in the reconstruction error. Figure 3.11 shows the reconstruction errors for different compression ratio's for temperature and humidity values

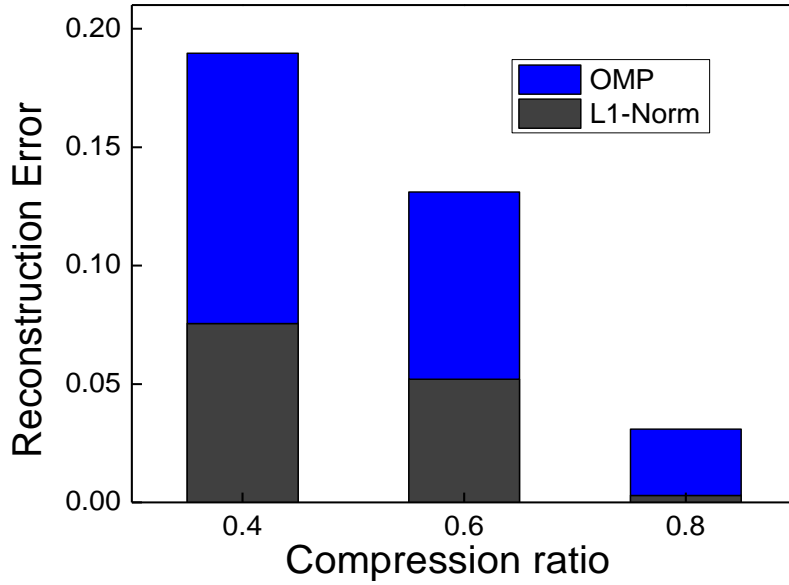


Figure 3.10: Comparison of Reconstruction error using l_1 Regularization and OMP

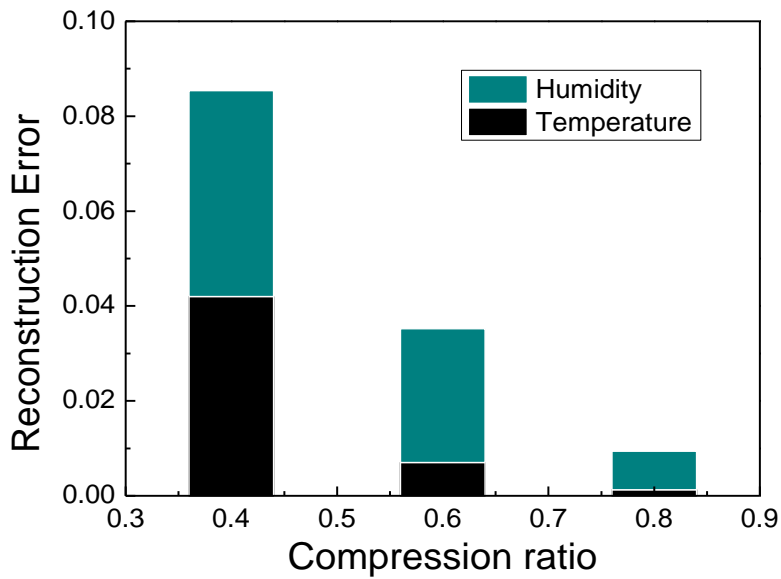


Figure 3.11: Reconstruction error for temperature and humidity values using l_1 -norm regularization

3.5.3 Experimental results of CS recovery based on Greedy Algorithms.

Greedy algorithms for CS reconstruction, give the similar results as compared with Convex relaxation methods by using different algebraic tools. Reconstruction of the data, is based on iteration procedure by projecting the data on significant column only. The categorization of greedy methods lies in the selection of the column (one or many), and the convergence rate, termination condition.

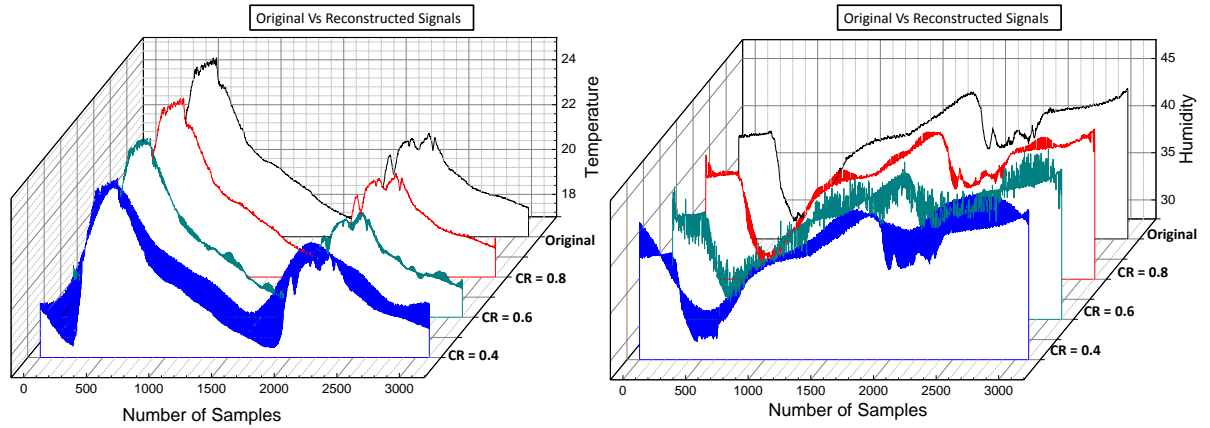


Figure 3.12: Reconstruction of temperature and humidity values using l_1 -norm regularization

To evaluate the performance, we have considered the data $N=512$ and the measurements varying in steps for different sparsity levels. The measurement matrix is independent and identically distributed (i.i.d) Gaussian matrix. The platform for simulation is MATLAB. Simulations are done based on greedy algorithms (OMP, RMP, StOMP, SP), and the plots depicts percentage of recovery and the error. Relative error is calculated as given in equation 3.10. The tolerance limit for perfect recovery is $1e-6$. The figures 3.13-3.20 shows recovery and error rate of the greedy methods (OMP, RMP, StOMP, SP).

$$relative\ error = \frac{\|x - \hat{x}\|_2}{\|x\|_2} \quad (3.10)$$

In Fig.3.21 we can see that the performance of the greedy methods for varying sparsity (k) and measurements (M). When the sparsity is less, almost all the methods convergence rate is similar except OMP. The time required to recover the data depends on the type of iteration, the algorithm performs. When the sparsity is less, all greedy methods convergence rate is almost same, but as K increases we can make a clear distinction of the convergence rate of various greedy methods.

Simulations are carried out by considering the dataset from Intel lab [Koushanfar et al. \(2006\)](#) which comprises of temperature and humidity data set. The signals are not sparse when they are acquired, we used DCT in order to sparsify the data. As we can see from Fig.3.23 and Fig.3.24 for temperature and humidity, SNR of all the methods are almost same but it differs with the rate of convergence. In order to meet the same SNR, the iterations taken by the greedy methods are different. WSN generates huge amount of data, it has to be processed and

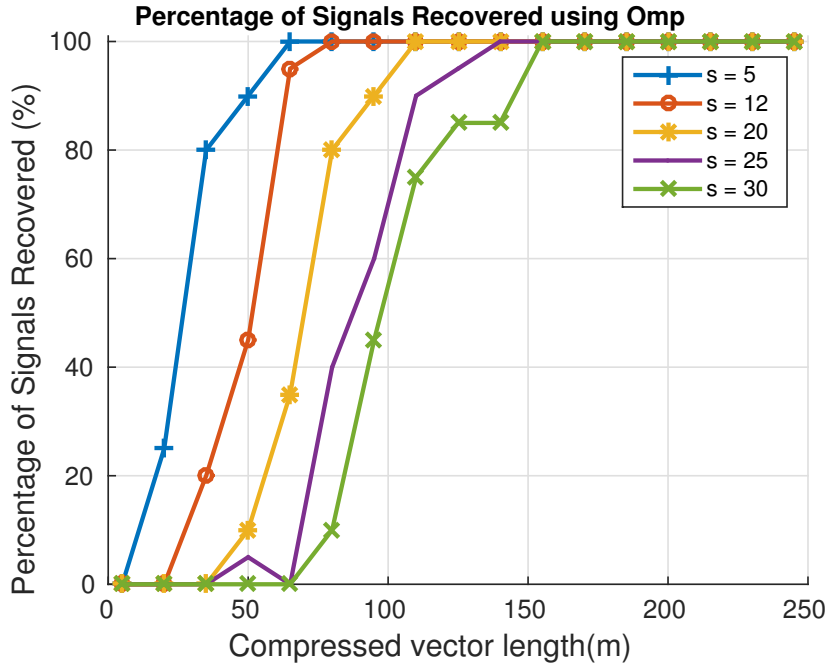


Figure 3.13: CS recovery using OMP.

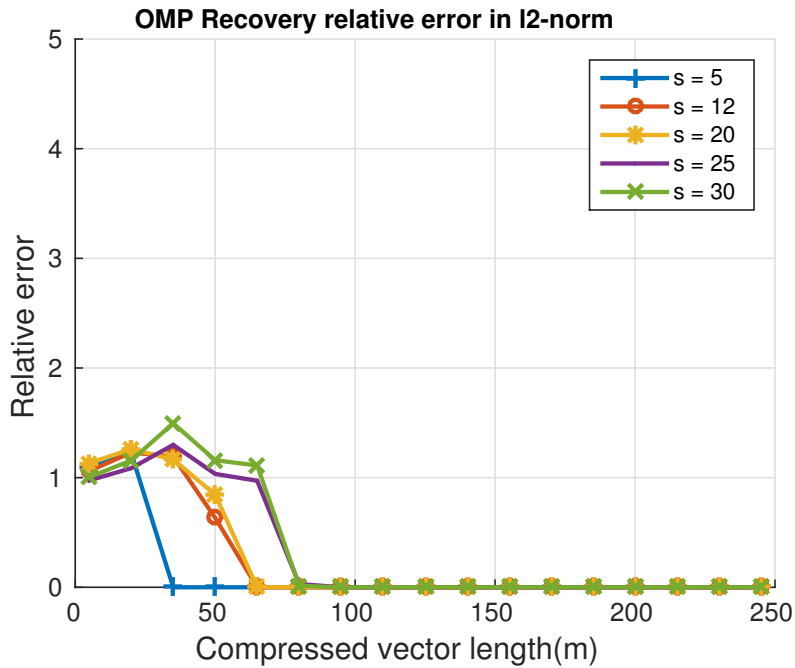


Figure 3.14: Relative error using OMP.

transmitted to the control unit in order to take necessary action. Further certain applications needs speedy processing and few may not. Thus based on the nature of the application suitable greedy algorithms can be selected.

If we consider CS as a data aggregation scheme along with the multi path routing strategy,

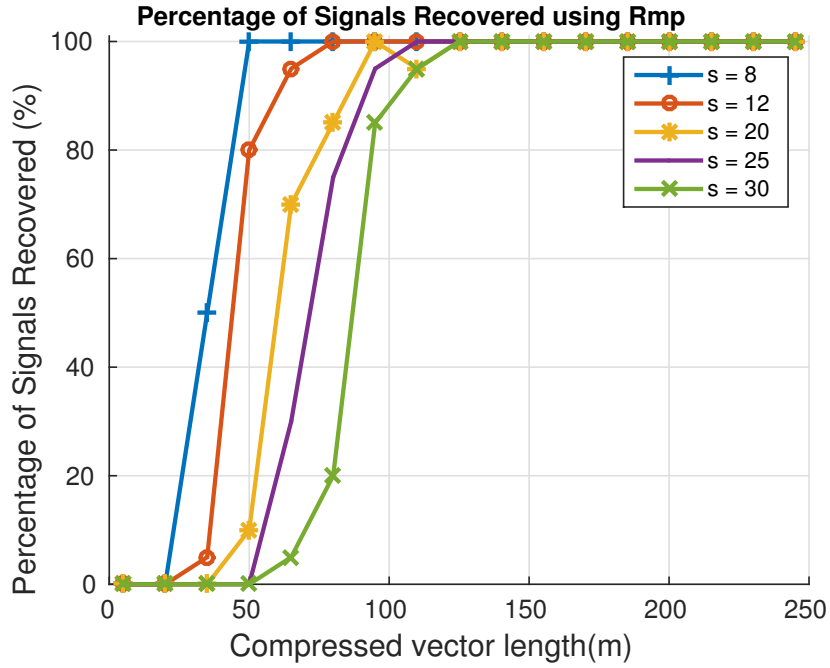


Figure 3.15: CS recovery using RMP.

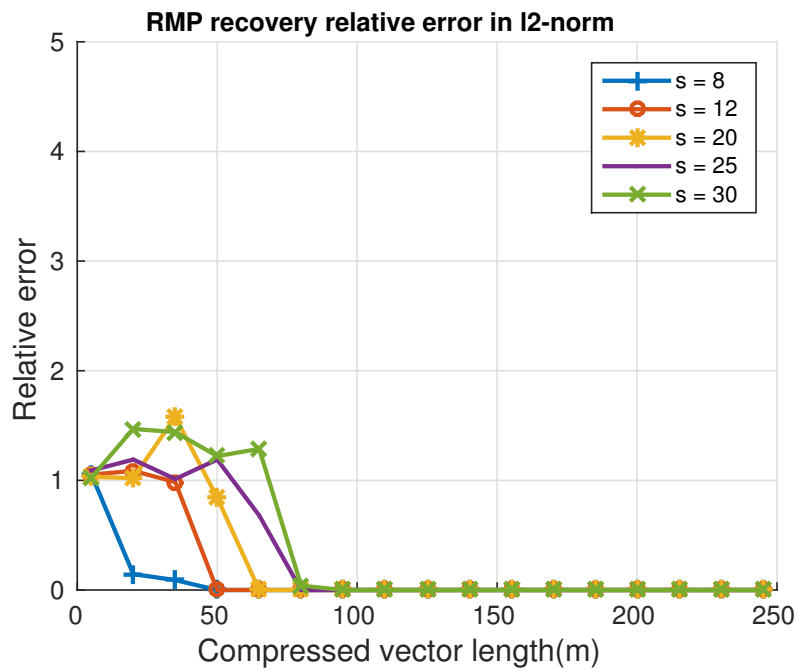


Figure 3.16: Relative error using RMP.

network lifetime can be enhanced. With CS, the number of packets to be transmitted can be significantly reduced. Since the nodes are deployed randomly, the data exhibits correlation either spatially or temporally. If the source nodes use lower intra transmission power levels to get connected to one hop CH, then we can significantly reduce the power consumption of the network.

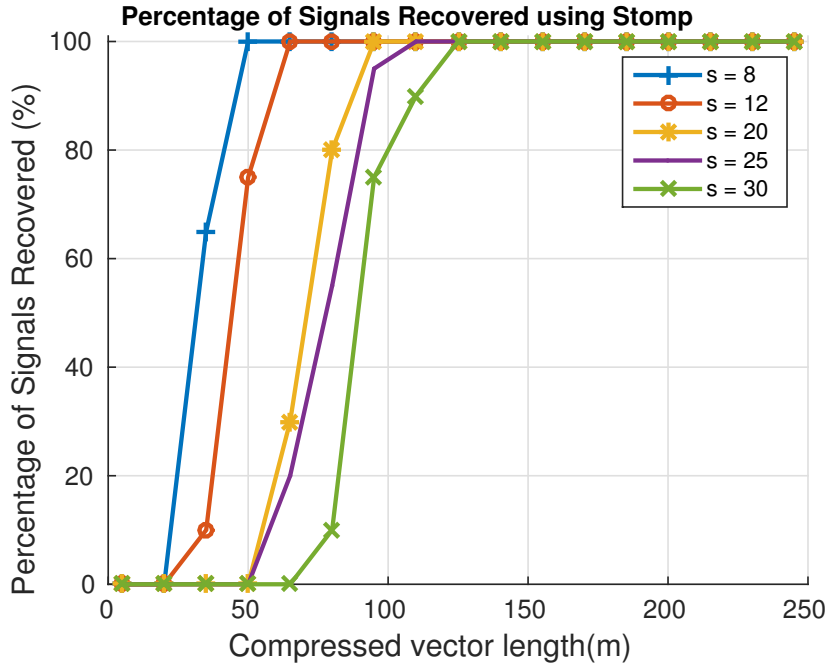


Figure 3.17: CS recovery using StOMP.

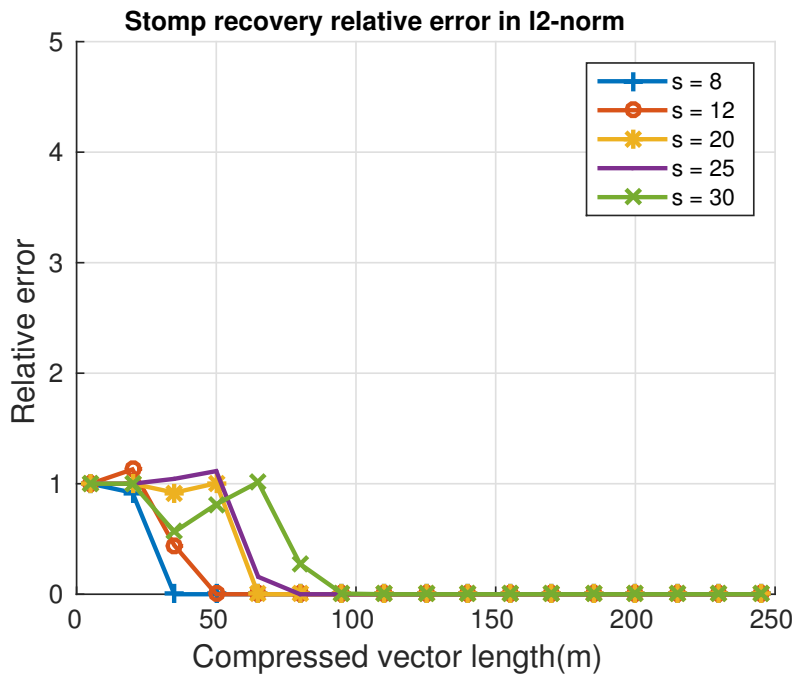


Figure 3.18: Relative error using StOMP.

3.6 Concluding Remarks

The proposed algorithm deals with data compression and reconstruction based on CS, which uses correlation property to compress the data. The traffic cost of the network has been reduced which reflects on the network lifetime. The performance metrics namely the lifetime

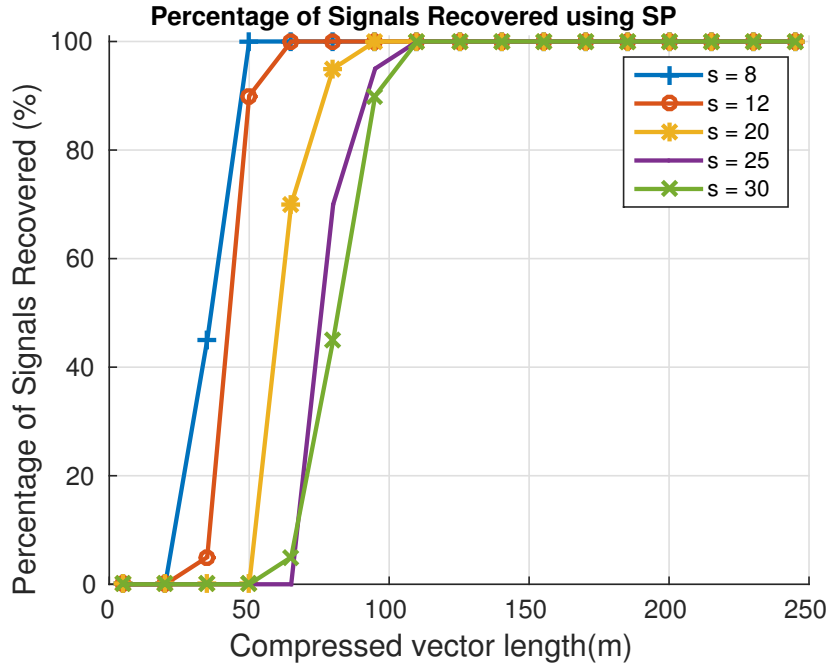


Figure 3.19: CS recovery using SP.

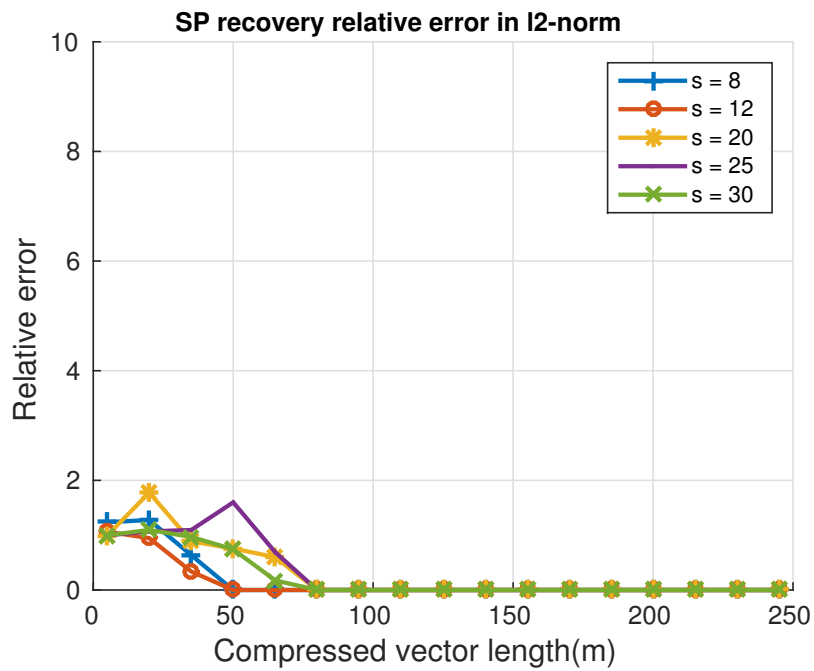


Figure 3.20: Relative error using SP.

of the network and the reconstruction error were analyzed and the results have been validated. It was observed that the network lifetime is doubled in both LEACH and MTRAC. M-TRAC algorithm proved to be a good algorithm over LEACH in terms of the lifetime of the network. Reconstruction is done using l_1 norm regularization and greedy methods. The outcome of our analysis is that applying CS may not bring the improvement in all cases, but could be applicable

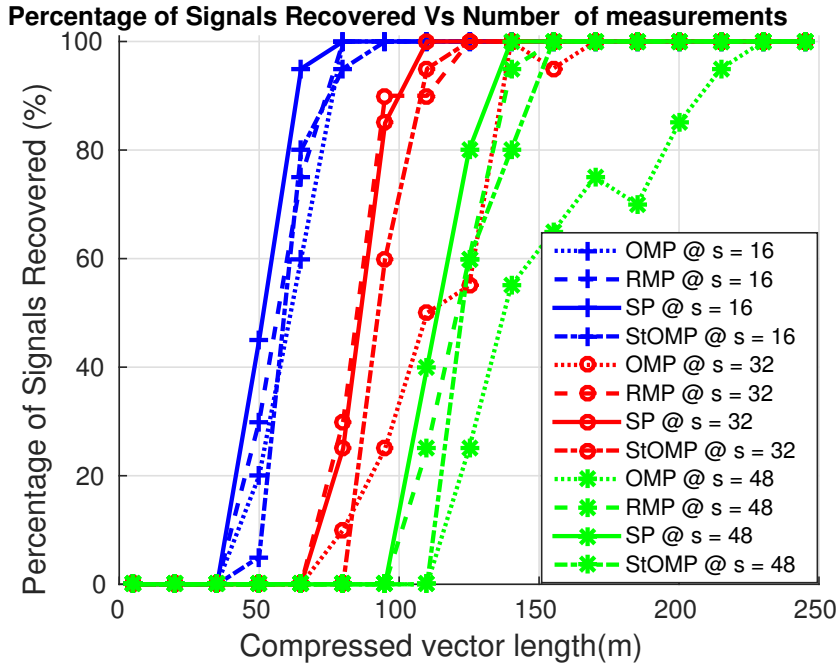


Figure 3.21: CS recovery using greedy methods.

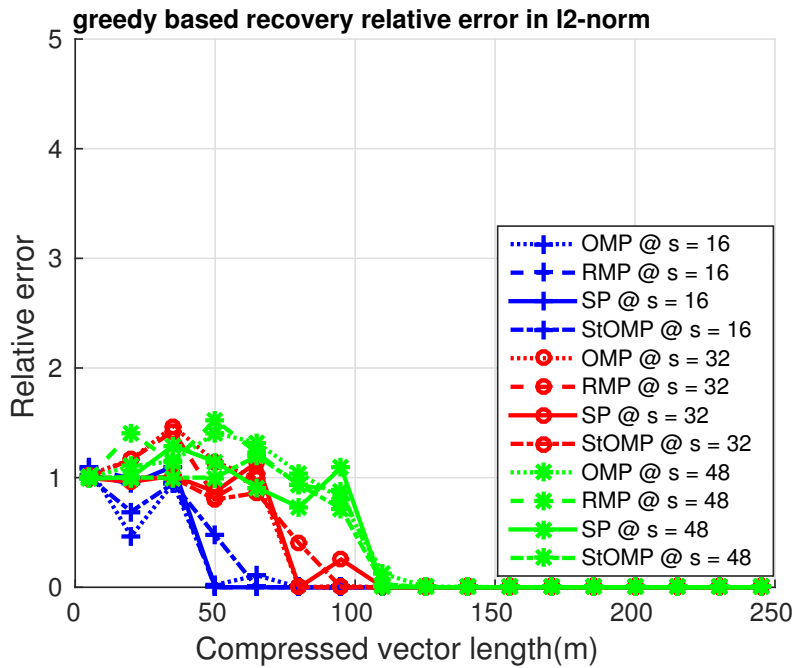


Figure 3.22: Relative error using greedy methods.

where correlation among nodes exists. The temperature values are highly correlated, thus the reconstruction error is comparably less. Analysis of correlations among nodes must be done before applying CS methods.

In order to improve network lifetime under multi-path strategy, we propose routing strat-

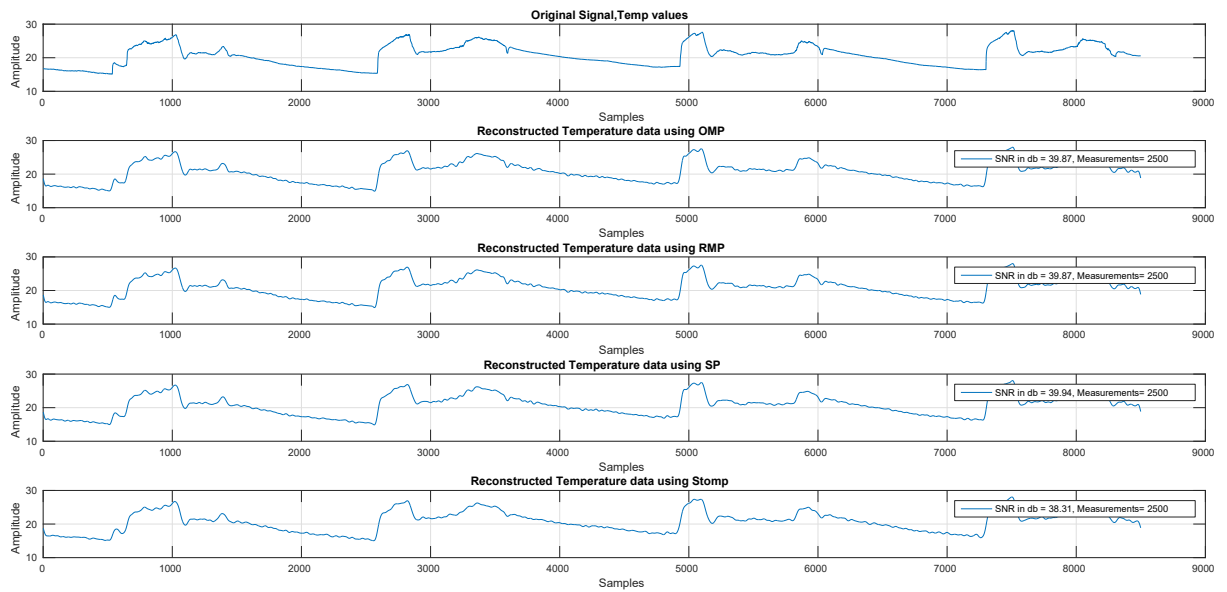


Figure 3.23: CS recovery using greedy methods for temperature data.

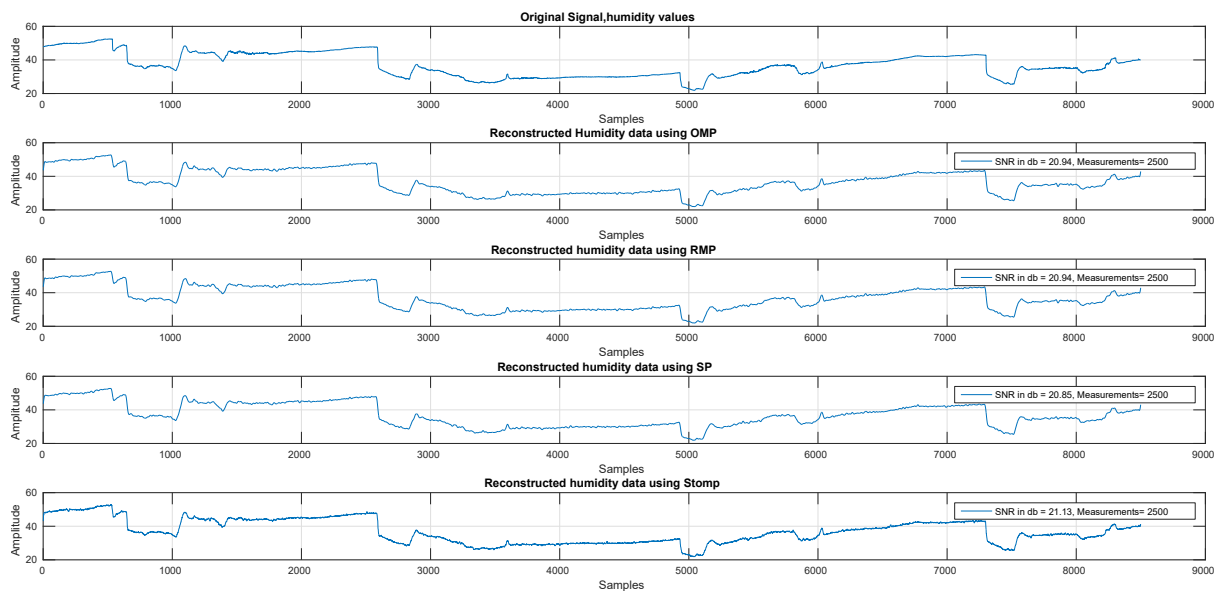


Figure 3.24: CS recovery using greedy methods for humidity data..

egy with tunable power levels along with data aggregation using CS. In order to achieve better network lifetime under node-disjoint multi path routing, sensor nodes adjust themselves to the best possible transmission power levels as per requirements. Since CS is used along with an efficient routing protocol, it is possible to reconstruct lost packets which gives an added advantage. CS reconstruction using greedy algorithm is found to be better than other reconstruction methods, in terms of computation time and complexity. Simulations prove that, recovery of synthetic and real data using greedy methods diminishes data stream into the network and in turn enhances the lifetime of the network.

CHAPTER 4

DISTRIBUTED COMPRESSIVE SENSING FOR WSNs

4.1 Introduction

WSNs consist of numerous nodes, randomly deployed in a geographical area that sense the environmental condition (temperature, humidity etc) and collectively work to manage and route the sensed signal to the BS. Data aggregation is a fundamental action in WSN's, where sensors are accountable for accumulating all the sensed values, and finally delivering them to the BS. In order to increase the life span of the network, we try to compress the data of individual sensor node. We use CS method to compress the data, further we emphasize with the method of Distributed Compressed Sensing (DCS) we can explore inter/ intra signal correlation with the concept of joint sparsity.

CS outperformed conventional compression methods and justifies a better trade-off between the quality of reconstruction and minimum power consumption. The high-dimensional signal which is sparse either in the acquired domain or transfer domain say $x \in \mathbb{R}^N$ can be compressed by projecting the signal into a low dimensional signal $y \in \mathbb{R}^M$. In order to carry this out the signal 'x' has to be multiplied with a projection matrix $A \in \mathbb{R}^{M \times N}$. At the receiver the estimation of \hat{x} has to be done from the under-determined system since fewer equations than unknowns. CS theory affirms that, the probability of having a unique solution depends on the sparsity of the signal vector x . The chances of having a unique solution increases if the signal is sufficiently sparse. In a WSN based on CS scheme, the sensed signals are compressed using linear projections and then transmitted. At the BS estimation of the signal is done through CS based recovery algorithms. In general, if we want to explore the inter- correlation among the signals, then we need to collect the samples in a single location and perform the compression. With the concept of Distributed Compressive Sensing (DCS), there is no need to gather the samples in one location, rather than that it needs a joint recovery at the decoding point. In case of DCS, each sensor senses the signals which are sparse in a particular basis and might be correlated from a sensor to another sensor. Thus, individual sensed signals are independently encoded

using a measurement matrix, which performs dimensional reduction of the sensed signal. Further these reduced samples are transmitted to the BS. At the receiving end, reconstruction takes place (exploring inter/ intra signal) by using one among the CS recovery algorithms. DCS theory relies on the concept of joint sparsity of the signal. The number of measurements in DCS is significantly less, as joint sparsity is usually smaller as compared to individual signal sparsity.

4.2 Intra and Inter correlation effects

The sensors are deployed randomly in the region of interest, thus the data values sensed from these sensors have either spatial or temporal correlation. In network data, aggregation and compression are the fundamental means to reduce communication cost and to extend network lifetime.

The inter node correlation $\Phi_{inter}(m)$ and the intra node correlation $\Phi_{intra}(m)$ which is calculated as given in equation 4.1 and 4.2 follows-

$$\Phi_{inter}(m^{(.)}) = \sum_{i=1}^{i=N} \sum_{j>i} \frac{(m_i^{(k)} - E[m_i])(m_j^{(k)} - E[m_j])}{\sigma_{m_i}\sigma_{m_j}} \quad (4.1)$$

$$\Phi_{intra}(m^{(.)}) = \sum_{i=1}^{i=N} \frac{(m_i^{(k)} - E[x_i])(m_i^{(k+t)} - E[x_i])}{\sigma_{x_i}^2} \quad (4.2)$$

4.3 Distributed Compressive Sensing.

The sensor nodes are randomly deployed in the area of interest, and a certain physical phenomenon are picked up by more than a single sensor node. CS can be employed to the sensed data from the sensor, if there is temporal correlation. But there might exist a spatial correlation among the sensed data, which further decreases the amount of data, thus improving the performance of WSN. CS based network can exploit either only temporal correlation or spatial and temporal depending on the reconstruction methods. It can exploit both correlation types if the decoding process is based on joint reconstruction methods. With out the nodes being communicating to each other, along with intra sensor correlation benefit with the help of joint reconstruction method we can exploit inter sensor correlation too. DCS is a strategy in which data is compressed depending upon, intra and inter correlations with out the sensors commu-

nicating to one another [Baron et al. \(2005\)](#); [Sarvotham et al. \(2005\)](#); [Wakin et al. \(2006\)](#) The underlying theory for DCS is ‘joint sparsity ‘ of the data signals. Let us go through the different joint sparsity models with the help of which we can recover the signals based on joint decoding at the receiver. Joint sparsity models show that even with out the sensors collaboration at the transmitter, it is possible to recover the data signals with reduced number of measurements. With the help of joint decoding we further can reduce the number of measurements than required if it is recovered separately.

4.3.1 Models based on joint sparsity

When we consider jointly sparse signals, there exists three separate models and each model fits for separate class of ensembles. Most of the time, signals in transformed domain is sparse thus signals can be encoded using CS which is termed as separate reconstruction which does not explore inter signal correlation. But there exists a joint sparse reconstruction which helps to recover the signal with less number of measurements.

We employ the following notation for the signal ensemble and the encoding/decoding model. The signal ensembles are denoted by x_l , with $l \in \{1, 2, 3, 4 \dots L\}$. There exists a sparse basis $\Psi \in R^N$ for each signal in which the signal ensemble x_l is represented sparsely. By considering suitable measurement matrix $\Phi \in R^{M_l \times N}$, the signals are compressed $y_l = \Phi x_l$ which posses the $M_l < N$ measurements of x_l

The Joint Sparsity Signal Models (JSM) are introduced in [Baron et al. \(2005\)](#); [Sarvotham et al. \(2005\)](#); [Wakin et al. \(2006\)](#). The joint sparsity model can be depicted as given in equation 4.3

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_l \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_l \end{bmatrix} \quad \Phi = \begin{bmatrix} \Phi_1, 0 \dots, 0 \\ 0, \Phi_2, \dots, 0 \\ \vdots \\ 0, 0, \dots, \Phi_L \end{bmatrix} \quad (4.3)$$

Signal modeling using DCS, consists of representing the signal using two components : common sparse components (present in all the signals under consideration) and sparse innova-

tion component (pertaining to individual signal).

$$x_l = Z_C + Z_l, l \in \{1, 2, \dots, L\} \quad (4.4)$$

with $Z_C = \Psi\theta_C, \|\theta_C\|_0 = K$ and $Z_l = \Psi\theta_l, \|\theta_l\|_0 = K_l$.

The signal which is split into two components, ' Z_C ' components is common to all the signal and Z_l is a special component of each signal.

Depending on the type of correlation, which exists among the sensors following joint sparsity models are configured namely Joint Sparsity Model (JSM) -1, JSM-2, JSM-3. In JSM-1 all the signals have similar set of sparsity (non zero components) which we can call it as common sparsity and an innovation component pertain to the individual signal. As in equation 4.4 The component which is common to all signals is Z_C , and innovation component is Z_l .

$$x_1 = Z_C + Z_1$$

$$x_2 = Z_C + Z_2$$

⋮

$$x_l = Z_C + Z_l, l \in \{1, 2, \dots, L\}$$

Where K and K_l are the corresponding values of Z_C and Z_l respectively, and $y_l = A \times x_l$. The encoded signal further transmitted to the BS. At the BS, joint recoveries of the signals are done, because of intra and inter correlation effects the original signals can be recovered with slightly reduced number of measurements than separate recovery. In order to recover the jointly sparse signals, we conduct simulations, considering the synthetic signals as well as considering the real data . In both cases joint recovery proves better than separate recovery as joint decoding exploits intra and inter correlations.

4.3.2 Recovery of Jointly sparse signals

In this section we discuss the recovery algorithm for JSM. In JSM-1 the signals will share the same common non zero coefficients (the location of these elements will be same) but the amplitude and phase might be different. We can term this as common support set. These signals further posses different non-zero components(location of these elements need not be same). We can term these elements as innovation sparsity. But where as in JSM-2, all the signals coefficients are different (common sparse+innovation) but the location of these components are exactly the same. Where as JSM-3 models, consist of a non-sparse common component

and a sparse innovation component [Hormati and Vetterli \(2008\)](#). Large scale WSN signals can be modeled using the JSM, where global variations i.e sun, temperature, humidity, wind affect the sensors collectively but local variations such as water flow, shade, animal/human presence affects the smaller group of sensors. Recovery of CS signals can be either based on greedy or gradient based algorithms.

For our initial simulation we consider the synthetic signals which represent the real WSN scenario, by considering multiple sensors and recovery is based on joint recovery. Each sensors sensed data is transmitted to the BS, and by joint recovery procedure the intra and inter correlations are explored. Simulations are carried out using YALL1 package [Zhang \(2009\)](#) in MATLAB which uses basis pursuit to recover the signals. In order to show the difference between joint recovery and separate recovery we have used Orthogonal Matching Pursuit (OMP) [Tropp and Gilbert \(2007\)](#) and Simultaneous Orthogonal Matching Pursuit (SOMP)[Tropp et al. \(2005\)](#) which is summarized below.

4.3.3 Separate recovery using OMP

In order to validate DCS we retrieve the data at the BS separately, i.e decoding each sensor data but result decision is on the basis of collective recovery. We decode each branch of the sensor data in a group of say 'N' sensors but we declare success after all the 'N' sensor data's are separately recovered using OMP. The 's' sparse signal $x_l \in R^n$, from l sensors, where $l = \{1, 2, 3, \dots, L\}$ are encoded using the measurement matrix Φ where $m < n$, generating the measurements $y_l = \Phi x_l \in R^m$. Let Φ_j be the jth column, $j \in n$, $n = \{1, 2, \dots, n\}$. The measurement vector y_l is generated by the linear combination of 's' columns of Φ . Thus estimation of 'x_l' is identifying those columns of Φ . In OMP this problem is solved in a greedy fashion, at each iteration of OMP algorithm selects the column which is mostly correlated with the residual of y_l . Further it eliminates the significance of this column to compute updated residual. Figures .4.1 and 4.2 show the result of separate recovery for L=8 and 16.

4.3.4 Recovery using Simultaneous- OMP

In order to recover jointly sparse signals simultaneous greedy approximation has been proposed [Tropp et al. \(2005\)](#). The algorithm is much similar to other greedy methods with minor changes. This in general is known as DCS-SOMP. The algorithm is as follows. The procedure

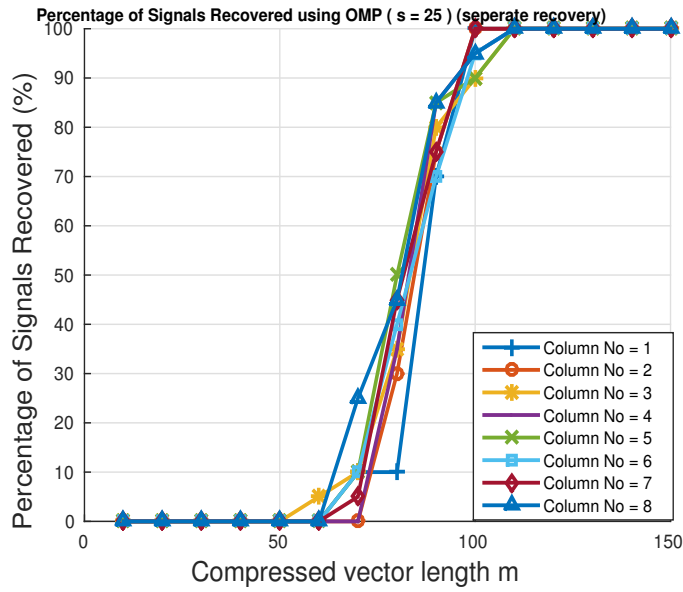


Figure 4.1: Separate recovery using OMP, convergence in case of L=8.

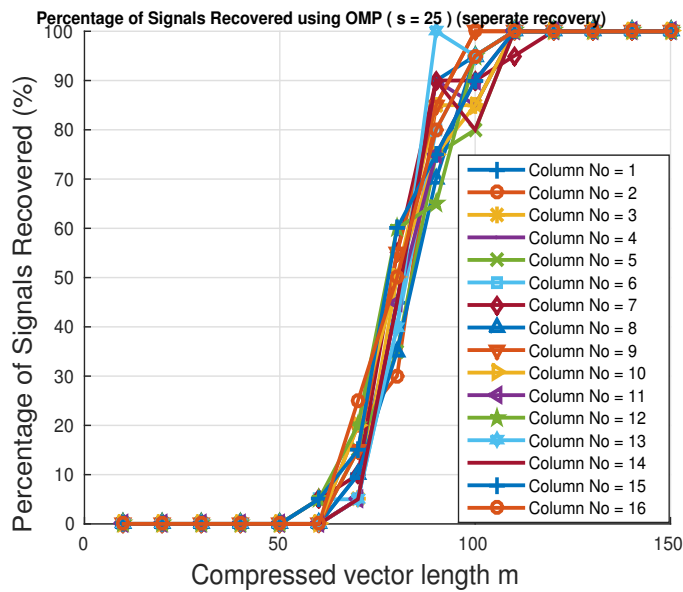


Figure 4.2: Separate recovery using OMP, convergence in case of L=16.

of S-OMP is similar to OMP except the fact that here there are 'L' compressed samples y_l where $l=\{1, 2, 3...L\}$. And SOMP reduces to OMP when $L=1$.

Simultaneous Orthogonal Matching pursuit for jointly sparse signals

Input: Measurement vector $\{y\}_l$, Measurement/sensing matrix $\{\Phi\}_l$,

sparse vector $\{x\}_l$, $l \in \{1.2.3...L\}$, sparsity = K

Initialization: residual $\{r_{l,0}\} = \{y_{l,0}\}$, $k = 1$, $\{\hat{x}\}_l = 0$, $\hat{\Phi}_0 = \emptyset$

index set = $\Delta_s = \emptyset$.

\emptyset = empty set, δ = small constant.

Repeat until stopping criteria holds

(1) Identify the column vector C_s which is highly correlated with the residual.

$$\delta_k = \underset{i=1,2,\dots,N}{\operatorname{argmax}} \left| \langle r_{l,k-1} \Theta_{li} \rangle \right|$$

(2) Update the index set by augmenting it with chosen column along with the matrix

$$\Delta_k = \Delta_{k-1} \cup \delta_k$$

$$\Phi_k = [\Phi_{k-1}, \Phi_{\delta_k}]$$

(3) Estimation of the sparse signal by computing least square problem

$$x_{l,k} = \underset{x}{\operatorname{argmin}} \|y_l - \Phi_k x_l\|_2$$

(4) Update the residue

$$\hat{y}_{l,k} = \Phi_k x_{l,k}$$

$$r_{l,k} = y_l - \hat{y}_{l,k}$$

(5) Advance the counter, if $k < K$ go to step 2

End

Output: $\{\hat{x}\}_l$, $\{\hat{\Phi}\}_l$

4.4 Results and analysis

To carry out the simulations, we consider the synthetic data, which is similar to the real world data with $N=512$ and $L=8$. The signal under consideration comprises a common component K_c , which is sparse in DCT basis which symbolizes to the common temperature as well as an innovation component which represents the abnormalities in the temperature reading. Plotting the values of signal recovery versus the measurements M required to compress the signal. The success is declared depending upon the recovered signal, i.e $\epsilon = \|\hat{x} - x\|_2 / \|x\|_2 \leq 10^{-2}$. Figures.4.3 and 4.4 depict the performance of YALL1 and SOMP for various values of K_c and K_l with $N=512$, sparsity of the signal = 75, $L=8$, across the number of measurements.

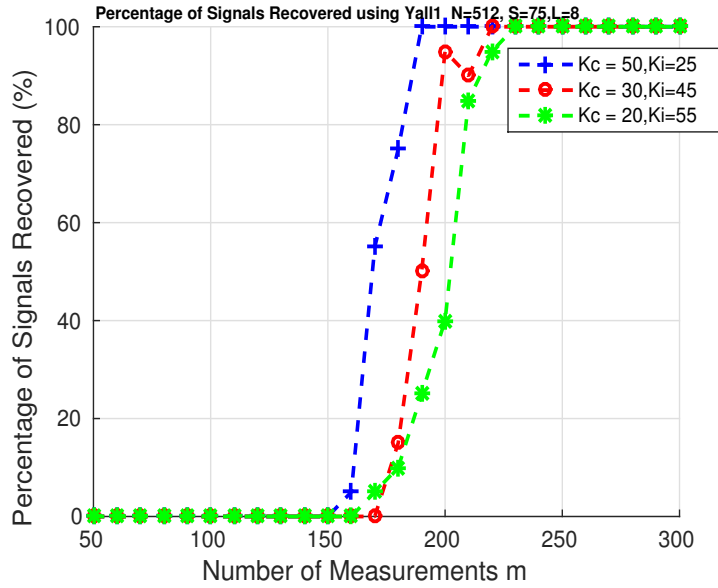


Figure 4.3: Joint recovery using YALL1 $L=8$, by considering different values of K_c and K_l

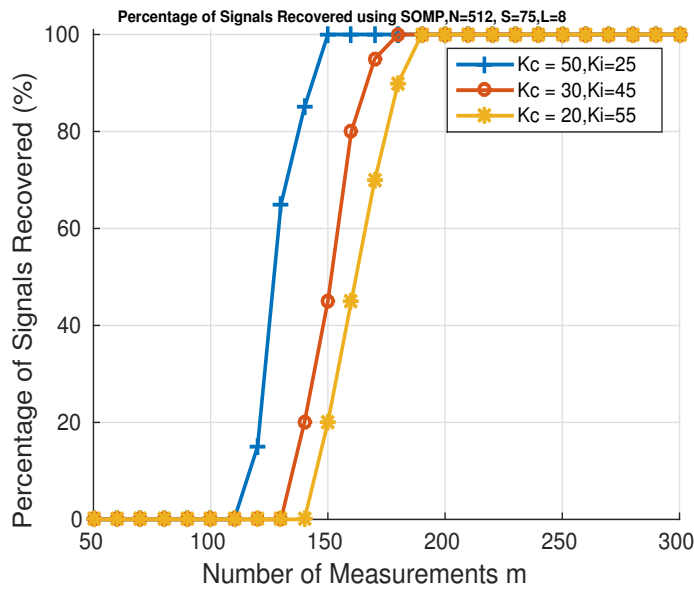


Figure 4.4: Joint recovery using Somp $L=8$, by considering different values of K_c and K_l

We know that in CS as the sparsity increases the probability of exact reconstruction decreases. Thus we try to figure out the relation of sparsity and exact reconstruction for jointly sparse signals and then compare the same with separate reconstruction. Recovery of the jointly sparse signals, using YALL1 and SOMP are presented in Fig.4.5, Fig.4.6 and Fig.4.7 by letting $N=512$, $M=256$ and varying 'L'. As the number of sensors (L) increases, even for less sparse data, recovery is assured which can be concluded as shown in figures 4.5 and Fig.4.6. But if you consider the separate recovery using OMP there is no considerable changes, and as L

increases the result is opposite as in the case of jointly sparse recovery. In both cases we try to solve $y_l = A_l X_l$, but in joint recovery the support set is shared, thus even though 's' increases recovery can be done with reduced values of measurement, but same is not true with separate recovery.

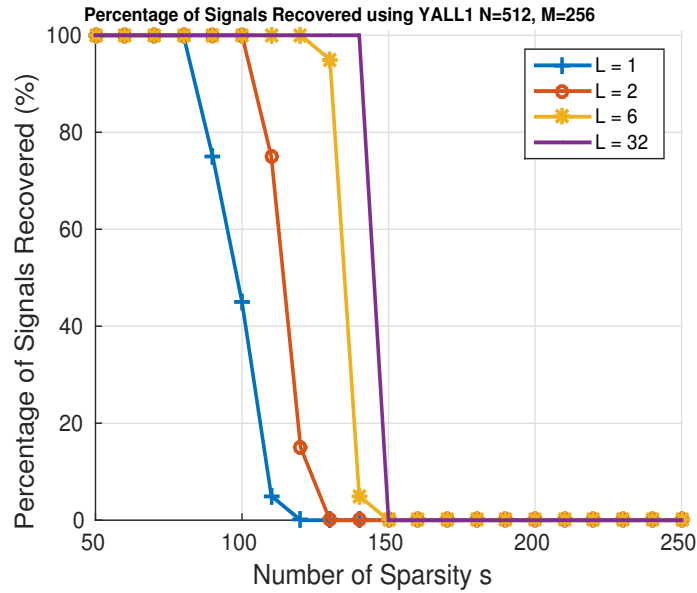


Figure 4.5: Joint recovery using YALL1 ,for varying values of L and sparsity

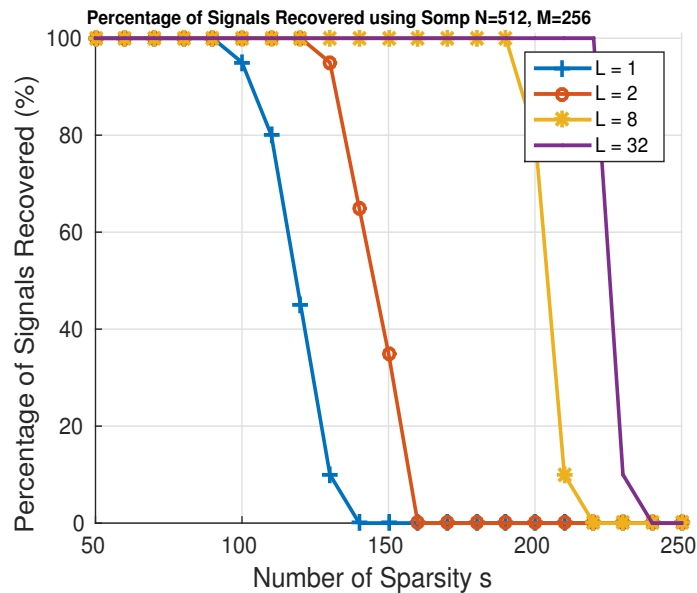


Figure 4.6: Joint recovery using SOMP,for varying values of L and sparsity

Figures.4.8 and 4.9 show the result of joint recovery using YALL1 and SOMP by varying the number of measurements and numbers of sensors (L). With the increase in the number of

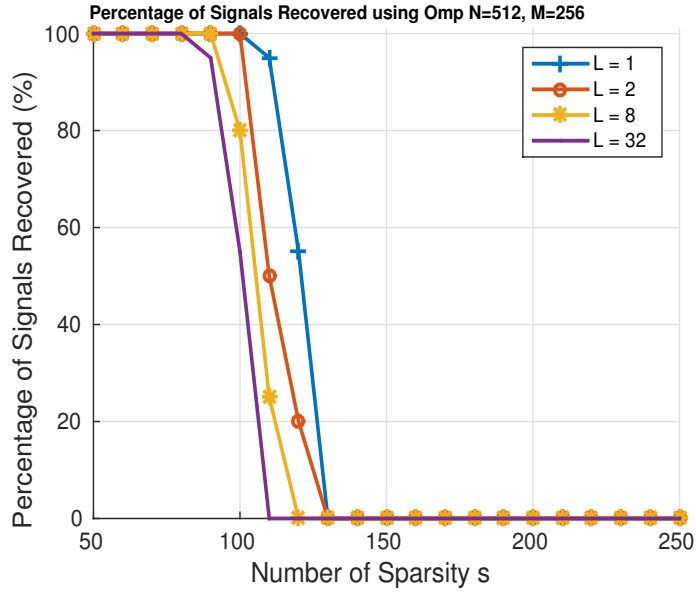


Figure 4.7: Joint recovery using OMP, for varying values of L and sparsity

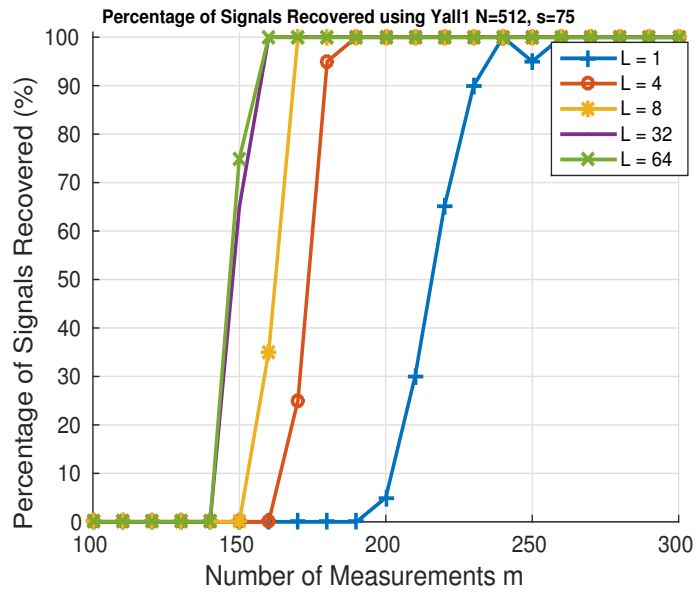


Figure 4.8: Joint recovery using YALL1 ,for L=1,4,8,32,64

sensors, there is a decrement in the number of measurements required for reconstruction. A sparse signal with sparsity $s=75$ has been considered, with L varying from L=1 4 8 32 64 as in Fig.4.8 and Fig.4.9. There is a drastic decrement when we consider L=1 and 4. Further when we consider L=32 and 64, the measurement required is almost same. More details about the lower bound for the measurement required for jointly sparse signals can be found in [Baron et al. \(2005\)](#)

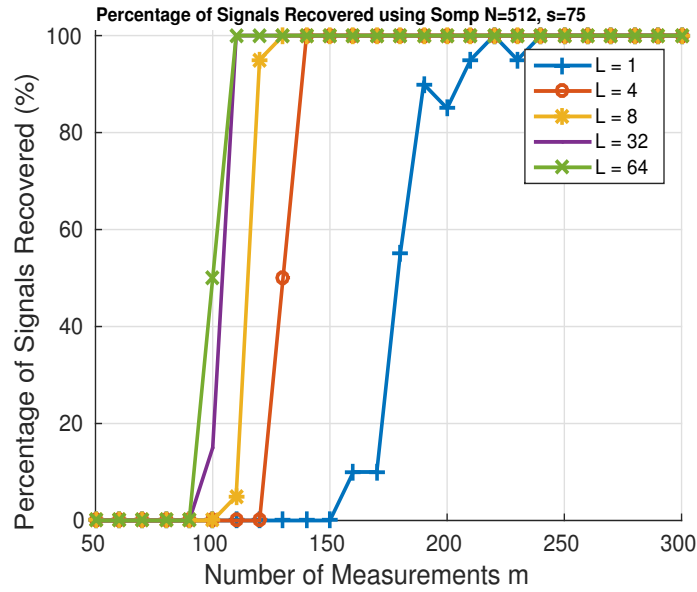


Figure 4.9: Joint recovery using Somp ,for L=1,4,8,32,64

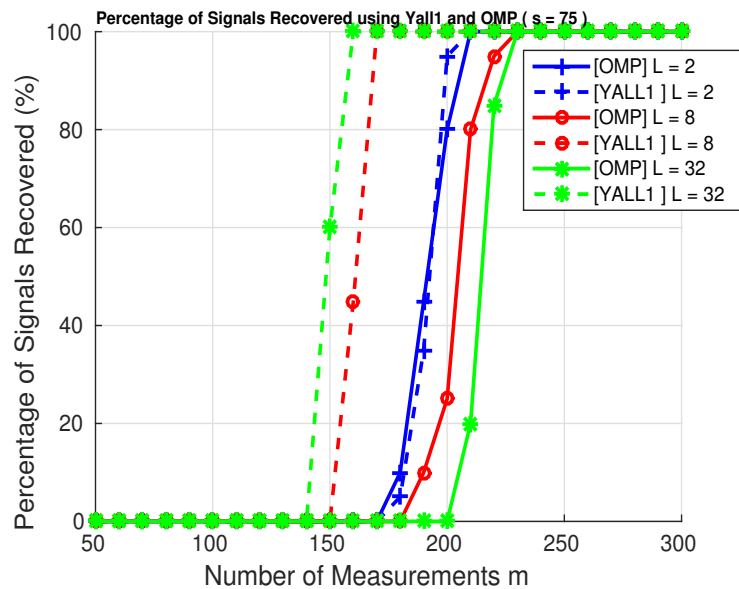


Figure 4.10: Joint recovery using YALL1 and separate recovery using OMP ,for L=2,8,32

Fig.4.10 shows recovery using YALL1 and separate recovery using OMP and Fig.4.11 using SOMP and separate recovery using OMP by varying number sensors (L) and the number measurement (M) by letting sparsity of the signal $s=75$. Thus if the signals are correlated then joint recovery promises reduction in the measurements required for reconstruction. Thus by joint recovery we can reduce the number of transmission required by the sensor with out communicating to each other. If the sensor signals are correlated then joint recovery has an advantage which in turn reduces the burden of sensor nodes thus helps to improve overall network

lifetime. Fig.4.12 shows joint recovery using SOMP and separate recovery using OMP for $L=1, 2, 4, 8, 16, 32$. As from the above results we can conclude if the recovery is separate then the number of measurements increases with number of sensors. In separate recovery only the intra correlation is explored, but in joint recovery intra as well as inter correlations are explored.

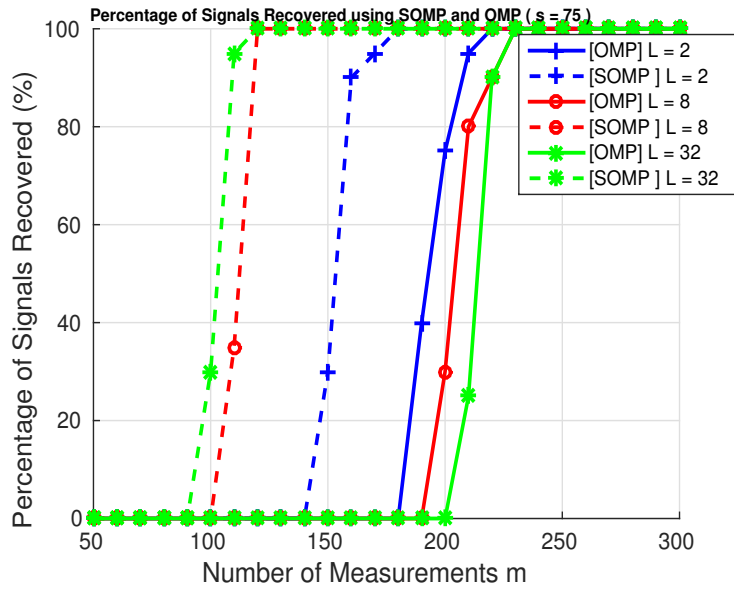


Figure 4.11: Joint recovery using SOMP and separate recovery using OMP ,for $L=2,8,32$

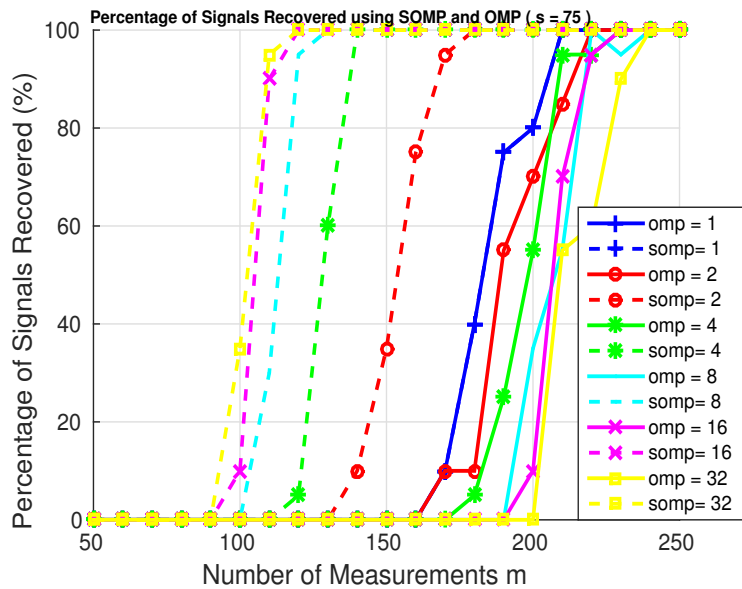


Figure 4.12: Joint recovery using SOMP and separate recovery using OMP, for $L=1, 2, 4, 8, 16, 32$

In this section we considered the data set from Berkley lab ,which is recorded in an office environment which exhibits regular variation during day and night time. Figures.4.13 and 4.14 show the result of joint recovery, by considering the real data set-I and Figures 4.15 and 4.16

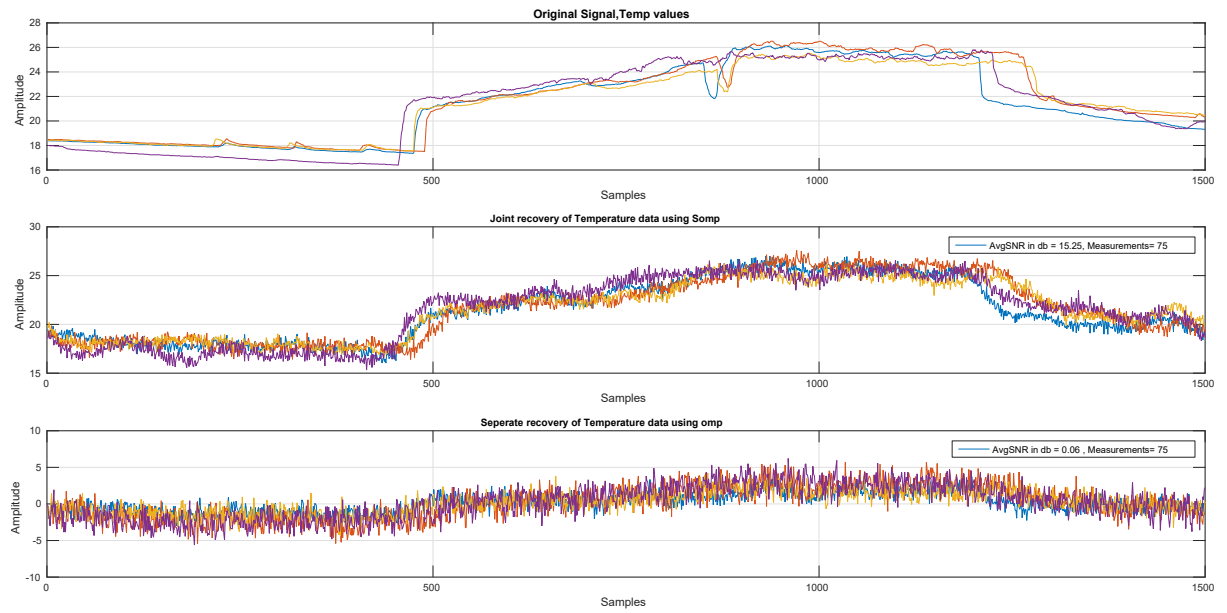


Figure 4.13: Joint recovery using SOMP by considering real data-I,M=75

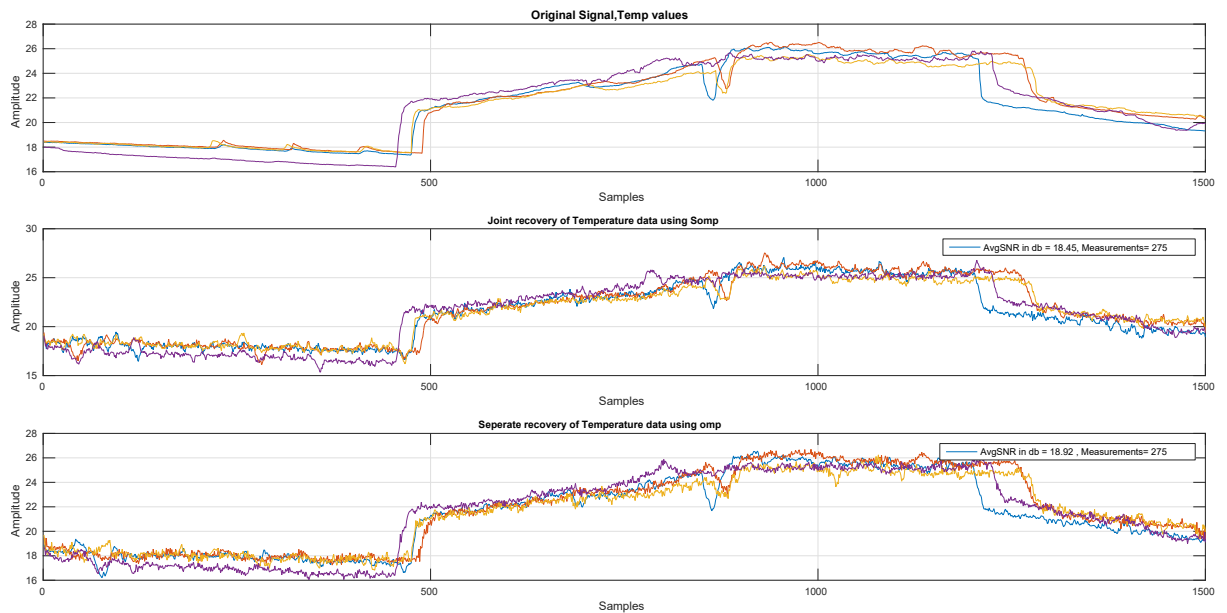


Figure 4.14: Joint recovery using SOMP by considering real data-I,M=300

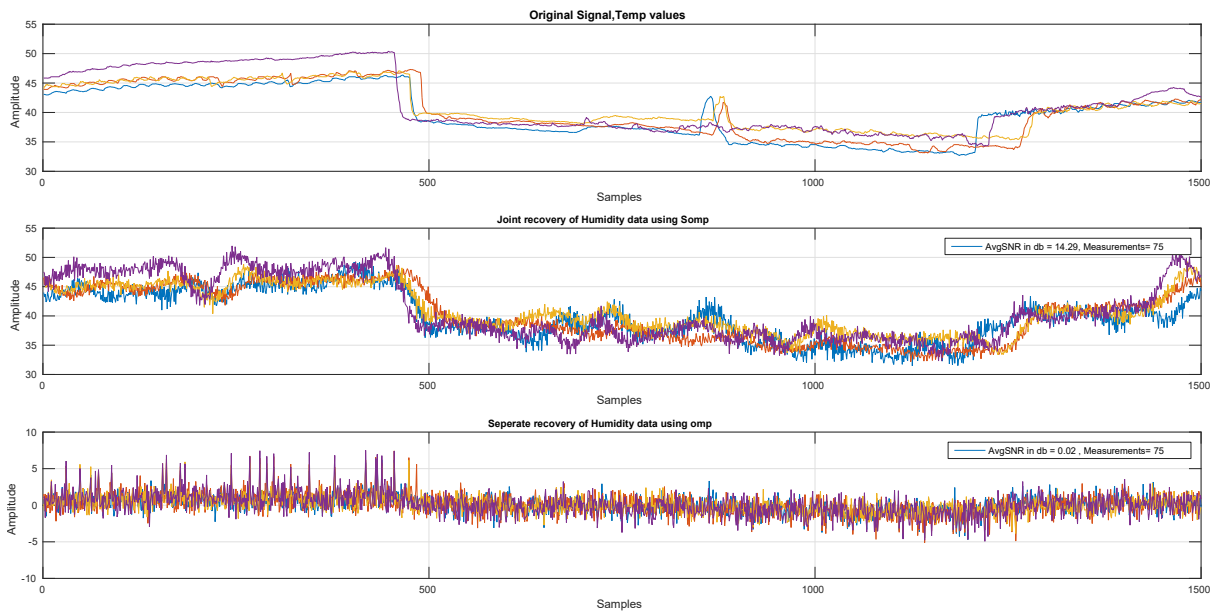


Figure 4.15: Joint recovery using SOMP by considering real data-II, $M=75$

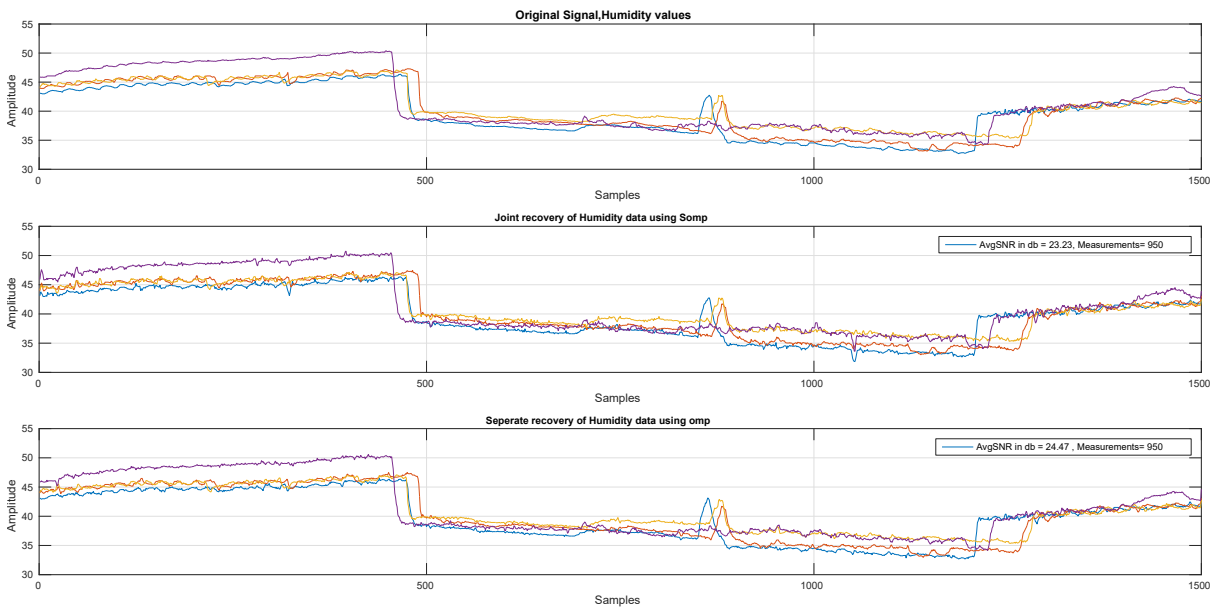


Figure 4.16: Joint recovery using SOMP by considering real data-II, $M=950$

for dataset-II. The signal is not exactly sparse. The signals have smooth variation in space and time. If we consider the DCT, the signal can be represented as sparse, and can be modeled using JSM-2 model.

Figures.4.13 and 4.14 depict recovery based on joint recovery and separate recovery techniques for dataset-I, with $N=1500$ and $M= 75$ and 275 respectively. With the SOMP as reconstruction methods we are able to recover the signal with SNR of 15.25 but if we consider separate recovery, the data is unrecoverable. Thus we increase the number of measurements to 275, we could successfully recover using separate reconstruction. But if we consider the data recovery using separate reconstruction, with the same constraints it is impossible to recover unless there is an increase in the length of compressed vector (y). When we considered dataset-II, with $N=1500$ and $M=75$, using SOMP we are able to recover even with reduced number of measurements, SNR= 14.29 and using separate recovery it s not possible to recover. But in this case, $M=950$, This depends on the sparsity of the data in the transformed domain. In this particular data set sparsity is low thus it requires more number of iterations to compute the coefficients while reconstruction using OMP.

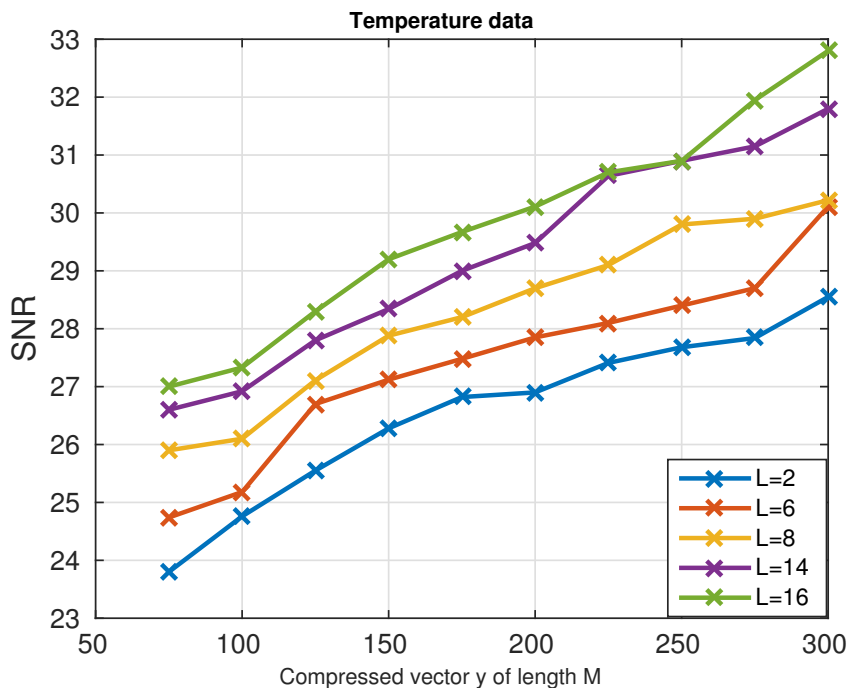


Figure 4.17: Joint recovery using SOMP by considering Temperature data (outdoor) $N = 1024$

Figures. 4.17 and 4.18 show the SNR versus compressed vector length (y) in case of real signal (outdoor). In this case we have considered $N=1024$ and varying values of M i.e com-

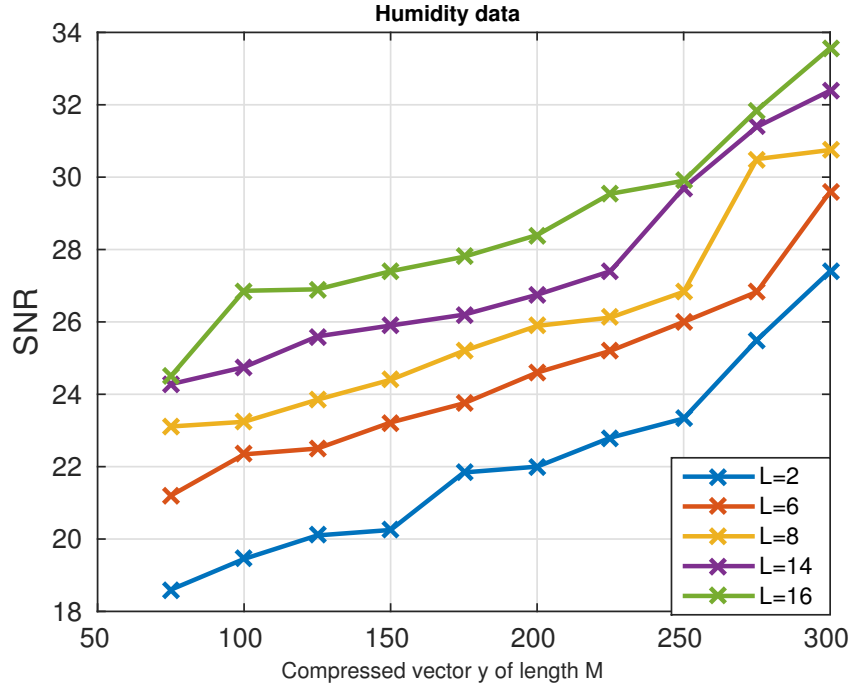


Figure 4.18: Joint recovery using SOMP by considering Humidity data (outdoor) $N=1024$

pressed vector length. This outdoor data set is taken from sensor-scope [Ingelrest et al. \(2010\)](#) which contains the temperature as well as humidity data sets. The plot depicts variation of SNR for different values of compressed vector for different values of L . As we can see from the plots using joint sparsity models it is able to recover the data with lesser values of y . There is an improvement in the quality of the signal when we increase the number of sensors. When we consider separate recovery it is not possible to recover the original signal with lesser value of the compressed vector y .

4.4.1 DCS for multi-channel EEG

The electrical activities related to the brain are measured using EEG signals. Various types of neurological disorders are detected using EEG which include epilepsy, sleep disorders, stroke, dementia etc. The activities of brain are recorded, through the electrodes which are attached to the scalp of the person. The multi-channel and multi trial EEG generates huge amount of data, which has to be either stored or transmitted. The aggregation method prior to storage/transmission motivates to employ CS to EEG signals. In literature there are several papers on, CS application in EEG signals. The performance evaluation of those CS-oriented system is dominated by two main metrics, employed recovery and the domain of sparsification. In this case for EEG signals, we consider wavelet transform as the sparsifying basis. By using

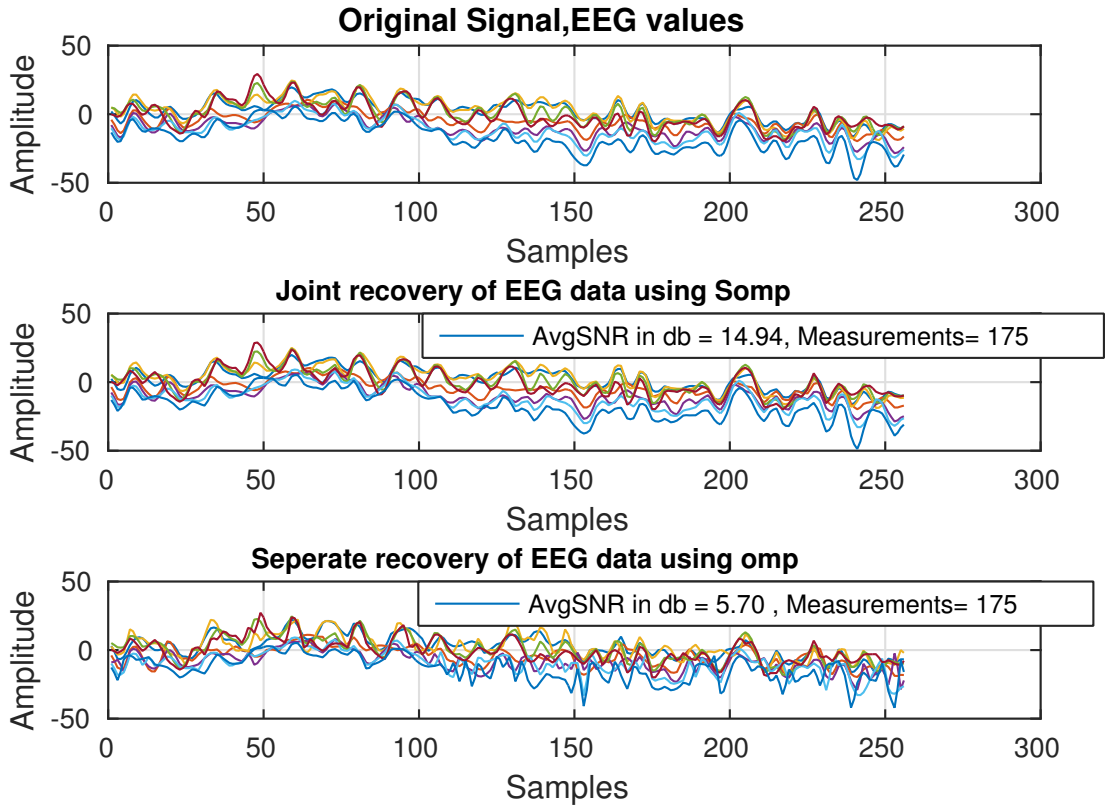


Figure 4.19: Joint recovery using SOMP by considering real data-III, $M=175$

CS frame work, we compress and reconstruct the multichannel EEG signals. Fig. 4.19 shows the 8 channel EEG signals with $N=256$, using wavelet transform as sparsifying basis.

4.5 Conclusion

The idea of joint sparse models, fits into WSNs, as the nodes are correlated among each other. In WSN power consumption is an important parameter to consider, as to reduce the exhaustion, and increase network lifetime. How we can save the transmission power by reducing the measurements needed to recover the signal at the receiver, has been presented in this work. Since we have used joint sparsity techniques, we could explore the intra as well as inter correlation among the data signals. By comparing the jointly sparse signals with separate recovery, we demonstrated how efficiently data can be reduced in joint sparsity techniques. We have simulated the results by considering the synthetic signals as well as real data, which proves that by using joint sparse model we can further reduce the number of measurements needed to recover the data.

CHAPTER 5

ENERGY EFFICIENT, SECURE AND RELIABLE DATA COLLECTION IN WSNs

5.1 Introduction

WSN applications are becoming ubiquitous in this era of highly networked life. With advances in technology, today's sensor nodes can gather and process more data; leading to the creation of new applications over recent years. To achieve cost effectiveness, sensor nodes are not typically equipped with tamper-proof facilities [Lou and Kwon \(2006a\)](#); [Liu et al. \(2012a\)](#); [Challal et al. \(2011\)](#); [Shi and Perrig \(2004\)](#); [Deng et al. \(2006\)](#). Due to the hostile and unattended environments, there is a very high chance of nodes being compromised in a WSN. Due to the relative ease with which node compromise can be achieved, it is a key part of many unique network insider attacks [Shi and Perrig \(2004\)](#). A Compromised Node (CN) attack is an attack in which an adversary compromises a certain subset of nodes to passively intercept data packets traversing the compromised nodes [Lou and Kwon \(2006a\)](#); [Liu et al. \(2012a\)](#).

We validate the vulnerability of secret sharing schemes under the relaxation of a secure area around the BS, by a combination of energy efficient Shamir's Ramp Secret Sharing (SRSS) method and round reduced AES symmetric encryption, termed as 'Split Hop AES (SHAES)' to address the CN attack problem. We analyse the energy efficiency and security of the proposed approach through theoretical analysis. It shows that the proposed combination achieves both semantic security and reliability in an energy efficient way.

Reliability in multipath routing is often achieved with the help of data redundancy. Typically, reliability is achieved in multipath routing by creating multiple copies of the same data and routing them in different paths. When security is combined with reliability in multipath routing, creating copies of data increases the chances of an adversary accessing the data, unless some security mechanism is used (like encryption). A common method for combining reliability and security in multipath routing is to split the data based on secret sharing schemes then send the shares on different paths to reach the BS. In order to achieve greater security, previous

works have used the approach of dispersing the shares randomly and then sending the data towards the BS. Original data is reconstructed only when the required number of shares reaches the BS. The higher the dispersion of shares, the higher the associated communications and thus, the higher the communication energy drain. Even after investing more communication energy in dispersion, security achieved may be lower because all the shares must converge at the BS. Therefore the dispersion of shares may not completely solve the security problem when multi-hop communication routing is used and when the shares converge to single BS. Therefore we opted not to disperse the data for security purposes and thereby reduced the communication energy drain. Instead of dispersion, the approach followed in this work invests a small amount of computation energy to achieve better security over the entire network including the area near the BS.

A brief overview of the approach followed in this work to achieve energy efficient secure reliable data collection is presented as follows. When any node generates data to send, it acts as the Source Node (SN). Every node has a routing table with entries of one-hop Cluster Heads (CHs) belonging to each path. The neighborhood CH table would be created during the route setup phase and is used when a node is acting as a source node. The SN splits the data into shares using SRSS and forwards the shares by adjusting itself to the minimum available intra cluster transmission power levels associated with each of its one-hop CHs belonging to different paths. After receiving the shares from the source node, the one-hop neighboring CHs of different paths will encrypt their shares using SH-AES (provided that the adversary is not located near the SN, otherwise SH-AES encryption is performed on shares at the source node itself). CHs adjust their transmission power levels to corresponding inter cluster transmission power levels associated with each path and then send the data to the next CHs of the same path. CHs on each path use higher inter cluster transmission power levels and sends the data to the next CHs until the shares reach the BS. CHs on each path are unique, in other words disjoint nodes. Therefore a node-disjoint multipath with variable multihop routing is used to route the data. The BS has complete information of symmetric keys associated with each sensor node. After receiving the encrypted shares, the BS will decrypt the shares using the same symmetric key associated with each node that has encrypted those shares. Now the required numbers of decrypted shares are used for the reconstruction of original data. Since only threshold numbers of shares are required for the reconstruction of data, the network sustains the loss of few data packets and achieves desired reliability. The complete process of combining SRSS and SH-

AES is presented graphically in the Figure 5.1.

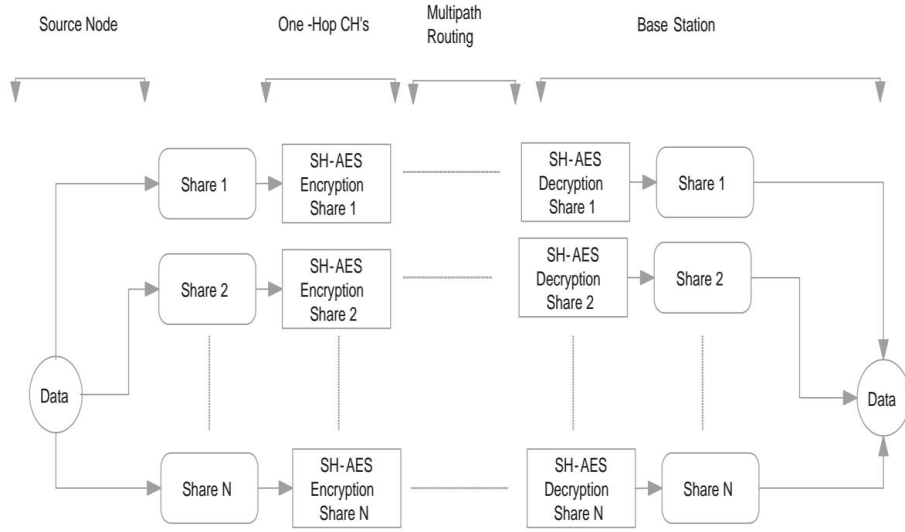


Figure 5.1: Overview of SRSS and SH-AES combination process

5.2 Secret Sharing Schemes - Overview

Shamir's (t, n) threshold secret sharing (SSS) scheme splits the secret data into n shares (Lou and Kwon, 2006a). Out of n shares, only t shares are required to reconstruct the complete original data. Any $t - 1$ shares reveal no information about the message. This desirable property of SSS scheme helps to generate the data redundancy required to achieve reliability and yet provides information theoretic security. Share generation is quite simple and is obtained by evaluating the polynomial of degree $t - 1$ under a Galois Field (GF) given by equation 5.1 Stinson (2005).

$$S = (a_0 + \sum_{i=1}^{t-1} a_i x^i) \quad (5.1)$$

a_0 the secret data.

a_i the random data and

S are the shares generated for each value of x

Reconstruction of shares can be achieved by Lagrange's interpolation method as explained by Stinson (2005), and is achieved at the BS, which is not usually constrained by resources. In

order to achieve information theoretic security, the data coefficients are used only at the a_0 position. This leads to an increase in the required communications to convey overall data.

In order to reduce the overall communications and thereby reduce excessive energy drain, Shamir's (t_1, t_2, n) Ramp Secret Sharing (SRSS) as explained by [Stinson \(2005\)](#) can be used in WSNs. SRSS allows more data to be used in the polynomial computations of S given by equation 5.1 by replacing t with t_2 . The first t_0 values of equation 5.1 are obtained from the secret data and remaining $t_1(t_1 = (t_2 - t_0))$ values are obtained from the random data. Thus, no information about the message is leaked until an adversary is able to access the t_1 shares. If an adversary has $t_1 + 1$ shares or more, then information leakage begins and increases until it reaches t_2 shares, where the complete information is obtained ([Stinson, 2005](#)).

Assumptions

Sensor nodes are not equipped with tamper-proof facilities and are prone to be compromised by adversary and can perform SRSS and SHAES operations having unique 128bit keys. The BS is always secure with unlimited energy, processing power and having the complete knowledge of the unique 128-bit key associated with each node. Compared to other previous related works, the assumption of the secure area around the BS is relaxed in our approach.

5.3 Proposed work

5.3.1 Near-Sink CN Attack

The security of secret sharing schemes lies in the divergence of the shares from the adversary. If the threshold shares converge near the adversary then secret sharing schemes can't provide any security, as the adversary can reconstruct the secret data. In sensor networks, all the data needs to be collected at the BS, therefore all shares need to converge at the BS. Since a wireless medium and multihop communication is used for the data communication, nodes near the BS will forward all the shares that reach the BS. If the adversary compromises a few nodes near the BS, then the adversary can attempt to get the required shares for reconstructing the complete data. We term this type of attack as a near-sink CN attack. The near-sink CN attack can compromise the security achieved by secret sharing schemes used in WSN applications. In order to validate the near-sink CN attack, two deployment strategies are considered:

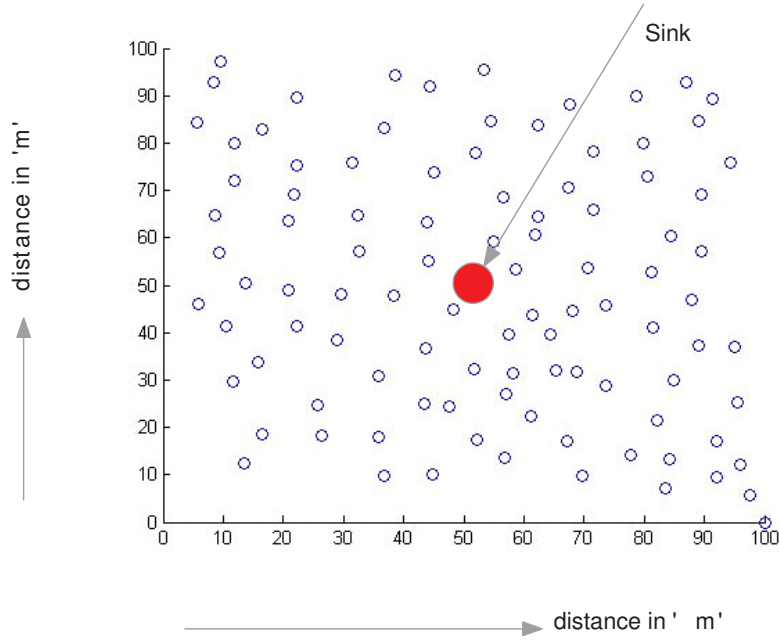


Figure 5.2: Centralized sink deployment

Centralized sink deployment, wherein the BS is located at the center of the random nodes, and Corner sink deployment, wherein the BS is located in the corner of the network area containing randomly deployed nodes. These strategies are shown in Figure 5.2 and Figure 5.3.

A single node is assumed to be compromised by the adversary near the BS. To study the effect of a near-sink CN attack with respect to its distance from the BS, the compromised node is located at different distances from the BS. The transmission power level and communication channel path loss model determines the range and successful reception of data. The near-sink CN attack is tested with different transmission power levels: 0,-1,-3,-5 in dBm. The channel is characterized by the log normal path loss model having a path loss exponent $\eta = 2.4$ and with slow fading characterized by different standard deviation σ values of 0,1,3,5 in dB. Simulations are carried out in the Castalia, a discrete network simulator, with nodes near the BS communicating the shares to the BS and having a receiver sensitivity of -95dBm. Figure 5.4 clarifies that under the centralized BS scenario, the circumference area around the BS having a radius of 20m can overhear 80% of the shares under all transmission power levels. With a transmission power level of 0 dBm, a compromised node can overhear 75% of shares in the perimeter area around the BS having a radius of 45m. Therefore a near-sink CN attack is a prominent attack in the applications of WSNs that use secret sharing schemes. Furthermore, from figure 5.4 it is evident that the monetary burden of equipping video cameras to monitor the

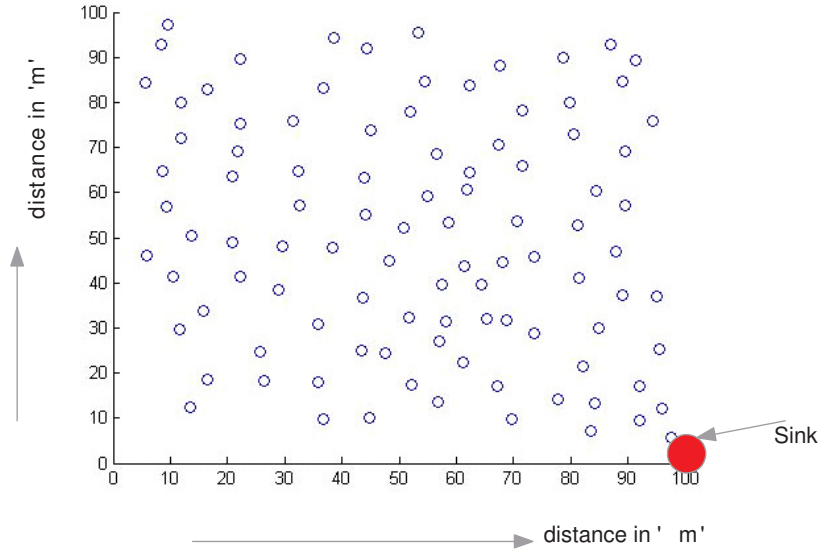


Figure 5.3: Corner sink deployment

area around the BS could be high, since a large area needs to be monitored. The near-sink CN attack is even more effective in the corner-deployed BS model compared to the model where the BS is deployed at the center of the network area. When the BS is deployed to any of its network corners, the area around the BS is smaller. Therefore, shares converge well before they reach the BS. Figure 5.5 justifies that at the 0 dBm power level, a compromised node that is 50m away from the BS can overhear as much as 90% of shares received by the BS. Although the security achieved by the SSS is higher than the SRSS, the communication energy efficiency required for sensor networks can not be achieved. If the required number of nodes is compromised anywhere in the network area or specifically the nodes near the BS (near-sink CN attack), then security is compromised. If we use SRSS communication, then energy efficiency can be achieved, but relatively smaller number of compromised nodes can compromise the security. Neither schemes are sufficient to address the problem of an adversary compromising nodes anywhere in the network (CN attack), occurring most often near the BS. From Figure 5.4 and 5.5, it is evident that the near-sink CN attack is effective with secret sharing schemes in WSNs under the relaxation of a secure area around the BS. Therefore secret sharing schemes alone can't provide security to the sensor network applications. One way to address this attack is to encrypt the shares. Therefore, in this research work a round reduced AES symmetric encryption termed as SHAES is considered to encrypt the shares.

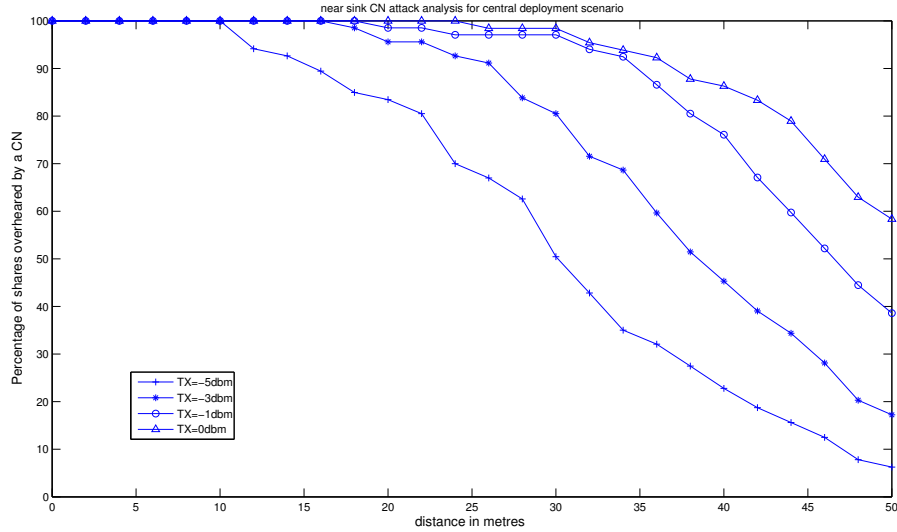


Figure 5.4: Near-Sink CN attack on Centralized Sink deployment under different power levels

5.4 SHAES scheme

In 2000, the NIST (US National Institute of Standards and Technology) selected the Rijndael algorithm submission as the AES (Advanced Encryption Standard) and since its adoption it has become a *de facto* cryptographic standard in many areas such as banking, administration and others (Daemen and Rijmen, 2002). The evaluation criteria for the AES selection were based on security, cost and algorithm and implementation characteristics. During the selection process and over the last decade, AES has proved its security through an extensive cryptanalysis. We make use of this extensive literature available on AES cryptanalysis for the security analysis of our round reduced AES.

AES is even suitable for devices that work on 8-bit processors or micro-controllers, having low program memory and a restricted amount of RAM (Random Access Memory) for working memory. An IEEE 802.15.4 standard that recommends the use of AES in low power devices has reiterated the versatility of this cipher. Today low power sensor nodes often follow the IEEE 802.15.4 standards and commercial manufacturers of sensor nodes are coming out with inbuilt AES modules in their sensor nodes. This creates a strong motivation for the use of an AES cipher in our approach for secure reliable data collection for sensor networks. Although AES in its original form is energy efficient, there is a scope to modify the AES cipher. Since sensor networks are application-specific, the security concerns and desired level of security may vary with the application-specific needs (Karlof and Wagner, 2003). For example, military and

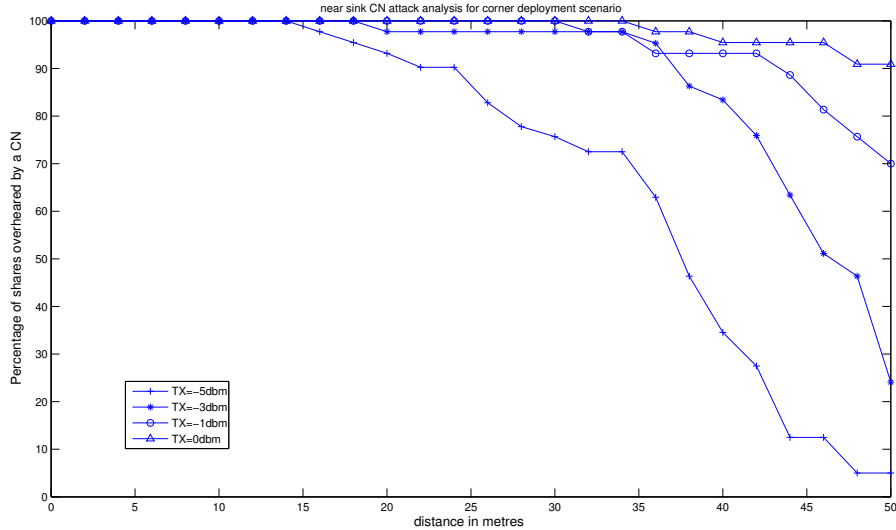


Figure 5.5: Near-Sink CN attack on Corner Sink deployment under different power levels

health care applications will require very high security compared to the environmental forest fire detection application.

AES is an iterated block cipher with round keys. It consists of the repeated use of a round transformation on the 128 bit plaintext represented in the form of state matrix having four rows and columns, where each entry in the matrix is of bytes having a value in $GF(2^8)$ (Daemen and Rijmen, 2002). We restrict the explanation to 128-bit key AES because our proposed variant is of a 128-bit key. The round transformation consists of four steps operated on the state matrix.

1. SubByte (SB): the same invertible S-box of matrix 16×16 is applied on each byte of the state matrix.
2. ShiftRows (SR): Each row of the state matrix is cyclically shifted left based on the offset i ($0 \leq i \leq 3$).
3. MixColumns(MC): Each column is multiplied by a constant 4×4 matrix having the values in $GF(2^8)$.
4. AddRoundKey (AK): The 128-bit state matrix is xored with a 128-bit round key generated by the key schedule.

There is one extra AK which operates on the initial 128-bit state matrix, acting as the key whitening, then the nine full rounds having all four steps mentioned above and in the final round the MC is omitted. The key expansion and remaining details of AES can be found in the book by Daemen and Rijmen (2002).

The SHAES is a reduced round version of AES with 3 full rounds and the MC omitted in the last round. The initial AK is retained and it acts as the key whitening. From the key schedule we require only 5 round keys to be generated to be used in the AK step. We use a separate name for this round-reduced version of AES:SHAES. Since sensor networks use multihop communication Hop is used. Rather than using complete AES, we split it into reduced round of 4 Split is used. This encryption is used in a somewhat different manner compared to the normal AES and is used in combination with secret sharing. The SHAES is pictorially represented in figure 5.6.

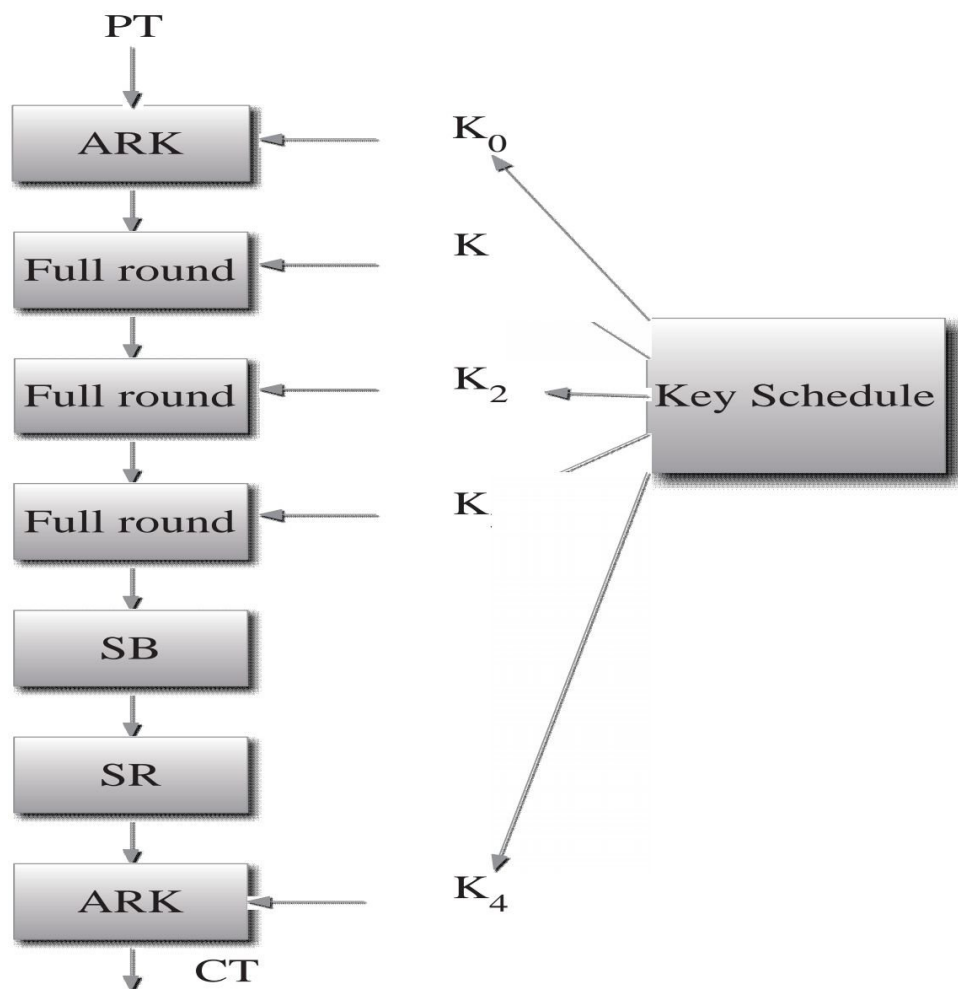


Figure 5.6: SHAES graphical representation

Doomun and Soyjaudah (2009) give a detailed analysis of the basic processing operations required for the AES cipher, and the processing overhead and energy consumption of the AES cipher has been presented. They also compare its performance with the RC5 encryption and conclude that the AES is best suited for resource constrained devices. The inventors of AES,

in their book on AES (Daemen and Rijmen, 2002), have mentioned that the MC operation is not preferred for 8-bit micro controllers or processors because it consumes the highest energy among all other operations. The same can be seen in the paper by Doomun and Soyjaudah (2009) where they show that the MC operation consumes 63 % of the total computational time, 75 % of the total AND operations, 72 % of the total OR operations and 95 % of the total shift operations required to complete the 10 round transformations. These basic operations: AND, OR and shift, indicate the energy consumption. It can be easily seen that MC is the highest energy consuming operation among the round transformation operations. Furthermore, the fourth round MC does not affect the linear and differential cryptanalysis bounds (Daemen and Rijmen, 2002). Therefore, the MC operation has been removed from the fourth round of the SHAES.

Doomun and Soyjaudah (2009) mention that 10 rounds of encryption plus the key schedule generating 11 subkeys of the AES consumes 3638 clock cycles and (23.6 μ J) of energy. These values are for the complete AES, but we are using only 4 rounds of the AES and the corresponding 5 round keys generated by the key schedule. There is a reduction of 6 round transformations from encryption and 6 round key generations from the key schedule compared to the original AES. Based on the results quoted by Doomun and Soyjaudah (2009), we can conclude that in the SHAES, 1560 clock cycles [obtained by $(3638/(10 + 11)) \times (5 + 4)$] and (10.11 μ J) of energy [obtained by $(23.6/(10 + 11)) \times (5 + 4)$] are consumed.

5.5 Security analysis of the SHAES:

The two rounds of AES provide full diffusion, i.e. any one change in a state bit affects half of the total state bits after two rounds (Daemen and Rijmen, 2002). Therefore in SHAES we use 4 rounds, i.e. two full diffusion steps. As mentioned in the work by Daemen and Rijmen (2002), the wide trail strategy is used to determine the bounds offered by the AES to provide resistance against differential and linear cryptanalysis. The bounds calculated on 4 rounds are given by Daemen and Rijmen (2002) as follows.

1. A minimum weight of 150 for the differential trail and a maximum of 2^{-75} for correlation contribution in the linear trail. These bounds are true for all blocks of length and are independent of round keys.
2. The number of active S-boxes for differential and linear trails is lower bounded by 25.

Table 5.1: Consolidated cryptanalysis attack complexity on 4 round AES

| Work | Adversary type | Attack type | Data Complexity | Time Complexity |
|-------------------------------------------|-------------------------------------|---------------|-----------------|-----------------|
| Bouillaguet et al. (2012) | Highly Resource bounded in data | Diff and MiTm | 2 CP | 2^{80} |
| | | | 4 CP | 2^{32} |
| Bouillaguet et al. (2010) | Highly Resource bounded in data | Diff and MiTm | 5 CP | 2^{64} |
| | | | 10 CP | 2^{40} |
| | | | $2^{54.5}$ KP | 2^{64} |
| Tunstall (2011) | Highly Resource bounded in data | Differential | 12 CP | 2^{55} |
| | | | 30 CP | 2^{54} |
| | Moderately Resource bounded in data | Differential | 2^{11} CP | 2^{52} |
| | | | $2^{14.4}$ CP | 2^{51} |
| Biham and Keller (2000) | Moderately Resource bounded in data | Square | 2^9 | 2^8 |
| Daemen and Rijmen (2002) | Moderately Resource bounded in data | Square | 2^9 CP | 2^9 |

3. The Mix column of the fourth round does not have any effect on this bound and is not considered in proving this bound.

The consolidated cryptanalysis attack and its complexity presented in the literature of 4 round AES is presented in a tabular form in table 5.1.

From table 5.1, it is clear that under the data resource bounded adversary, the complexity of breaking the 4 round cipher is very high. However, for an adversary who is moderately resource bounded in data, the complexity is very low in a square attack. These results are based on the attacks performed on normal deterministic AES encryption either under the Known Plaintext (KP) or Chosen Plaintext (CP) attack model. Table 5.1 presents the security level achieved from normal 4-round deterministic AES.

5.6 SRSS and SHAES Combination

The combination of secret sharing schemes with encryption is one possible way to overcome the near-sink CN attack. The other advantage of this combination is reliability. With $n \geq t$ in equation 5.1, this method can provide the desired reliability. Properly grouping shares and then encrypting them can provide both reliability and security. Figure 5.7 shows the proper way of combining the SRSS scheme with the SHAES encryption, providing both reliability and security. Sensor nodes can select proper values for t_1, t_2 and n based on the application requirements and generate the shares. The shares are grouped based on the share numbers as

explained in figure 5.7. Only $n - t_1$ grouped shares are encrypted. Since an adversary having only t_1 unencrypted shares can't learn any information about the secret data, we encrypt the remaining shares (Stinson, 2006). The proposed optimized SRSS and SHAES (SRSS + SHAES) combination to provide both security and reliability is as shown in figure 5.7. Nodes transmit the encrypted shares to the BS. The combination scheme provides a resilience of $n - t_2$ share losses. Since only t_2 shares are required to reconstruct the transmitted data, the BS does the operations as shown in figure 5.8 to obtain the original data. If shares are encrypted then they are decrypted and finally, using interpolation, the data is recovered. All the operations are performed under $GF(2^8)$.

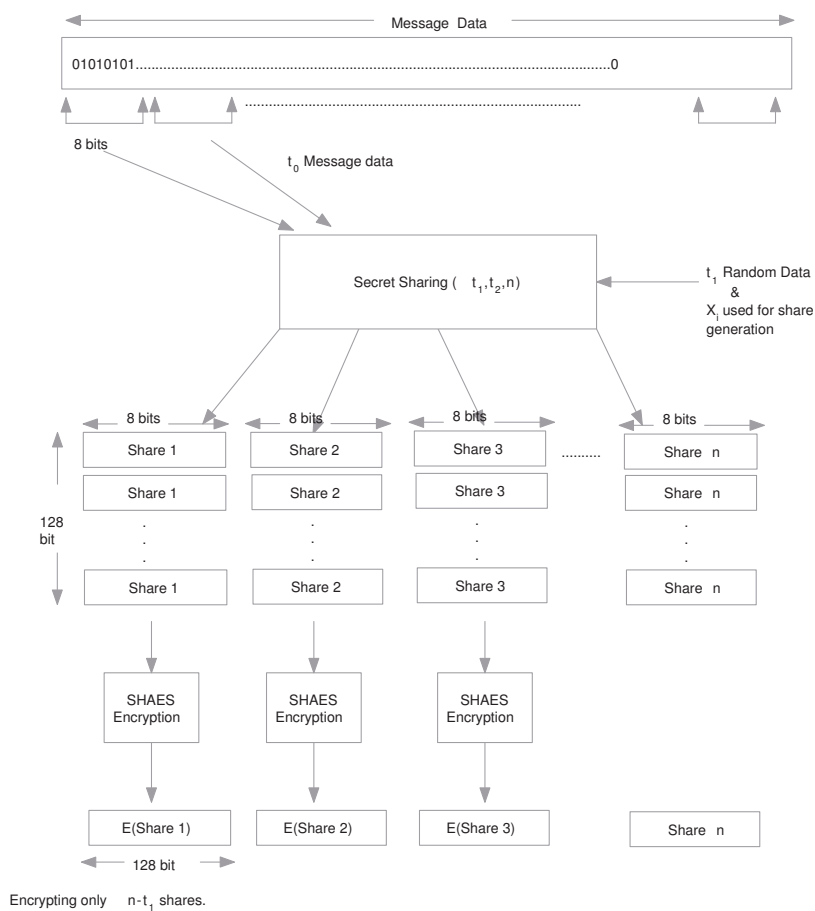


Figure 5.7: Graphical representation of achieving the Secret sharing and SHAES combination

5.6.1 SRSS+SHAES Security Analysis

The definition for the probabilistic public-key encryption is provided in the book by Stinson (2006) and if we extend the same definition to probabilistic symmetric encryption it would be

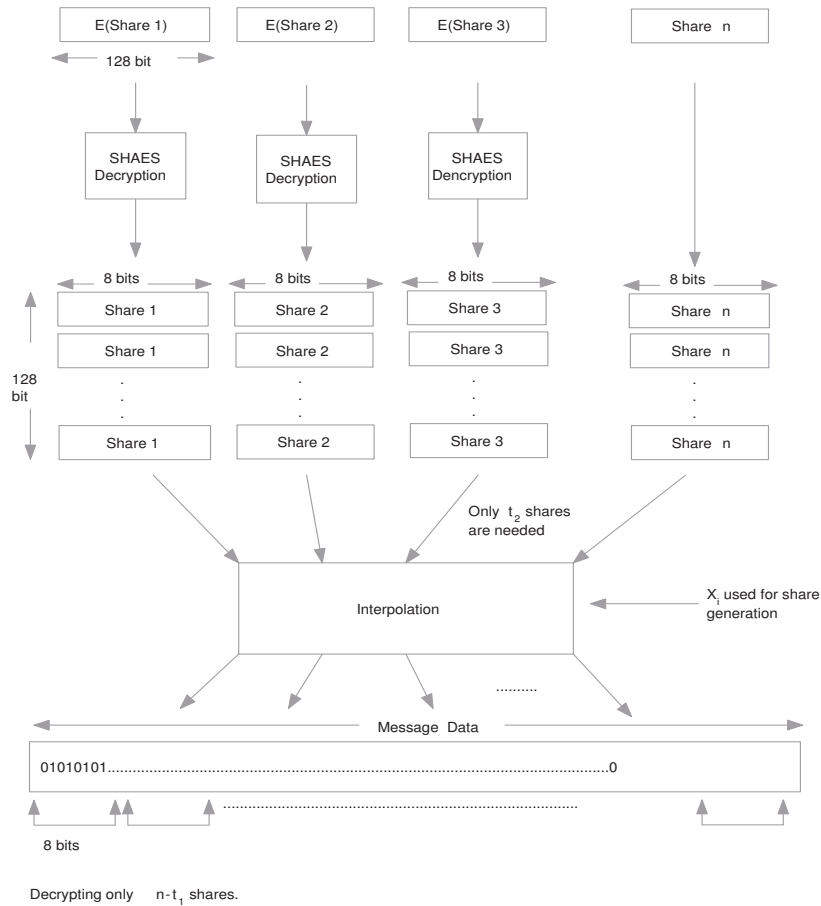


Figure 5.8: Graphical representation of retrieving the original message from the Secret sharing and SHAES combination at the BS

as follows.

A probabilistic Symmetric encryption can be defined as being six-tuple (P, C, K, E, D, R) where P is the Plaintext set space, C is the set of Ciphertext space, K represents the set of Key space, R is the set of Randomizer space, and for each key $k \in K$, $e_k \in E$ is the encryption rule and $d_k \in D$ is the decryption rule. The following properties should be satisfied:

1. For each $e_k : (P, R) \rightarrow C$ and $d_k : C \rightarrow P$ are functions such that $d_k(e_k(p, r)) = p$ for every plaintext $p \in P$ and $r \in R$

This implies that $e_k(p, r) \neq e_k(p_1, r)$ if $p \neq p_1$

2. For any fixed $k \in K$ and for any $p \in P$, define a probability distribution $f_{(k,p)}(y)$ on C where $f_{(k,p)}(y)$ denotes the probability that y is the ciphertext, given that k is the key and p is the plaintext (probability should be computed on all random choices of $r \in R$). Suppose $p, p_1 \in P$, $p \neq p_1$ and $k \in K$. The probability distributions $f_{(k,p)}$ and $f_{(k,p_1)}$ are not δ distinguishable in polynomial time. If δ is specified security parameter, then this is how the security of the scheme is defined.

Property 2 states that ciphertexts encrypting any two plaintexts should be indistinguishable in polynomial time. This is the desired feature for any security system and it provides strong semantic security or message indistinguishability (often used as interchangeable).

Any block cipher permutation function needs to be bijection (i.e. one-to-one and onto). AES is a block cipher and for any $k \in K$, $e_k()$ the encryption function is also bijection. Therefore property 1 is satisfied.

Proposition 1 : SRSS+SHAES is semantically secure

proof : Let M be the set of all possible messages.

Share generation is a function that maps messages to shares using random co-efficients from random space and for different values of x as given in equation 5.1.

Let $S(m_i)$ be the set of all possible shares generated using share generation function for m_i .

Let $A = \{s_1, s_2, \dots, s_n\} \subset S(m_i)$, $v = \{s_1, s_2, \dots, s_q\} \subseteq A$, $w = \{s_1, s_2, \dots, s_z\} \subseteq A$, where $z \leq t_1 < q \leq n$

SSS scheme is a perfectly secure sharing scheme and from the definition of perfect security as given by [Stinson \(2006\)](#). SRSS scheme with $(t_2 - 1, t_2, n)$ is same as the SSS scheme. Furthermore, SRSS with (t_1, t_2, n) does not reveal any information to an adversary having only t_1 shares. Therefore,

$$Pr(m = M | w = S(m_i)) = Pr(m = M) \tag{5.2}$$

This implies that the adversary having less than t_1 shares does not learn anything new from the message.

From Bayes' theorem we can compute

$$Pr(w = S(m_i) | m = M) = Pr(w = S(m_i)) \tag{5.3}$$

Since share generation is uniformly distributed, the shares in $S(m_i)$ are similar. $n - t_1$ shares are encrypted using computationally secure SHAES. Encrypted shares reveal no further information about the original message. Therefore, an adversary can't distinguish which cipher

text has resulted from which plain text. Thus, the scheme achieves semantic security and property 2 is satisfied.

The SRSS+SHAES combination achieves semantic security and comes under probabilistic symmetric encryption. Therefore, none of the cryptanalysis results of normal deterministic 4-round AES presented in Table 5.1, hold as it is for the proposed approach. However, cryptanalysis of the proposed probabilistic symmetric encryption that provides semantic security could be studied separately and this issue is not addressed in this thesis.

If the adversary using a CN attack or a near-sink CN attack overhears the transmitted message, then he can't learn any information from the t_1 unencrypted shares, and therefore can't reconstruct the original data. The adversary needs to successfully decrypt the encrypted shares to know the information about the data. Since the SRSS+SHAES combination provides semantic security, the adversary does not learn any new information from the encrypted shares and will therefore be unsuccessful in decrypting the shares. Thus the proposed optimized SRSS+SHAES combination overcomes the CN attack problem in WSNs.

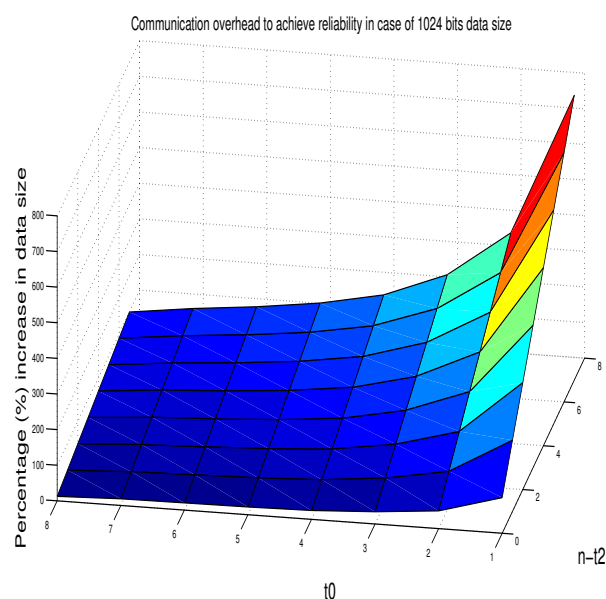


Figure 5.9: Reliability analysis of SRSS and SSS schemes

5.6.2 Energy Efficiency and Reliability Analysis of SRSS+SHAES

Energy efficiency is a crucial requirement for sensor networks. As multihop communication is often used to communicate data, communication energy efficiency becomes critical.

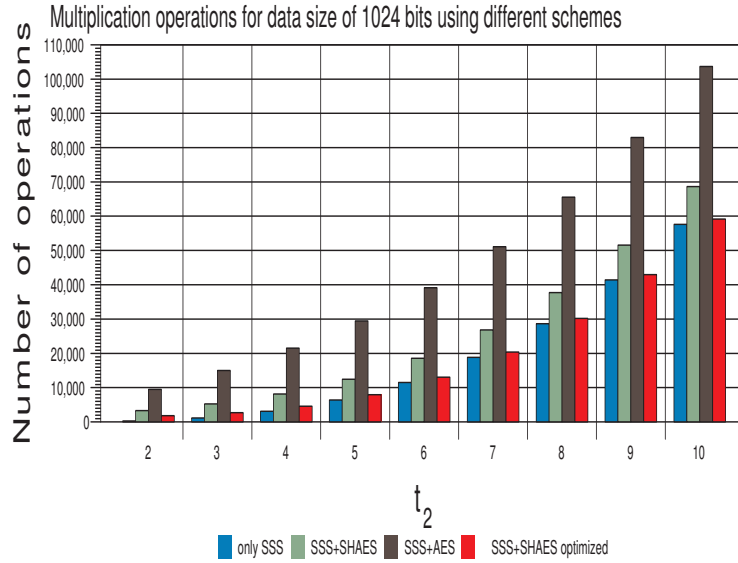


Figure 5.10: Computational overhead analysis in terms of number of multiplication operations for different approaches

Data redundancy is necessary to achieve reliability, therefore as the data size and redundancy increases, communication energy drain also increases. The requirement is to achieve reliability with a minimum increase in data size. Figure 5.9 shows the percentage increase in data size for a 1024 bits data using SRSS and SSS schemes in order to meet various reliability requirements indicated by $n - t_2$. For instance, in order to being able to afford to lose 7 shares (i.e. $n - t_2 = 7$), while the SRSS scheme with $t_0 = 8$ presents 100% percentage increase in data size, the SSS scheme with $t_0 = 1$ would have 800% increase in data size. Note that SRSS with $t_0 = 1$ is same as the SSS scheme. Further, reliability requirements depend on the wireless channel properties and therefore vary based on the channel conditions.

Computation energy depends on the complexity of the schemes, hardware implementation and also on the processor. One general way to analyze computation overhead is to analyze the complexity of the schemes by calculating the number of complex operations. Major complex operations involved in SRSS and SHAES are addition and multiplication operations under $GF(2^8)$. Among these multiplication operations are the most computationally expensive operations (Daemen and Rijmen, 2002). Therefore more importance is given to multiplication operations. There are different ways to realize the efficient hardware implementation of multiplication and addition operations in $GF(2^8)$ (Guajardo *et al.*, 2006). In the AES, the MC involves multiplication operations. Addition operations are involved in AK, MC and key schedule. The

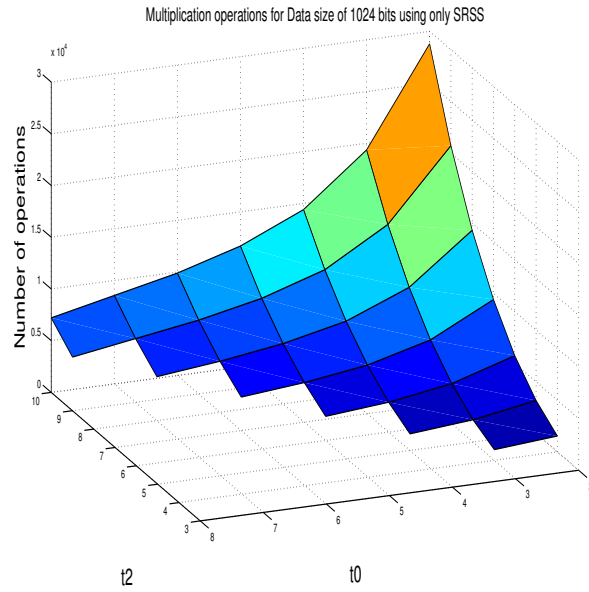


Figure 5.11: Computational overhead analysis in terms of number of multiplication operations using only SRSS

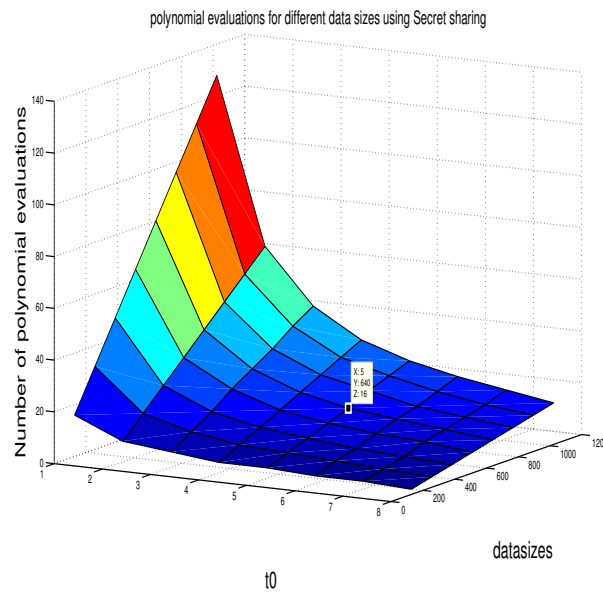


Figure 5.12: Polynomial evaluations for different data sizes using various SRSS parameters

number of addition and multiplication operations required to realize AES are calculated and are equal to 752 and 576 operations respectively. Referring to the AES, the number of addition and multiplication operations needed to realize SHAES are equal to 272 and 192 respectively.

The polynomial evaluation of secret sharing schemes also involves addition, multiplication and exponential operations. The exponential operations can also be realized with repeated

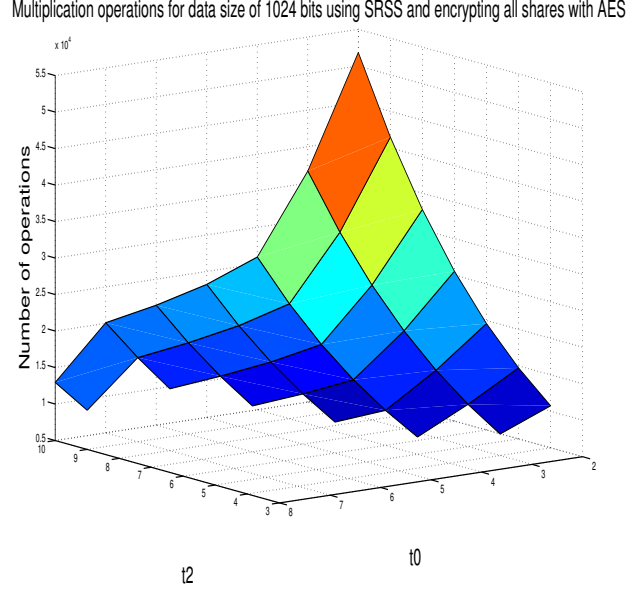


Figure 5.13: Computational overhead analysis in terms of number of multiplication operations using SRSS+AES

multiplication operations. The number of multiplication and addition operations needed to communicate the data size of \dot{D} bytes using SRSS are given by equations 5.4 and 5.5, respectively. Figures 5.10 and 5.11 show the multiplication operations needed to communicate the data size of 1024 bits using only the SSS and the SRSS schemes, respectively. For instances, at $SSS(t_2 = 5, n = 5)$ and $SSS(t_2 = 10, n = 10)$ the numbers of multiplication operations needed are 6400 and 57600, respectively. At $SRSS(t_0 = 4, t_2 = 5, n = 5)$ and $SRSS(t_0 = 8, t_2 = 10, n = 10)$ the numbers of multiplication operations needed are 1600 and 7200, respectively. As t_2 increases, the number of multiplication operations also increases, as indicated in figures 5.10 and 5.11. Therefore, to reduce the computation burden under secret sharing schemes, one needs to select parameters with lesser t_2 resulting in less multiplication operations.

$$\dot{D}/t_0 \times N \times \sum_{i=1}^{t_2-1} i \quad (5.4)$$

$$\dot{D}/t_0 \times N \times (t_2 - 1) \quad (5.5)$$

The proposed optimized SRSS and SHAES combination encrypts minimum group of shares (total size = 128 bits) as explained in figure 5.7. If the encrypting share group size is less than

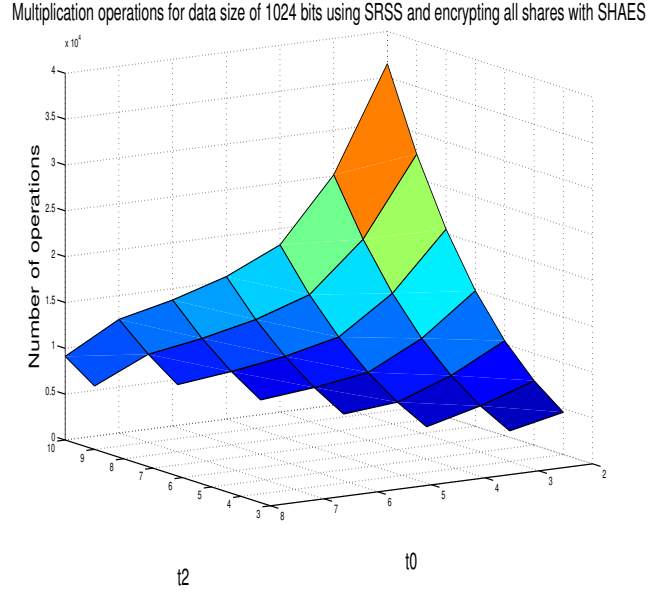


Figure 5.14: Computational overhead analysis in terms of number of multiplication operations using SRSS+SHAES

128 bits then extra bits must be added. As operations are performed in $GF(2^8)$, each share will be of 8 bits. A group of 16 shares would be equal to the required 128 bits. Therefore, proper selection of SRSS parameters based on the data size lengths helps to reduce the extra bits and achieves better communication energy efficiency. Figure 5.12 shows the number of polynomial evaluations for different data sizes using various SRSS parameters. The SRSS parameter combinations for different data sizes in the lower half of the demarcated line of figure 5.12 are not efficient as they require extra bits to make the share group size reach 128 bits. In the SRSS analysis, we have restricted up to $t_0 = 8$, the lowest data size that results in an efficient combination using $t_0 = 8$ is equal to 1024 bits. Therefore, the results of 1024 bit data size are presented in this Thesis.

Figure 5.10 shows the total number of multiplication operations required to realize the combination of SSS with the AES and the SHAES, encrypting each group of shares. The numbers of multiplication operations needed to realize the combination of SRSS with the AES and the SHAES, encrypting each group of shares are shown in figures 5.13 and 5.14, respectively. For instances, at SRSS[$(t_0 = 4, t_2 = 5, n = 5)$ and $(t_0 = 8, t_2 = 10, n = 10)$] the numbers of multiplication operations needed for SRSS+SHAES and SRSS+AES are [3520 and 9120] and [7360 and 12960], respectively. The combination of SSS + SHAES with minimum share group encryption has high computational overhead and communication overhead as indicated in figures

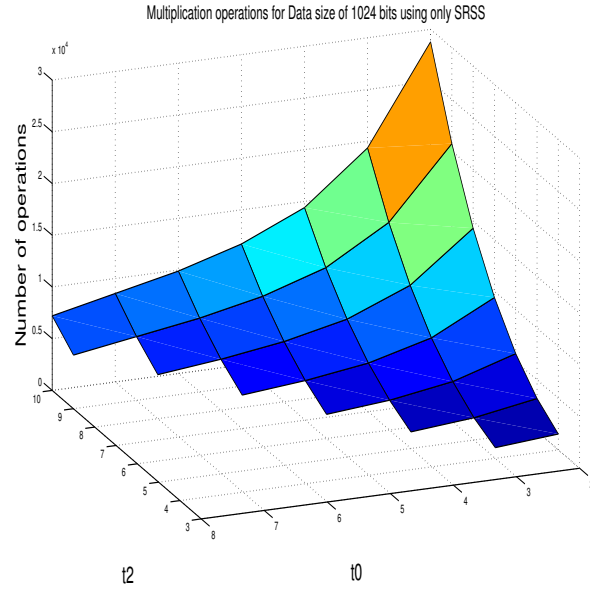


Figure 5.15: Computational overhead analysis in terms of number of multiplication operations using optimized SRSS+SHAES

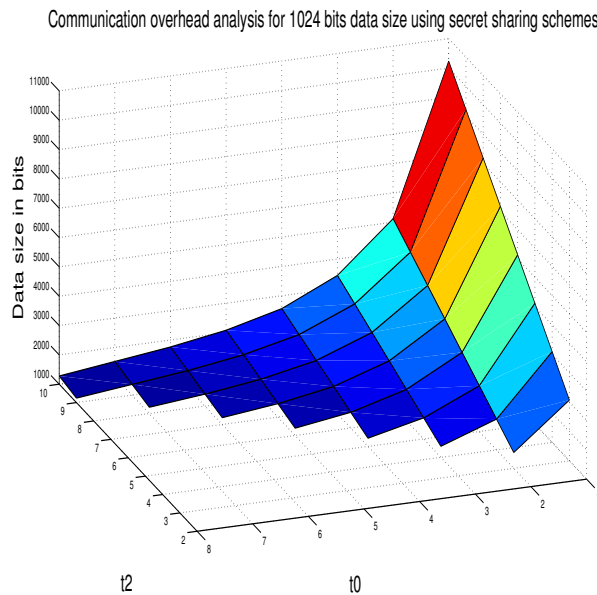


Figure 5.16: Communication overhead analysis in terms of data size expansion using SSS and SRSS schemes

5.10 and 5.16. The proposed optimized combination of SRSS+SHAES achieves low computational and communication overhead as explained in figures 5.15 and 5.16. For instances, at SRSS $[(t_0 = 4, t_2 = 5, n = 5)$ and $(t_0 = 8, t_2 = 10, n = 10)]$ the numbers of multiplication operations needed for optimized SRSS+SHAES is [3136 and 8736].

The SSS+AES combination is least energy efficient and the proposed optimized SRSS +

Table 5.2: Consolidated analysis of different approaches

| The Works | Objective | CN attack | Computation Overhead in terms of (x) operations | Communication Overhead in terms of data size | |
|-------------------------------------------------|----------------------------------------------------------------------------|-----------|-------------------------------------------------|----------------------------------------------|------|
| Previous Works | Lou and Kwon (2006b), Shu <i>et al.</i> (2010b), Liu <i>et al.</i> (2012b) | Only SSS | Yes | Medium | High |
| | Hsu <i>et al.</i> (2011b) | Only SRSS | Yes | Lowest | Low |
| Different approaches examined in proposed works | SSS+AES | No | Highest | High | |
| | SSS+SHAES | No | High | High | |
| | SRSS+AES | No | High | Low | |
| | SRSS+SHAES | No | Medium | Low | |
| | SSS+AES optimized | No | High | High | |
| | SRSS+SHAES optimized | No | Low | Low | |

SHAES with minimum share group encryption is highly energy efficient compared to other combinations. Computational overhead of different approaches are analyzed using number of multiplication operations. With lot of options in selecting the parameters of SSS and SRSS, it is difficult to exactly analyze and compare the different schemes. The different approaches are given different levels of computational overhead ranging from lowest to highest based on the overall trend observed with increase in t_2 and data size. The communication overhead of different approaches depends on the selection of secret sharing schemes. SSS has the high communication overhead and the SRSS has the low communication overhead. The consolidated analysis of different combination objectives is presented in tabular form in table 5.2. The proposed optimized SRSS+SHAES combination approach achieves energy efficiency and also overcomes the CN attack.

5.7 conclusions

In this chapter, the vulnerability of secret sharing schemes under the relaxation of a completely secure area around the BS to near-sink CN attacks was validated through simulation results. Simulations were carried out using MATLAB and Castalia, a discrete network simulator. The theoretical analysis validated the achieved energy efficiency and desired semantic security. The proposed combination works independently from the underlying routing schemes.

Therefore, it can be easily incorporated into existing related works in secure data collection of WSNs.

CHAPTER 6

CONCLUSIONS AND FUTURE SCOPE

6.1 Conclusions

The following research objectives have been achieved.

- A data aggregation scheme is proposed, which deals with data compression and reconstruction based on Compressed Sensing (CS), which uses correlation between and within nodes.

In this thesis, with incorporation of CS as a data aggregation scheme, the information contained in the signal is safely maintained through its projections which can be reconstructed later. Inclusion of CS for an energy efficient routing technique further enhances the lifetime of the network. To improve network lifetime, CS has been employed at level 1 (at the leaf nodes). The dimensionality reduction at the transmitter is done using a measurement matrix. At the receiver data is recovered using l_1 magic (convex relaxation) and greedy based methods. Greedy based method offers low complexity and low implementation cost. The success rate of greedy method depends on the sparsity of the data. Performance evaluation of greedy based methods is analyzed by considering varying sparsity, and plotted against number of measurement required to reconstruct the signals along with reconstruction errors. The same is analyzed by considering the real temperature and humidity data sets.

- Another data aggregation scheme is proposed exploring the intra and inter correlations, through the concept of Joint Sparse Models (JSM) to reduce the amount of redundant data into the WSN.

The inter-signal and intra-signal correlations are explored in DCS through the concept of joint sparse models. We conducted an analysis of joint sparse models and reconstruction of the

signal using joint recovery. Using synthetic signals which possess the inherent qualities of natural signals, we analyzed reconstruction performance using joint recovery (using S-OMP) and separate recovery (using OMP). Further we employed DCS on real data to evaluate the amount of data reduction, by comparing the same using separate recovery. Depending on the amount of intra and inter correlation DCS proves to be a better data aggregation technique. Simulation results show that even in less sparse environment, DCS performs better than separate recovery, which is well suited for real signals. Simulations also prove that with DCS, we can further reduce the number of measurements (compressed vector length), required for data reconstruction as compared to separate recovery. Simulation results show nearly 50% reduction in the data required for reconstruction in case of synthetic signals and 27% in case of real signals. We also considered a data set with EEG signals, with 8 channel DCS Compression. In this case identifying the correct sparsifying basis is very important. For EEG signals, we considered wavelet transform as the sparsifying basis.

- A new scheme that is energy efficient, reliable, and secure against CN attacks is proposed by combining Shamir's Ramp Secret Sharing (SRSS) and a round-reduced AES cipher, which we call split hop AES (SHAES).

In this thesis, an energy efficient SRSS and SH-AES is combined to provide both strong semantic security and reliability in an efficient way. A possible way of combining two schemes is proposed, and the combination works independently of the underlying routing schemes. Therefore, it can be easily accommodated with the existing related works in the area of secure data collection of WSNs. Extensive theoretical analysis of which also considers the vast available literature on crypt analysis validates the achieved energy efficiency and the desired strong security. The unique characteristic and practicality of a CN attack in sensor networks raises some serious threats on data security. Hence we addressed the CN attack by providing the strong semantic security in an energy-efficient way. As future extension of this work, the following ideas can be considered.

6.2 Future Scope

- While considering multi hop WSNs, CS can be applied at the leaf nodes which can help to reduce the network burden.
- Further by considering cross-layer design the performance can be improved.
- Estimation of measurement matrix can be done so as to get the optimum result of Data aggregation using CS/DCS.
- The objective is to relax the assumption of secure area around the BS and to analyze the vulnerability of secret sharing schemes under CN attack. Proposed a method that combats CN attack in an energy efficient way.
- Optimized selection on the ramp secret sharing parameters based on the sensor network routing constraints.
- Further crypt analysis of the proposed probabilistic symmetric encryption scheme.

REFERENCES

- Akkaya, K.** and **M. Younis** (2005). A survey on routing protocols for wireless sensor networks. *Ad hoc networks*, **3**(3), 325–349.
- Alippi, C., S. Ntalampiras,** and **M. Roveri**, Model ensemble for an effective on-line reconstruction of missing data in sensor networks. *In The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013.
- Baraniuk, R. G.** (2007). Compressive sensing. *IEEE signal processing magazine*, **24**(4).
- Baron, D., M. F. Duarte, S. Sarvotham, M. B. Wakin,** and **R. G. Baraniuk**, An information-theoretic approach to distributed compressed sensing. *In Proc. 45rd Conference on Communication, Control, and Computing*. 2005.
- Baron, D., M. F. Duarte, M. B. Wakin, S. Sarvotham,** and **R. G. Baraniuk** (2009). Distributed compressive sensing. *arXiv preprint arXiv:0901.3403*.
- Biham, E.** and **N. Keller**, Cryptanalysis of reduced variants of rijndael. *In 3rd AES Conference*, volume 230. 2000.
- Bouillaguet, C., P. Derbez, O. Dunkelman, P. Fouque, N. Keller,** and **V. Rijmen** (2012). Low-data complexity attacks on aes. *Information Theory, IEEE Transactions on*, **58**(11), 7002–7017. ISSN 0018-9448.
- Bouillaguet, C., P. Derbez, O. Dunkelman, N. Keller,** and **P.-A. Fouque** (2010). Low data complexity attacks on aes.
- Caione, C., D. Brunelli,** and **L. Benini** (2013). Compressive sensing optimization for signal ensembles in wsns. *IEEE Transactions on Industrial Informatics*, **10**(1), 382–392.
- Caione, C., D. Brunelli,** and **L. Benini** (2014). Compressive sensing optimization for signal ensembles in wsns. *Industrial Informatics, IEEE Transactions on*, **10**(1), 382–392. ISSN 1551-3203.

Candès, E. J., J. Romberg, and T. Tao (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, **52**(2), 489–509.

Candes, E. J., J. K. Romberg, and T. Tao (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, **59**(8), 1207–1223.

Cao, G., P. Jung, S. lawomir Stanczak, and F. Yu (2008). Data aggregation and recovery in wireless sensor networks using compressed sensing.

Challal, Y., A. Ouadjaout, N. Lasla, M. Bagaa, and A. Hadjidj (2011). Secure and efficient disjoint multipath construction for fault tolerant routing in wireless sensor networks. *Journal of network and computer applications*, **34**(4), 1380–1397.

Claveirole, T., M. D. De Amorim, M. Abdalla, and Y. Viniotis (2008). Securing wireless sensor networks against aggregator compromises. *IEEE Communications Magazine*, **46**(4), 134–141.

Daemen, J. and V. Rijmen, *The Design of Rijndael*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002. ISBN 3540425802.

Dai, W. and O. Milenkovic (2009). Subspace pursuit for compressive sensing signal reconstruction. *IEEE transactions on Information Theory*, **55**(5), 2230–2249.

Deng, J., R. Han, and S. Mishra (2006). Insens: Intrusion-tolerant routing for wireless sensor networks. *Computer Communications*, **29**(2), 216–230.

Djelouat, H., H. Baali, A. Amira, and F. Bensaali (2017). An adaptive joint sparsity recovery for compressive sensing based eeg system. *Wireless Communications and Mobile Computing*, **2017**.

Dong, M., K. Ota, and A. Liu (2016). Rmer: Reliable and energy-efficient data collection for large-scale wireless sensor networks. *IEEE Internet of Things Journal*, **3**(4), 511–519.

Donoho, D. L. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, **52**(4), 1289–1306. ISSN 0018-9448.

- Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck** (2012). Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, **58**(2), 1094–1121.
- Doomun, M. R. and K. M. S. Soyjaudah** (2009). Analytical comparison of cryptographic techniques for resource-constrained wireless security. *I. J. Network Security*, **9**(1), 82–94.
- Duarte, M. F., M. B. Wakin, D. Baron, and R. G. Baraniuk**, Universal distributed sensing via random projections. *In Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006.
- Fragkiadakis, A., I. Askoxylakis, and E. Tragos**, Joint compressed-sensing and matrix-completion for efficient data collection in wsns. *In Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2013 IEEE 18th International Workshop on*. IEEE, 2013.
- Gu, L., Y. Pan, M. Dong, and K. Ota** (2013). Noncommutative lightweight signcryption for wireless sensor networks. *International Journal of Distributed Sensor Networks*, **9**(3), 818917.
- Guajardo, J., T. Güneysu, S. S. Kumar, C. Paar, and J. Pelzl** (2006). Efficient hardware implementation of finite fields with applications to cryptography. *Acta Applicandae Mathematica*, **93**(1-3), 75–118.
- Haenggi, M. and D. Puccinelli** (2005). Routing in ad hoc networks: a case for long hops. *IEEE Communications Magazine*, **43**(10), 93–101.
- Heinzelman, W. B., A. P. Chandrakasan, H. Balakrishnan, et al.** (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications*, **1**(4), 660–670.
- Hormati, A. and M. Vetterli**, Distributed compressed sensing: Sparsity models and reconstruction algorithms using annihilating filter. *In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008.
- Hsu, C.-F., G.-H. Cui, Q. Cheng, and J. Chen** (2011a). A novel linear multi-secret sharing scheme for group communication in wireless mesh networks. *Journal of Network and Computer Applications*, **34**(2), 464–468.

Hsu, C.-F., G.-H. Cui, Q. Cheng, and J. Chen (2011b). A novel linear multi-secret sharing scheme for group communication in wireless mesh networks. *Journal of Network and Computer Applications*, **34**(2), 464 – 468. ISSN 1084-8045.

Hu, F. and Q. Hao, ntelligent sensor networks: The integration of sensor networks, signal processing and machine learning. In **Q. H. Fei Hu** (ed.), *The Oxford Handbook of Innovation*, chapter 10. CRC Press, Oxford, 2012, 266–290.

Ingelrest, F., G. Barrenetxea, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange (2010). Sensorscope: Application-specific sensor network for environmental monitoring. *ACM Transactions on Sensor Networks (TOSN)*, **6**(2), 17.

Jung, W.-S., K.-W. Lim, Y.-B. Ko, and S.-J. Park (2011). Efficient clustering-based data aggregation techniques for wireless sensor networks. *Wireless Networks*, **17**(5), 1387–1400. ISSN 1572-8196. URL <http://dx.doi.org/10.1007/s11276-011-0355-6>.

Karlof, C. and D. Wagner, Secure routing in wireless sensor networks: attacks and countermeasures. In *Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on.* 2003.

Kim, S., H. Joh, S. Choi, and I. Ryo (2015). Energy efficient mac scheme for wireless sensor networks with high-dimensional data aggregate. *Mathematical Problems in Engineering*, **2015**, Pages 13, Hindawi Publishing Corporation.

Koushanfar, F., N. Taft, and M. Potkonjak (2006). Sleeping coordination for comprehensive sensing using isotonic regression and domatic partitions.

Liu, A., L. T. Yang, M. Sakai, and M. Dong (2013). Secure and energy-efficient data collection in wireless sensor networks.

Liu, A., Z. Zheng, C. Zhang, Z. Chen, and X. Shen (2012a). Secure and energy-efficient disjoint multipath routing for wsns. *Vehicular Technology, IEEE Transactions on*, **61**(7), 3255–3265. ISSN 0018-9545.

Liu, A., Z. Zheng, C. Zhang, Z. Chen, and X. Shen (2012b). Secure and energy-efficient disjoint multipath routing for wsns. *Vehicular Technology, IEEE Transactions on*, **61**(7), 3255–3265. ISSN 0018-9545.

- Liu, J., K. Huang, and X. Yao** (2018). Common-innovation subspace pursuit for distributed compressed sensing in wireless sensor networks. *IEEE Sensors Journal*, **19**(3), 1091–1103.
- Liu, Y., W. Zhao, L. Zhu, B. Ci, and S. Chen**, *Advances in Wireless Sensor Networks: The 8th China Conference, CWSN 2014, Xi'an, China, October 31–November 2, 2014. Revised Selected Papers*, chapter The Method of Data Aggregation for Wireless Sensor Networks Based on LEACH-CS. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-46981-1, 489–498. URL http://dx.doi.org/10.1007/978-3-662-46981-1_47.
- Liu, Z., Ren P.and Rosberg, C. Collings, Iain B.and Wilson, A. Y. Don, and S. Jha** (2009). Energy efficient reliable data collection in wireless sensor networks with asymmetric links. *International Journal of Wireless Information Networks*, **16**(3), 131. ISSN 1572-8129. URL <https://doi.org/10.1007/s10776-009-0103-3>.
- Lou, W. and Y. Kwon** (2006a). H-spread: a hybrid multipath scheme for secure and reliable data collection in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, **55**(4), 1320–1330.
- Lou, W. and Y. Kwon** (2006b). H-spread: a hybrid multipath scheme for secure and reliable data collection in wireless sensor networks. *Vehicular Technology, IEEE Transactions on*, **55**(4), 1320–1330. ISSN 0018-9545.
- Madden, S., M. J. Franklin, J. M. Hellerstein, and W. Hong** (2002). Tag: A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, **36**(SI), 131–146.
- Marcelloni, F. and M. Vecchio** (2008). A simple algorithm for data compression in wireless sensor networks. *Communications Letters, IEEE*, **12**(6), 411–413. ISSN 1089-7798.
- Needell, D. and J. A. Tropp** (2009). Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and computational harmonic analysis*, **26**(3), 301–321.
- Needell, D. and R. Vershynin** (2009). Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of computational mathematics*, **9**(3), 317–334.
- Nishant, J. D., K. Shivaprakasha, and M. Kulkarni**, Multi threshold adaptive range clustering (m-trac) algorithm for energy balancing in wireless sensor networks. *In Wireless and*

Optical Communications Networks (WOCN), 2012 Ninth International Conference on. IEEE, 2012.

Pan, Y., Y. Yu, and L. Yan (2013). An improved trust model based on interactive ant algorithms and its applications in wireless sensor networks. *International Journal of Distributed Sensor Networks*, **9**(7), 764064.

Perrig, A., R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler (2002). Spins: Security protocols for sensor networks. *Wireless networks*, **8**(5), 521–534.

Qin, S. and J. Yin, *Advances in Wireless Sensor Networks: The 8th China Conference, CWSN 2014, Xi'an, China, October 31–November 2, 2014. Revised Selected Papers*, chapter A Robust Sparsity Estimation Method in Compressed Sensing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-46981-1, 481–488. URL http://dx.doi.org/10.1007/978-3-662-46981-1_46.

Quer, G., R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi (2012). Sensing, compression, and recovery for wsns: Sparse signal modeling and monitoring framework. *Wireless Communications, IEEE Transactions on*, **11**(10), 3447–3461. ISSN 1536-3461.

Rossi, M., M. Hooshmand, D. Zordan, and M. Zorzi, Evaluating the gap between compressive sensing and distributed source coding in wsn. *In Computing, Networking and Communications (ICNC), 2015 International Conference on.* IEEE, 2015.

Sarvotham, S., D. Baron, M. Wakin, M. F. Duarte, and R. G. Baraniuk, Distributed compressed sensing of jointly sparse signals. *In Asilomar conference on signals, systems, and computers.* 2005.

Shaheen, J., D. Ostry, V. Sivaraman, and S. Jha, Confidential and secure broadcast in wireless sensor networks. *In 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications.* IEEE, 2007.

Shi, E. and A. Perrig (2004). Designing secure sensor networks. *IEEE Wireless Communications*, **11**(6), 38–43.

Shivaprakasha, K., M. Kulkarni, and N. Joshi (2013). Improved network survivability using multi-threshold adaptive range clustering (m-trac) algorithm for energy balancing in wireless sensor networks. *Journal of High Speed Networks*, **19**(2), 99–113.

- Shu, T., M. Krunz, and S. Liu (2010a).** Secure data collection in wireless sensor networks using randomized dispersive routes. *IEEE transactions on mobile computing*, **9(7)**, 941–954.
- Shu, T., M. Krunz, and S. Liu (2010b).** Secure data collection in wireless sensor networks using randomized dispersive routes. *Mobile Computing, IEEE Transactions on*, **9(7)**, 941–954. ISSN 1536-1233.
- Slepian, D. and J. Wolf (1973).** Noiseless coding of correlated information sources. *IEEE Transactions on information Theory*, **19(4)**, 471–480.
- Song, H., G. Wang, and Y. Zhan,** Sparse target localization in rf sensor networks using compressed sensing. *In 2013 25th Chinese Control and Decision Conference (CCDC)*. IEEE, 2013.
- Stinson, D.,** *Cryptography: Theory And Practice*. The CRC Press Series on Discrete Mathematics and its Applications. Chapman & Hall/CRC, 2006. ISBN 9781584885085. URL <http://books.google.ca/books?id=Yz551PEuzckC>.
- Stinson, D. R.,** *Cryptography: theory and practice*. CRC press, 2005.
- Sundman, D., S. Chatterjee, and M. Skoglund,** Greedy pursuits for compressed sensing of jointly sparse signals. *In 2011 19th European Signal Processing Conference*. IEEE, 2011.
- Tropp, J. A. and A. C. Gilbert (2007).** Signal recovery from random measurements via orthogonal matching pursuits. *IEEE Transactions on information theory*, **53(12)**, 4655–4666.
- Tropp, J. A., A. C. Gilbert, and M. J. Strauss,** Simultaneous sparse approximation via greedy pursuit. *In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 5. IEEE, 2005.
- Tunstall, M. (2011).** Practical complexity differential cryptanalysis and fault analysis of aes. Cryptology ePrint Archive, Report 2011/453. <http://eprint.iacr.org/>.
- Wakin, M. B., M. F. Duarte, S. Sarvotham, D. Baron, and R. G. Baraniuk,** Recovery of jointly sparse signals from few random projections. *In Advances in Neural Information Processing Systems*. 2006.
- Wang, X., Z. Zhao, Y. Xia, and H. Zhang,** Compressed sensing based random routing for multi-hop wireless sensor networks. *In Communications and Information Technologies (ISCIT), 2010 International Symposium on*. IEEE, 2010.

Wimalajeewa, T. and **P. K. Varshney**, Robust detection of random events with spatially correlated data in wireless sensor networks via distributed compressive sensing. *In 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2017.

Xiang, L., J. Luo, and **C. Rosenberg** (2013). Compressed data aggregation: Energy-efficient and high-fidelity data collection. *Networking, IEEE/ACM Transactions on*, **21**(6), 1722–1735.

Xing, X., D. Xie, and **G. Wang** (2015). Energy-balanced data gathering and aggregating in wsns: A compressed sensing scheme. *International Journal of Distributed Sensor Networks*, **2015**, Article ID 585191, 10 pages, doi:10.1155/2015/585191.

Xu, X., R. Ansari, and **A. Khokhar**, Power-efficient hierarchical data aggregation using compressive sensing in wsns. *In Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013.

Zhang, Y. (2009). User's guide for yall1: Your algorithms for l1 optimization. Technical report.

Zheng, H., F. Yang, X. Tian, X. Gan, X. Wang, and **S. Xiao** (2015). Data gathering with compressive sensing in wireless sensor networks: a random walk based approach. *Parallel and Distributed Systems, IEEE Transactions on*, **26**(1), 35–44. ISSN 1045-9219.

Zhou, J. (2013). Efficient and secure routing protocol based on encryption and authentication for wireless sensor networks. *International Journal of Distributed Sensor Networks*, **9**(4), 108968.

Zhu, Y.-h., Y.-y. Wang, and **L. Chi, Kai-kai and**, *Advances in Wireless Sensor Networks: The 8th China Conference, CWSN 2014, Xi'an, China, October 31–November 2, 2014. Revised Selected Papers*, chapter An Energy-Efficient Data Gathering Scheme for Unreliable Wireless Sensor Networks Using Compressed Sensing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-46981-1, 444–455. URL http://dx.doi.org/10.1007/978-3-662-46981-1_43.

Publications based on the thesis

Refereed International Journals/Conferences

1. Deepa Puneeth and Muralidhar Kulkarni, Book Chapter, “Data Aggregation Using Distributed Compressive Sensing in WSNs”, **SPRINGER NATURE Singapore Pte Ltd 2020, S. M. Thampi et. al.(Eds.): SIR’S 2019, Communications in Computer and Information Science (CCIS), 1209, pp:276-290, 2020. <https://doi.org/10.1007/978-981-15-4828-4-23>. (Scopus)**
2. Deepa Puneeth, Muralidhar Kulkarni, “Data Aggregation Using Compressive Sensing for Energy Efficient Routing Strategy”, **Elsevier Procedia Computer Science Journal, Vol: 171: pp: 2242-2251, 2020. (Scopus)**
3. D. Puneeth, N. Joshi, Pradeep Kumar Atrey and Muralidhar Kulkarni, “Energy efficient and reliable data collection in wireless sensor networks”, **Turkish Journal of Electrical Engineering and Computer Science, Vol: 26: pp: 138 - 149, 2018. (SCI-E)**
4. Deepa Puneeth, R Ruthwik, Muralidhar Kulkarni, “Data Aggregation using Compressive Sensing for Improved Network Lifetime in Large Scale Wireless Sensor Networks”, **International Journal of Control Theory and Applications, Vol:No.9, Issue:17, pp: 8651-8657, 2016. (Scopus)**
5. Deepa Puneeth and Muralidhar Kulkarni, “Data Aggregation using Distributive Compressive Sensing in WSNs”, **Proceedings of the Fifth International Symposium on**

Signal Processing and Intelligent Recognition Systems (SIRS'19), Trivandrum, Kerala, India, Dec.18-21, 2019.

6. Deepa Puneeth and Muralidhar Kulkarni, "Data Aggregation using Compressive Sensing for Energy Efficient Routing Strategy", , **Proceedings of the Third International Conference on Computing and Network Communications (CoCoNet'19), Trivandrum, Kerala, India, Dec.18-21, 2019.**

7. Deepa Puneeth, R Ruthwik, Muralidhar Kulkarni, "Data Aggregation using Compressive Sensing for Improved Network Lifetime in Large Scale Wireless Sensor Networks", **International Conference on Sustainable Computing Techniques in Engineering, Science and Management (SCESM 2016), September 9-10, 2016.**

BIO-DATA

NAME : DEEPA PUNEETH
DATE OF BIRTH : 21st December, 1983
CONTACT ADDRESS : W/O Puneeth,
Manjunivas, Mundalachil,
Kulai, Mangalore, Dakshina Kannada - 575019
Karnataka, India
CONTACT NUMBER : +91-9611496321
EMAIL ID : deepapuneethk@gmail.com

EDUCATION QUALIFICATIONS :

Bachelor of Engineering (B.E)

Institution : P.A.College of Engineering, Mangalore
University : Visvesvaraya Technological University, Belgaum
Discipline : Telecommunication Engineering
Year of Passing : 2005

Master of Technology (M.Tech)

Institution : NMAMIT, Nitte
University : Autonomous Institute, Nitte
Discipline : Digital Electronics & Communication Engineering
Year of Passing : 2009