To the Graduate Council:

I am submitting herewith a thesis written by Olawoye A. Oyeyele entitled "A Robust Node Selection Strategy for Lifetime Extension in Wireless Sensor Networks". I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Hairong Qi

Major Professor

We have read this thesis
and recommend its acceptance:

Daniel B. Koch

Mongi A. Abidi

Accepted for the Council:

Anne Mayhew

Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

A Robust Node Selection Strategy for Lifetime Extension in Wireless Sensor Networks

A Thesis
Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

Olawoye A. Oyeyele

August 2004

Acknowledgments

I am deeply grateful to my advisor, Professor Hairong Qi, for her excellent guidance, advice and endless support during the course of my study and research at The University of Tennessee, Knoxville. I also wish to thank Professor Mongi A. Abidi and Professor Daniel B. Koch for serving on my thesis committee.

I am indebted to my parents (Moses and Deborah Oyeyele) and siblings (Tade, Soji, Ronke, Tola and Diran) for the understanding, support and encouragement they gave me to explore higher levels of education. Thank you.

My ultimate praise goes to the Almighty God, the creator of heaven and earth, in whom all things consist. Thank you for loving me first and for your continual grace over me.

Abstract

Distributed Wireless Sensor Networks (WSNs) consist of energy-constrained sensor nodes that may be deployed in large numbers in order to monitor a given area, track military targets, detect civilian targets or for other purposes. In such densely deployed environments, multiple transmissions can lead to collisions resulting in packet losses and network congestion. This can increase latency and reduce energy efficiency. These networks also feature significant redundancy since nodes close to each other often sense similar data. Therefore, it may be adequate to employ only a subset of the deployed nodes at any given time in the network. In this thesis, node subsets are selected in a manner that coverage and connectivity are consistently achieved. The working subsets are changed after predetermined durations.

A framework using concepts from spatial statistics is developed as an approach to selecting the subset of sensor nodes. Proximal nodes negotiate with each other using energy information, to decide which nodes stay working while others go to sleep mode. The algorithm is executed autonomously by the network. The approach presented ensures that the selected subsets while not necessarily exclusive of previous selections covers the region of interest. Simulation results show that the algorithm is robust and retains some level of redundancy. The algorithm shows significant improvement in energy consumption compared with a network with no selection. The selected subset is shown to be able to withstand significant levels of fault in the network. Conclusions regarding the flexibility and application scenarios of the algorithm are drawn and opportunities for future work indicated.

Contents

1 Introduction **1**

1.1 Sensor Node Architectures 5

1.2 Wireless Sensor Networks 6

1.3 Applications of Wireless Sensor Networks 10

1.4 Energy Consumption in Wireless Sensor Networks 13

1.5 Self Organization in Wireless Sensor Networks 14

1.6 Contribution of Research 14

1.7 Structure of the Thesis 15

2 Theoretical Background **16**

2.1 Spatial Statistics 16

2.1.1 Geostatistical Analysis 19

2.1.2 Intrinsic Stationarity 20

2.1.3 Estimating Spatial Dependence 21

2.1.4 Stationary Processes 21

2.1.5 Variogram 22

2.2 Signal Detection and Classification Theory 28

2.2.1 Detector Performance 31
2.3 Network Protocol Design 34

3 State of the Art: Efficient Approaches in Sensor Networks 39

3.1 Energy-Efficient Techniques 40
3.1.1 Clustering 40
3.1.2 Power-Saving Modes of Operation 41
3.1.3 Power Control 42
3.2 Bandwidth-Efficient Techniques 42
3.2.1 Mobile Agent Computing Paradigm 43
3.2.2 Clustering 44
3.2.3 Resource Selection 44
3.3 Coverage and Connectivity 45
3.3.1 Coverage 45
3.3.2 Connectivity 46
3.4 Related Work 46

4 Spatial Selection 49

4.1 Development of the Spatial Selection Algorithm 50
4.1.1 Determination of Correlation Distance 55
4.2 Details of the Selection Algorithm 58
4.3 Implementation Approach 61
4.3.1 Pre-deployment Phase 62
4.3.2 Post-deployment Phase 62

5 Performance Evaluation **63**

5.1 Simulation Parameters 63

5.2 Simulation Experiments 65

5.2.1 Effect of the variation in Range parameter 65

5.2.2 Effect of variation in Deployment Density 67

5.2.3 Effect of varying t_{sleep} 68

5.2.4 Selection Energy Consumption with $r_{negotiate}$ 68

5.2.5 Selection Energy Consumption with Network Density . . . 69

5.2.6 Sensing Redundancy 71

5.2.7 Uniformity of Coverage 73

5.2.8 Fault Tolerance 76

5.3 Detection Performance 79

5.4 Simulation Notes 85

6 Conclusions and Future Work **86**

Bibliography **89**

Appendix **98**

A **99**

A.1 Coverage Maps - No Faulty Nodes 99

A.2 Coverage Maps - With Faulty Nodes 102

Vita **104**

List of Tables

2.1 Theoretical Variogram Models [40] 28

2.2 OSI layers and typical Network Protocols 36

2.3 OSI layers and typical Sensor Network Protocols 38

3.1 Power Saving Modes of Operation [53] 41

3.2 Power consumed in different sleep states [53] 42

5.1 Simulation Parameters used for performance evaluation 65

5.2 Total Energy Consumption for one round of selection with $r_{negotiate}$ (200 nodes) 70

5.3 Energy Consumption for one round of selection with different Node Densities 70

5.4 Redundancy parameters for sensing Coverage 73

5.5 Parameters for uniformity of Coverage 74

5.6 Number of selected nodes with Node Failures ($r_{negotiate} = 3m$) . . 77

5.7 Redundancy parameters for sensing Coverage ($r_{negotiate} = 3m$) . . 78

5.8 Spatial parameters for uniformity of Coverage 79

List of Figures

1.1 OSI Network Protocol Stack 4

1.2 Sensor Network Protocol Stack [8] 4

1.3 Sensor Node Architecture 6

1.4 Sensor Network Architecture 7

2.1 Generalized Shape of the Variogram 24

2.2 Examples of Theoretical Variogram models 26

2.3 Examples of Theoretical Variogram models (contd.) 27

2.4 PDFs under a binary hypothesis test showing the threshold 32

2.5 ROC for the detection of a constant dc signal 34

4.1 Random deployment of sensor nodes in a dense network 52

4.2 Systematic sampling of a sensor field 54

4.3 Systematic sampling with discs 54

4.4 Semi-variogram for independent Gaussian noise 56

4.5 Power-law semi-variogram for model in Eq. 4.1. Produced using GSTAT [4] 58

4.6 State Machine for Selection Algorithm 59

4.7 Spatial selection algorithm 60

5.1 Simplified First Order Radio Model [24] 65

5.2 Node Selection using $r_{negotiate}$=3m 66

5.3 Selection using $r_{negotiate}$=5m 66

5.4 Plot of number of nodes selected for different $r_{negotiate}$ values. . . . 67

5.5 Energy consumption for different node densities using $r_{negotiate}$=3m and t_{sleep}= 5 seconds. 68

5.6 Effect of varying the sleep time $r_{negotiate}$= 3m. 69

5.7 Cluster of spatial uniformity values for different $r_{negotiate}$ 75

5.8 Positions of 20 randomly failed nodes 76

5.9 Positions of selected sensors when 10% nodes failed 77

5.10 ROC curves showing detection performance for different values of N and $\sigma_s^2/\sigma_n^2 = 1$ 84

Chapter 1

Introduction

The development of low-cost, low-power, multifunctional sensor nodes has been made possible by recent advances in micro-electro-mechanical systems (MEMS) technology, wireless communications and digital electronics. These sensor nodes feature increasingly smaller form factors and are typically capable of sensing, processing and communicating data. Due to the small size of these sensor nodes a spectrum of applications have become possible. In most of these applications, a large number of the nodes are deployed in order to monitor an area, detect a phenomenon or announce the presence of a target. However, sensor networks represent a departure from the conventional networking paradigm. Recent research has focused on conquering the challenges posed by sensor networks. This multidisciplinary research area promises to produce the next generation of intelligent networks with capabilities that transcend the feat achieved by the Internet only a decade ago. Many of the applications that have been envisioned will leverage the connectivity of the Internet to interact with a remote network of sensors while

carrying out a specific task. General uses include the detection, classification and tracking of targets or phenomena in the environment [33],[15],[16], [38],[59]. In order to fully take advantage of the ubiquitous nature of sensor networking most applications will require deployment of multiple sensors in order to achieve spatially dense sampling of the environment or targeted phenomena. The deployment of sensors in large numbers offers the following advantages [19],

- **Distributed Sensing:** Usually the exact location of a target is unknown in a region of interest. The deployment of multiple sensors ensures that a set of sensors is guaranteed to be closer to the phenomenon than when a single sensor is used. This results in higher signal to noise ratio (SNR) and an improved opportunity for line of sight. Hence robust observation of the environment is possible with multiple sensors.

- **Distributed Processing:** Sensor nodes run on battery power (dc energy) which is non-renewable. Thus, there is a limited energy capacity for the sensor nodes to operate with. Communications has been identified as the major consumer of power in sensor networks. If nodes must transmit data to a centralized location for processing there may be a need to partially process the captured data at the nodes before transmitting to the central location. Such processing will in general, reduce the volume of data to be transmitted.

- **Redundancy:** As noted above, many applications require that large numbers of sensor nodes be deployed in order to perform a given task, e.g. monitor a given area, detect a target, etc. Such applications however, may

not permit careful placement of the sensor nodes, especially in hazardous environments as encountered in millitary applications [8]. In such situations, the number of nodes may be large and it may be difficult to control the density of the deployed nodes. Therefore, there is generally, a measure of redundancy in the network. This redundancy is desirable for robust sensing of the phenomenon.

In comparison with traditional networks, sensor networks represent a new paradigm. Conventional networks use a layered approach to protocol design which aims to provide modular interaction between the different layers of the protocol stack. In addition, traditional networks are designed to perform generic networking tasks. This model is not attractive in the context of sensor networks due to different challenges faced in sensor networks. The International Systems Organization (ISO) Open System Interconnection (OSI)network protocol stack is as shown in Figure 1.1, a generalized sensor network protocol stack is shown in Figure 1.2.

Due to the peculiar characteristics, operating environment in which sensor networks are used and the possiblity of a plethora of applications, research needs to be carried out by large groups of people [20]. A research agenda of this sort requires involvement by multiple disciplines. Many useful research findings have been published, more research is underway. Interesting topics such as multiple target detection and classfication continue to challenge researchers, minimizing the level of false alarms in detection algorithms remains a research agenda and strategies for the deployment of these networks to achieve sufficient round-up of the phenomenon of interest are still being investigated.

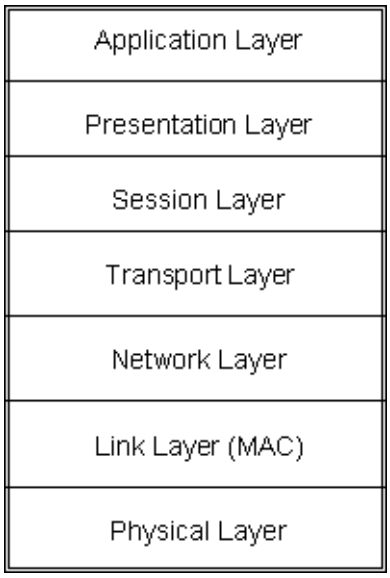


Figure 1.1: OSI Network Protocol Stack

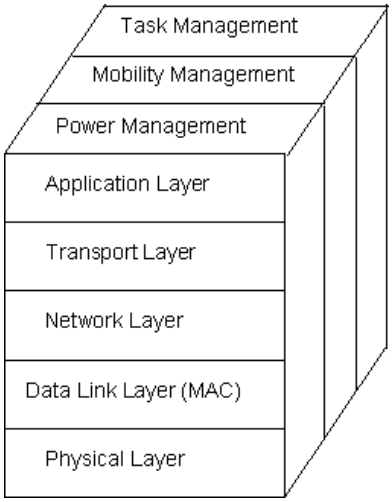


Figure 1.2: Sensor Network Protocol Stack [8]

To derive useful information about the environment, individual sensors are required to communicate captured data either to a remote base station or to nearby sensor nodes. Innovative data routing and communication protocols have been proposed [57],[24]. The need for sensor networks to operate in an autonomous fashion has led to the development of several self-organization methods including winner election, clustering and traffic-steered models as discussed in [46]. Although progress in research has resulted in the development of economical sensor nodes with multiple capabilities, the small size of these nodes come at a performance trade off [28]. This introduces challenges in the design of sensor networks.

1.1 Sensor Node Architectures

Sensor nodes have witnessed a progressive miniaturization over the past few years. A generic sensor node architecture is illustrated in Figure 1.3. However, different configurations are possible in which some of the components shown may be omitted e.g. the Global Positioning System (GPS) may not be required on some nodes or multiple sensing elements like acoustic and seismic sensors can be used on a single node. The particular configuration usually depends on the application for which the sensor node is intended. Configurable architectures in the form of unused connection buses to accomodate more components are also being designed [28]. The PC-104 based sensor nodes offered high computing capability on a relatively large computer board. The development of WINS sensoria nodes marked a new turn in sensor node development. An attempt to further reduce the physical size of sensor nodes led to the μAMPS initiative carried out at Mas-

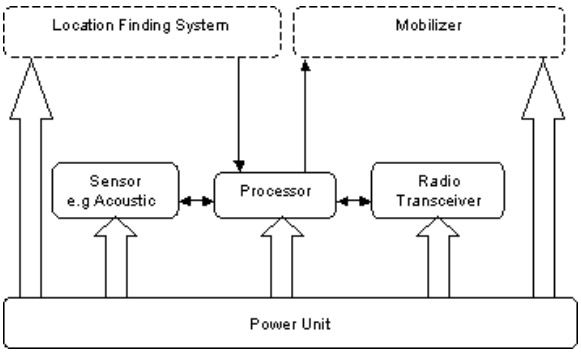


Figure 1.3: Sensor Node Architecture

sachussets Institute of Technology (MIT) [53]. More recently the development of the motes platform highlighted the opportunities for smaller nodes [28]. Further size reduction may be expected [1].

1.2 Wireless Sensor Networks

Wirelessly connected distributed sensor nodes form a unique class of networks known as Wireless Sensor Networks(WSN). A typical sensor network architecture is illustrated in Figure 1.4. The sensors are deployed in a field or in an environment where an event is to be detected. The user sends a query for data over a wireless medium. The choice of the distance between the base station or querying point and the network is based on many variables including accessibility and the transmitting range of the sensor network.

We summarize the key issues posed when designing networks of sensor nodes.

- **Fault Tolerance**

  Fault tolerance refers to the ability to sustain the functionality of a sensor network in spite of contingencies. Faults may come in the form of node fail-

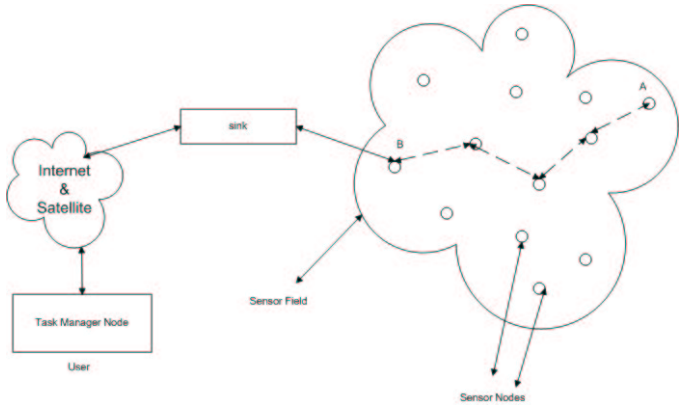


Figure 1.4: Sensor Network Architecture

ures, energy depletion or intentional destruction by hostile or enemy forces. Sensor networks must exhibit some fault tolerance to perform its internded purpose such as detection, tracking, classification [33].

- **Scalability**

  Depending on the application, a sensor network may consist of varying number of nodes from tens to thousands. Networks of millions of nodes have been envisioned [12]. A useful algorithm in sensor networks therefore, needs to be able to maintain its performance characteristics in networks of different sizes.

- **Production Costs**

  Since some applications of sensor networks may require a large number of nodes, the cost of each sensor node is a major determinant of the cost of the network. It is desirable to keep the cost of a sensor network to a minimum. The major cost component of a sensor network is the cost of manufacturing a sensor node. Recent advances in MEMS technology have resulted in the

development of low-cost nodes compared with traditional systems. As an example, the bluetooth radio system is manufactured for less than $10 [45] while sensor nodes are expected to be less than $1 [44]. This will make the deployment of sensor networks feasible for most proposed applications.

- **Operating Environment**

  Densely deployed sensor nodes are subject to various environmental conditions depending on the nature of the application. In general, nodes can be deployed either close to or inside a phenomenon. Examples include, attaching nodes to the body of animals, deploying in a biologically or chemically contaminated field, at the bottom of an ocean, etc. The design of the network needs to take into consideration the unique characteristics of the particular environment in which the network is to be deployed.

- **Sensor Network Topology**

  Hundreds or thousands of sensor nodes may be densely deployed in a sensor field. Node densities can be as high as 20 nodes/m^3 [52]. Deployment may be carried by different means including dropping them from an airplane, through a catapult or by other suitable means. The potentially large numbers of these sensor nodes preclude a carefully planned deployment policy. In addition, since the nodes need to operate unattended, unfavorable conditions such as energy depletion, destruction of a node or malfunctioning of a node may affect the topology of the network. This requires the maintenance of network topology.

- **Hardware Constraints**

A sensor node typically consists of a sensing unit, a processing unit, a transceiver unit and a power unit. Optionally, a location finding system or a power generator may be included depending on the intended application. For many sensor network applications, a large number of sensor nodes need to be deployed for reasons discussed previously. The different hardware components affect the characteristics of the sensor node hence the sensor network. Thus, the hardware that constitutes the sensor node needs to be carefully selected. For example, it is usually impractical to replace the power unit once the sensor node is deployed. The transceiver unit is known to consume relatively high amount of energy compared with the other hardware components.

- **Transmission Media**

  In the design of a sensor network, the transmission media is an important consideration. In WSNs, the nodes communicate using the wireless medium. The transmission can be carried out using radio, infrared or optical modes. It is important that the chosen media are available worldwide in order to enable global operation of these networks [8]. This indicates that the nodes manufactured for use in sensor networks should employ international standards.

- **Power Consumption**

  Nodes in a WSN are typically equipped with batteries and it may be impossible to recharge them. However, the energy consumption of a sensor network is related to the properties of the hardware used to manufacture

the nodes in the network. This has led to the design and production of ultra-low power sensor nodes. The choice of algorithms, sensors and the duty cycle of the sensor network also determine the power consumption. Power consumption is important because once a node runs out of energy, it affects the balance of the whole network in terms of its ability to deliver service.

These issues are important because they impact the design of protocols and algorithms for sensor networks. In addition, these factors can be used to compare different algorithms designed to be used in a sensor network.

1.3 Applications of Wireless Sensor Networks

Sensor networks have potential application in a myriad of situations. These applications usually require the capture of data about the environment including event detection, target localization or tracking where sensors can reliably record data about a phenomenon and subsequently provide the data or decision for analyses. These may be real-time applications, mission-critical or simply ad-hoc where the network is setup only when needed. Example applications are discussed below

- **Environmental Applications**

  Many fascinating phenomena in the environment can be better understood by using sensor networks. The behavior of birds, small animals and insects can be studied elaborately by deploying sensors in strategic locations. Sensors also make the monitoring of environmental conditions that affect crops and livestock easier. Other environmental applications of sensor networks

include irrigation, precision agriculture, forest fire detection, flood detection and pollution studies. Some sensor network based systems are already in place for flood detection. For example the ALERT system [2] consists of sensors that provide information about rainfall, water level and weather.

- **Health Applications**

  WSNs are capable of playing important roles in the delivery and managment of health care. These applications include the provision of more intelligent and interactive interfaces for the disabled, disease diagnosis, administration of drugs in hospitals, tracking and monitoring patients inside the hospital and telemonitoring of human physiological data. Other health applications of sensor networks are monitoring the movements and internal processes of insects and other small animals. Some of these applications require that the subject carry the sensing device which is designed to be light-weight but capable of transmitting sensed data in a flexible manner.

- **Home Applications**

  Sensors and Actuators have been found to be useful in home automation and the design of smart environments. Home automation has been envisioned in various forms in fictional movies and by researchers [41]. Sensors can be embedded in home appliances like VCRs, refrigerators and ovens which may be connected to an outside network such as the Internet and thus enable remote monitoring of the home. This form of remote monitoring can also be leveraged in security systems for the home. The building of smart environments can also result in a more intelligent interaction between

facilities in a building enabling them to offer better services either in a human-centered manner or a technology-centered way as discussed in [27]

- **Commercial Applications**

  Commercial applications of sensors and sensor networks include the monitoring of material fatigue, managing inventory, automatic control and guidance of manufacturing systems, factory process control and automation, detecting and monitoring car thefts, vehicle tracking and detection [5],[52],[60]. These applications add value in different forms when applied appropriately. For example, attaching a sensor to each item in the warehouse makes it easy to locate the item and tally the number of items in each category. It also becomes easier to track the depletion of inventory and plan to prevent stock-outs.

- **Millitary Applications**

  Wireless Sensor Networks have been identified as an integral part of millitary command, control, communications, computing, intelligence, surveillance, reconnaisance and targeting (C4ISRT) systems [8]. Millitary operations are varied and sensor networks are capable of providing intelligence about the battlefield terrain, enemy forces or environmental conditions in a timely fashion. Millitary use of sensor networks include monitoring friendly forces, ammunition and equipment, battlefield surveillance, targeting in guidance systems, battle damage assessment, nuclear, biological and chemical attack detection and reconnaisance. Details of these millitary applications may be found in [8]

In the applications discussed above, the quality of service provided by the sensor network depends largely on the reliability and lifetime of the network. For the network to be able to provide consistent service, it must be able to cancel out the effect of adverse circumstances such as node failures, energy depletion, topology changes by itself. This autonomous behavior is termed *self-organisation*.

1.4 Energy Consumption in Wireless Sensor Networks

The lifetime of a sensor network is a critical parameter in evaluating the usefulness of the sensor network. The lifetime of a sensor network is determine by the energy consumption of the network. The energy consumption in a sensor network has been studied extensively and key findings are presented in this section. Energy consumed in WSNs can be divided into three main areas,

- *Sensing*: The energy consumed by a sensor in sensing an event is related to the nature of sensing. Sporadic event sensing generally consumes less energy than continuous event sensing.

- *Communication*: Communication has been identified as the major energy consumer in sensor networks. The characteristics of the transceiver used by a sensor node strongly affect the energy consumption characteristics of the sensor node, hence the network. In general, the energy utilized in transmission of data is more than the energy used to receive the same data. The design of extremely low power transceivers is discussed in [**?**].

- *Data Processing*: The amount energy consumed in data processing is significantly less than the energy used for data communication. However, as the complexity of the computations to be performed increases the energy used for processing the data also increases [43]. Hence, it is important to manage the amount of energy consumed in data processing.

1.5 Self Organization in Wireless Sensor Networks

Due to the unreliable operating environment typical of sensor networks, there is a need for a sensor network to reorganize in order to nullify any adverse effect when necessary. Node failures, energy depletion and changes in task dynamics are possible reasons why a network may need to reorganize.

The concept of self-organisation refers to the ability of a system to adaptively restructure itself in response to an interfering mechanism or condition and to maintain a balance with respect to given metrics of performance [6]. This has become a goal in sensor network research as it is important in order for the network to operate autonomously with an acceptable Quality of Service (QoS). Insights into self-organization in sensor networks are provided in [13].

1.6 Contribution of Research

This research presents a novel approach for selecting a subset of deployed nodes in a densely deployed WSN. This selection procedure is motivated by the need to manage bandwidth and prolong the lifetime of a dense network. The proposed algorithm provides a framework for determining the distance to which data cap-

tured by a sensor network remains strongly correlated. This correlation distance is used to select node subsets in a manner that coverage and connectivity are consistently achieved. Some level of redundancy is retained in the network and the coverage provided by these subsets is somewhat uniform. The algorithm potentially achieves a high degree of fault tolerance. The algorithm also provides free parameters to the network designer who is at liberty to determine the operating lifetime of the network. A new set of metrics are also employed in evaluating different aspects of the algorithm.

This thesis documents the details of the algorithm and the results of simulations carried out in an attempt to evaluate the performance of the proposed algorithm.

1.7 Structure of the Thesis

The organization of this thesis is described in this section. This chapter introduced the area of sensor networks and developments that enabled its advancement. Chapter 2 overviews relevant subject areas that informed the approach proposed in this dissertation. A review of related literature is undertaken in Chapter 3. Chapter 4 details the proposed algorithm while the simulation framework and performance evaluation results are presented in Chapter 5. Conclusions, discussions and indications for future work are noted in Chapter 6.

Chapter 2

Theoretical Background

In this chapter, the subject areas used in the derivation of the proposed algorithm are overviewed briefly. Sensor networks as a unique field demands the use of tools from a variety of disciplines. This multidisciplinary approach has produced useful algorithms spanning the entire protocol stack of sensor networks. Principles from areas such as spatial statistics, detection theory and network protocol design provided ideas for the development of the protocol described in this thesis.

2.1 Spatial Statistics

The subject of spatial statistics evolved out of the need to account for the spatial behavior of data. The concepts that grew into spatial statistics were originally applied to time series data in order to understand the interdependence between data over time and create a model for predicting future data based on observed trend. Most statistical analyses assume that data obtained at different times or locations are independent mainly for the purpose of making such analysis tractable.

However, it has been observed in many physical phenonmena that data sample taken over space exhibit a relationship with each other. This location-dependent relationship is generally known as *spatial dependence*. It is characterized by the similarity between data recorded at close locations in statistical terms. The classical assumption that data is drawn from independent and identically distributed (i.i.d) distributions is a convenient one that is done to make much of mathematical and statistical theory easy to track [14]. However, models that incorporate spatial dependence between data are often more realistic. Using notation taken from [14], a general spatial model can be constructed as follows.

Let $\mathbf{s} \in \mathbb{R}^{\mathbf{d}}$ be an arbitrary data location in d-dimensional Euclidean space and suppose the potential datum $\mathbf{Z}(\mathbf{s})$ at spatial location $\mathbf{s}$ is a random quantity. Let $\mathbf{s}$ vary over the index set $D \subset \mathbb{R}^d$ so as to generate a multivariate random field (or random process):

$$\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D; \tag{2.1}$$

Three types of spatial data distributions are well-known depending on the manner in which they occur. They include Geostatistical data, Lattice data and Point patterns.

- **Geostatistical Data**

  This type of data is commonly found in mineral ore deposits. Early work [21] attempted to use the concept of spatial dependence to model variations within the deposit bed. The field of Geostatistics emerged from years of research in the analysis of such underground deposits and has become a useful

tool in fields like mining engineering, geology, mathematics and statistics. The key goal in Geostatistics is to predict the grade of minerals at different points using a limited number of observed samples. The methods of Geostatistics have been used in other fields like soil science where it may be desired to characterize soil properties from a small number of soil samples taken from a field. Geostatistical data is different from other forms of spatial data in that it is spatially continuous over the spatial index $\mathbf{s}$.

- **Lattice Data**

  As the term lattice implies, lattice data includes all data arrangements that form a pattern in which neighbors of each data point bear a similar relationship to each other. Although the term lattice evokes a sense of a regular structure, there are irregular lattices. In the case of irregular lattices the patterns are unpredictable. An example of regular lattice data in $\mathbb{R}^2$ is the remotely sensed data generated by a weather satellite. In a weather image large regular block sections of the earth are represented by a pixel in the final image. In such an image, the pixels form a regular grid. This pattern is referred to as a *lattice structure*.

- **Point Patterns**

  Point patterns arise when the important variable to be analyzed is the location of events or objects of interest [14]. For example, in the analysis of vegetation, trees grow at the points to be analyzed, hence they form a point pattern. Point patterns may exhibit randomness, clustering behavior or regularity. The node distribution in a sensor network can be modeled as

a point pattern.

The signals emitted by a target or phenomenon can be viewed as geostatistical in nature because it is continuous over space. This is especially true if the signal is examined at a given time instant. For the purposes of this discussion, it may be assumed that the spatial variables are isotropically generated in the region of interest.

2.1.1 Geostatistical Analysis

The analysis of geostatistical data has been used extensively in understanding geographic data, estimation of underground data deposits such as mineral deposits and in image analysis. A common approach to geostatistical analysis is to obtain data samples from the deposit in a systematic way. The most important problem in the analysis of geostatistical data is to predict data values at unsampled locations from the observed samples. This process of prediction has been termed *kriging*. In order to carry out spatial analysis, a model of the observed phenomenon is required. This is necessary in order to incorporate space parameters into the analysis. This method of analysis, taking spatial characteristics into characteristics into consideration, deviates from the popular assumption of independence between data samples. It is based on the premise that data that are close together in space (and/or time) are often more alike than those that are far apart. A spatial model must therefore take this spatial variation between samples into account. The classical nonspatial model is a special case of the spatial model when the spatial parameters are removed. In the past few years, significant consideration has been given to incorporating spatial models into scientific as well as

statistical observations in general [14]. Analysis of geospatial data is exploratory by nature and provides a mechanism for making important decisions like whether or not to mine. The estimation of spatial dependence is the first step in most geostatistical analysis. It is based on the assumption that the geostatistical data is intrinsically stationary.

2.1.2 Intrinsic Stationarity

In this subsection, a statistical model is introduced (see [14]). Suppose that the grade of an ore body at a point $\mathbf{s}$ in $\mathbb{R}^2$ is the realization of a random process $\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D$ and that it is observed on a square grid at points $\mathbf{s_i} : i = 1, ..., n$. Then intrinsic stationarity is defined through first differences as:

$$E(Z(\mathbf{s}+\mathbf{h})) - Z(\mathbf{s})) = 0; \tag{2.2}$$

$$var(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}); \tag{2.3}$$

The quantity $2\gamma(\mathbf{h})$ is known as the *variogram* and is an important parameter in geostatistical analysis. The assumption of intrinsic stationarity refers to the property that the process has a constant mean as shown in Eq. 2.2 and a variance that depends on distance as shown in Eq. 2.3. A rearrangement of Eq. 2.3 leads to the classical estimator of the variogram (Eq. 2.4) as proposed by Matheron [21]. This form of variogram has the advantage that it is unbiased but it possesses very poor resistance properties since it is affected by unusual data and outliers due to the $(.)^2$ term. Other forms of the variogram have been proposed. [36] proposed

a robust variogram estimator that is more resistant than the classical variogram.

2.1.3 Estimating Spatial Dependence

Spatial dependence reflects the correlation between data located in close proximity. This is generally estimated using the variogram. The variogram is a plot of the variation in covariance structure of the underlying spatial process with distance. It can be obtained by rearranging Eqs. 2.2 and 2.3 to derive

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(s_i) - Z(s_j))^2, \tag{2.4}$$

where $Z(s_i)$ and $Z(s_j)$ are signal energies at any two sensor nodes in the network. $N(\mathbf{h}) \equiv (i,j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}; i,j = 1,...n$ and $|N(h)|$ is the number of distinct elements in $N(\mathbf{h})$.

Note that the variogram as defined above is a second-order moment. The relationship defined by Equation 2.4 is a scale-factor of a related relationship called the *semi-variogram*. The relationship between the variogram and the semi-variogram is given as

$$semi - variogram = 0.5 * variogram \tag{2.5}$$

2.1.4 Stationary Processes

Although the assumption of intrinsic stationarity was made in the derivation of Eq. 2.4 in order to create a geostatistical model, further assumptions are needed

to correctly fit this model into practical situations in which only a finite set of samples of the phenomenon is available. This relaxed set of assumptions allow the definition of a set of processes called *second-order* (or *wide-sense* or *weak*) stationary.

Second-Order Stationarity

A random function $Z(.)$ is said to be second-order stationary if it satisfies the following conditions.

$$E(Z(\mathbf{s})) = \mu \tag{2.6}$$

$$\text{cov}(Z(\mathbf{s_1}), Z(\mathbf{s_2})) = C(\mathbf{s_1} - \mathbf{s_2}),\ \forall \mathbf{s_1}, \mathbf{s_2} \in D, \tag{2.7}$$

The function $C(.)$ is called a stationary covariance function or *covariogram* [14]. Further assumptions about the process may be necessary.

2.1.5 Variogram

The variogram as defined in Eq. 2.4 is a crucial tool in spatial statistics as it provides the basic parameter necessary for most operations including prediction and estimation of data values. Suppose,

$$var(Z(\mathbf{s_1}) - Z(\mathbf{s_2})) = 2\gamma(\mathbf{s_1} - \mathbf{s_2}),\ \forall\, \mathbf{s}_1, \mathbf{s}_2 \in D; \tag{2.8}$$

The quantity $2\gamma(.)$ is a function of the incremental distance $\mathbf{s}_1 - \mathbf{s}_2$.

If the variogram is strictly a function of the distance between data samples and is not affected by the direction in which it is estimated then the variogram so obtained is said to be an *isotropic variogram*. If the resulting variogram varies with the direction of estimation, then the variogram is said to be an *anisotropic variogram*.

Estimation of the Variogram

The variogram is estimated by varying the distance between the data samples used in Eq. 2.4. Due to spatial spread and continuous nature of the data under consideration, data samples are taken over the spatial index $\mathbf{s}$.

The classical variogram estimator shown in Eq. 2.4 requires that the data samples to be used are placed at evenly spaced distances from each other. This constant distance may be regarded as the granularity of the estimation since it controls the number of points that are used in the estimation of the variogram. The separation distance between any two points is often called the *lag*. The variogram is estimated by computing the covariance between all data points separated by the same distance at any given separation $\mathbf{s_i} - \mathbf{s_j}$. The results obtained are then summed and averaged over the number of pairs that are separated by that distance $N(\mathbf{h})$. A plot of the resulting averages for all lags against the lag represents the variogram. Figure 2.1 shows a generalized ideal variogram shape with typical parameters. The *Sill* is the variance (γ) value at which the variogram levels off indicating no further variations. The distance corresponding to the first intersection of the sill is called the *Range of Influence* and it is the effective distance beyond which the samples cease to be strongly correlated. The *Nugget* is an effect usually

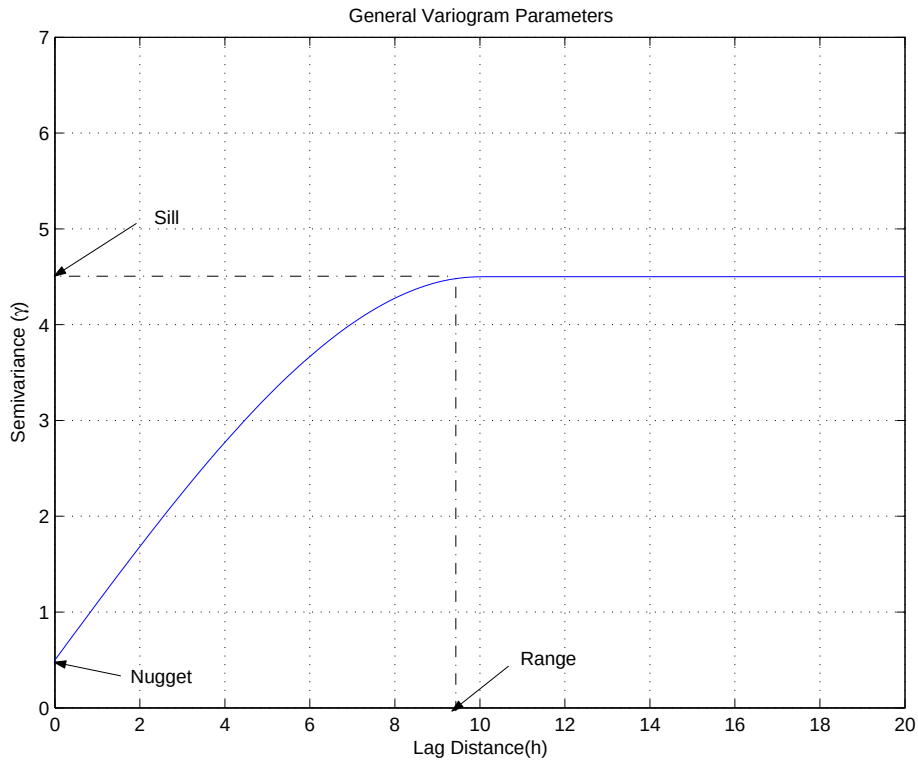


Figure 2.1: Generalized Shape of the Variogram

attributed to small scale variation in the data or due to measurement error [14].

Variogram computation can be done using either a *directional* or *omni-directional* approach. The directional variogram estimation takes point pairs in specific directions, for example, a North-South directional variogram may be computed for a given deposit. The omni-directional variogram estimation considers the points in the given region as a single set. Point pairs are determined by considering all possible directions (0 to 360^0) up to a given tolerance in order to accomodate points that may be slightly out of reach. The kind of variogram obtained from recorded data is known as the *Experimental Variogram* and often deviates markedly from the general shape shown in Figure 2.1. Therefore, theoretical variograms are of-

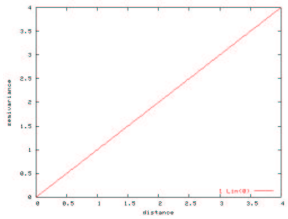
ten fitted to experimental variogram models for the purpose of determining the variogram parameters.

Theoretical Variogram Models

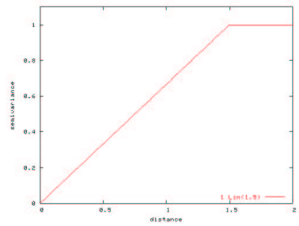
The natural variogram estimator due to Matheron [21] is a second-order moment which is typically used in creating the experimental variogram. In many practical cases, the estimated variogram does not follow a precise parametric form. However, theoretical variogram models can be used to approximate these practical variograms by *fitting* them to the practical variograms. Some theoretical variograms are shown in Figure 2.2 and Figure 2.3.

The theoretical variograms defined in Table 2.1 have been normalized so that the amplitude in each case is 1. Nugget parameters which generally appear as an added constant are not shown. K_1 is the modified Bessel function of the second kind. The parameter $'a'$ is the horizontal distance scaling factor and is related to the Range. The exponential, Bessel and Gaussian functions reach their sill asymptotically. The 'effective range' is the distance where the variogram reaches 95% of its maximum value ($3a$ for exponential, $\sqrt{3}a$ for Gaussian and $4a$ for the Bessel).

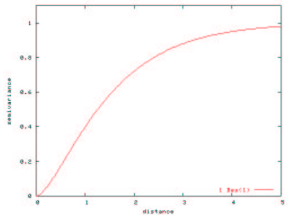
Open source and commercial packages are available for spatial analysis. GSTAT is an example in the open-source category and provides an easy interface for various geostatistical analysis functions [4].



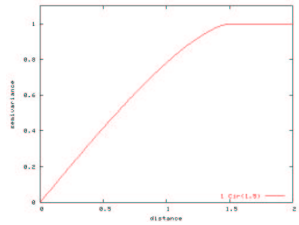
(a) Linear variogram (no sill)



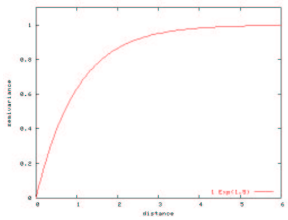
(b) Linear variogram (with sill)



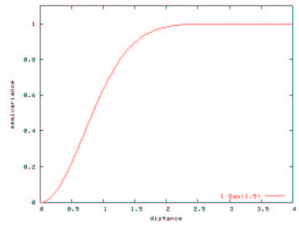
(c) Bessel variogram



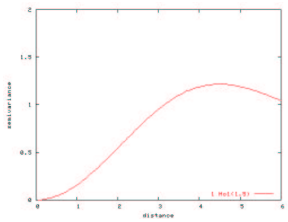
(d) Circle variogram



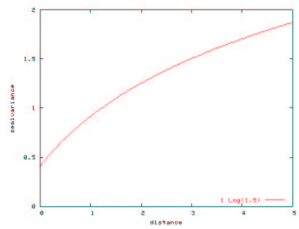
(e) Exponential variogram



(f) Gaussian variogram

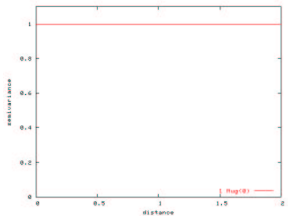


(g) Hole variogram

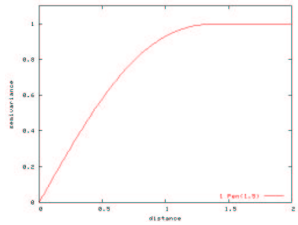


(h) Logarithmic variogram

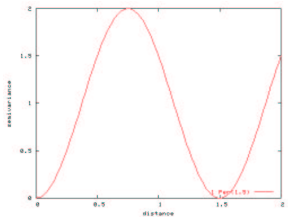
Figure 2.2: Examples of Theoretical Variogram models



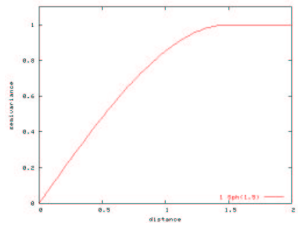
(a) Nugget variogram



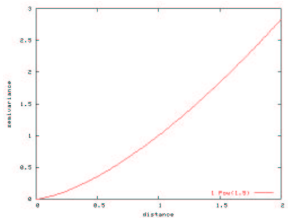
(b) Pentaspherical variogram



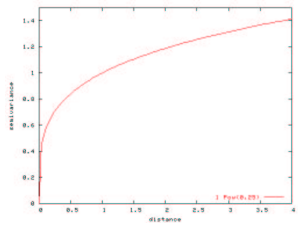
(c) Periodic variogram



(d) Spherical variogram



(e) Power variogram ($a < 1$)



(f) Power variogram ($a > 1$)

Figure 2.3: Examples of Theoretical Variogram models (contd.)

Table 2.1: Theoretical Variogram Models [40]

Model	$\hat{(}\gamma(h))$	Valid h range
Nugget	0 1	$h = 0$ $h > 0$
Spherical	$\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3$ 1	$0 \leq h \leq a$ $h > a$
Exponential	$1 - \exp(\frac{-h}{a})$	$h \geq 0$
Linear	h	$h \geq 0$
Linear-with-sill	$\frac{h}{a}$ 1	$0 \leq h \leq a$ $h > a$
Circular	$\frac{2h}{\pi a}\sqrt{1 - (\frac{h}{a})^2} + \frac{2}{\pi}\text{arcsin}\frac{h}{a}$ 1	$0 \leq h \leq a$ $h > a$
Pentaspherical	$\frac{15h}{8a} - \frac{5}{4}(\frac{h}{a})^3 + \frac{3}{8}(\frac{h}{a})^5$ 1	$0 \leq h \leq a$ $h > a$
Gaussian	$1 - \exp(-(\frac{h}{a})^2)$	$h \geq 0$
Bessel	$1 - \frac{h}{a}K_1(\frac{h}{a})$	$h \geq 0$
Logarithmic	0 $\log(h + a)$	$h = 0$ $h > 0$
Power	h^a	$h \geq 0, 0 < a \leq 2$
Periodic	$1 - \cos(\frac{2\pi h}{a})$	$h \geq 0$

2.2 Signal Detection and Classification Theory

Detection theory is a body of techniques used to determine whether or not a signal is present. In general, three types of signals are encountered [42]; deterministic signals in which all related signal parameters are known, completely random signals which require that their probability distributions are known and a hybrid in which the signal is known except for a set of unknown (possibly random) parameters. The problem of signal detection can be viewed as one involving the discrimination between the presence and absence of a signal. For example, there may be a need to decide whether an observed waveform consists of "noise alone"

or "signal masked by noise". The preceding example falls under signal detection. Signal classification involves the decision relating to identifying the class that a detected signal belongs to out of a prespecified set of signals [26]. The objective of signal detection and classification theory is to provide strategies for designing algorithms that minimize the average number of decision errors. The techniques used in detection and classification theory come from mathematical statistics and decision theory [25].

In the context of statistical hypothesis testing, a pair of hypothesis is presented and a decision needs to be made about the prevailing hypothesis given the observed signal samples.

Consider a general binary detection model in which the task is to detect the presence or absence of a target or phenomenon. Assume that the signals captured are Gaussian distributed and can be modeled as follows:

$$\begin{aligned} H_0 : x_k &= n_k \\ H_1 : x_k &= s_k + n_k \end{aligned} \tag{2.9}$$

Here, H_0 is the null hypothesis, H_1 is the alternative hypothesis, n_k is additive white Gaussian noise with a mean of zero and independent of signal samples (s_k) which may be non-zero mean signals. The optimal detector for a binary hypothesis testing model of this form is the likelihood ratio test [42]. The test compares the ratio of the conditional densities under each hypothesis with a fixed threshold denoted by η. The likelihood ratio test (LRT) may be written as:

$$L(x) \gtrless \eta. \tag{2.10}$$

When $L(x) > \eta$ the detector decides H_1 else it decides H_0. In the following example, the LRT is applied in the detection of an arbitrary constant signal in additive white Gaussian noise having zero mean. The binary hypothesis can then be written as [58]:

$$\begin{aligned} H_0 : x_k &= n_k \\ H_1 : x_k &= A + n_k \end{aligned} \tag{2.11}$$

where A is the value of the constant voltage signal.

The probability densities under these hypotheses are respectively

$$p_{x|H_0}(x|H_0) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{(2\pi)}} exp(-x_i^2/2\sigma^2) \tag{2.12}$$

$$p_{x|H_1}(x|H_1) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{(2\pi)}} exp(-(x_i - A)^2/2\sigma^2) \tag{2.13}$$

substituting into the likelihood ratio test, Eq. 2.10, gives

$$L(x) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{(2\pi)}} exp(-(x_i - A)^2/2\sigma^2) / \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{(2\pi)}} exp(-x_i^2/2\sigma^2) \tag{2.14}$$

simplifying Eq. 2.14 results in the form given below

$$L(x) = \prod_{i=1}^{N} exp(\frac{2Ax_i - A^2}{2\sigma^2}) \tag{2.15}$$

2.2.1 Detector Performance

The preceding discussion introduced the hypothesis testing approach to the detection of signals using the likelihood ratio test. In general, there is a need to evaluate the performance of a detector. This can be done using the Receiver Operating Characteristic (ROC). Given two probability density functions under the hypotheses H_0 and H_1. First, consider two Gaussian densities of a random variable as shown in Fig. 2.4. The two densities have different means (μ_1 and μ_2) and equal variance (σ). Let the densities correspond to different hypothesis. The problem of detection can then be viewed as determining which hypothesis is true based on the result of comparing the likelihood ratio with a given threshold (x^* in Fig. 2.4).

A quantity referred to as the *discriminability* describes the ease with which signals can be distinguished. It is related to the distance between the distributions under each hypothesis. Discriminability, d', is defined as:

$$d' = \frac{|\mu_2 - \mu_1|}{\sigma}. \tag{2.16}$$

As shown in Fig. 2.4, some probability parameters relevant to this model [17] can be defined as follows

- $P(x > x^*|x \in H_1)$: a *hit* – the probability that the signal is present when the detector says it is present. It is generally termed the probability of

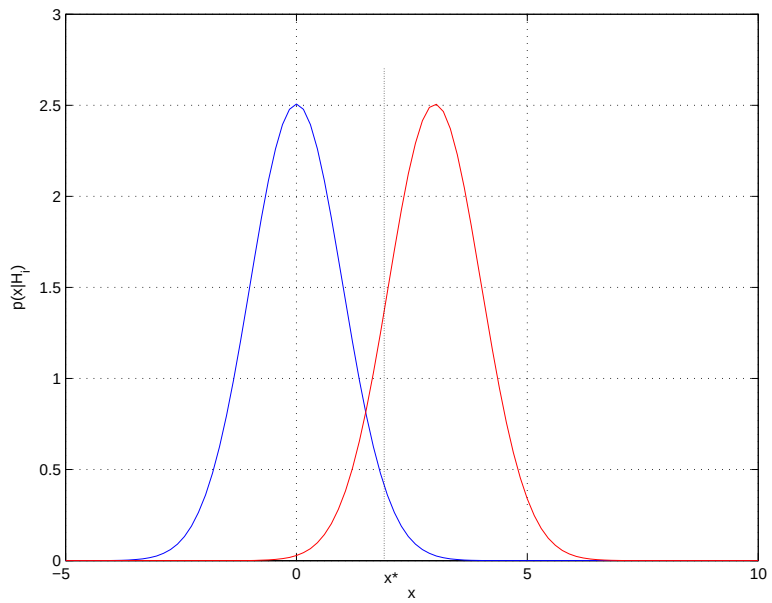


Figure 2.4: PDFs under a binary hypothesis test showing the threshold

detection (P_d).

- $P(x > x^*|x \in H_0)$: a *false alarm* – the probability that the detector says the signal is present although it is not. This is also called the probability of false alarm (P_f).

- $P(x > x^*|x \in H_1)$: a *miss* – the probability that the detector says the signal is absent when it is actually present.

- $P(x > x^*|x \in H_0)$: a *correct rejection* – the probability that the signal is absent when the detector says it is absent.

Given the parameters defined above then it is possible to plot a curve that describes the performance of a detector designed to detect a signal under the hypothesis test. The ROC is a plot of the Probability of detection(P_d) against the Probability of false alarm (P_f). For the constant signal detection given in Eq. 2.11

the ROC can be plotted by varying the threshold x^* from 0 to ∞ and integrating over the H_1 decision region (i.e. to the right of the threshold).

Since the LRT is a ratio of two random variables, it is also a random variable. The logarithm of a likelihood ratio is an equivalent test to that given in Eq. 2.10 because the logarithm is a monotonic function and helps simplify the resulting product. Taking the logarithm of the likelihood ratio obtained in Eq. 2.15 gives

$$\frac{A}{\sigma^2}\sum_{i=1}^{N} x_i - \frac{NA^2}{2\sigma^2} \gtreqless \ln\eta \tag{2.17}$$

which can be re-written as

$$l = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N} \gtreqless \frac{\sigma}{A\sqrt{N}}\ln\eta + \frac{A\sqrt{N}}{2\sigma} \tag{2.18}$$

obtained by multiplying Eq. 2.17 by $\sigma/A\sqrt{N}$. Under H_0, l is normal ($N(0,1)$) and $N(A/\sigma, 1)$ under H_1 . Therefore, l is subject to the distribution described in Fig. 2.4. Hence, the probability of detection (P_D) and false alarm (P_F) are obtained by integrating $p_{l|H_1}(L|H_1)$ and $p_{l|H_0}(L|H_0)$ respectively to the right of the threshold for each value of the threshold, L.

$$P_D = \int_{(\ln\eta)/d+d/2}^{\infty} \frac{1}{\sqrt{(2\pi)}} exp[-\frac{(x-d)^2}{2}]dx \tag{2.19}$$

$$P_F = \int_{(\ln\eta)/d+d/2}^{\infty} \frac{1}{\sqrt{(2\pi)}} exp(-\frac{x^2}{2})dx \tag{2.20}$$

where $d = A\sqrt{N}/\sigma$ corresponds to the discriminability defined in Eq. 2.16. The two integrals, Eqs. 2.19 and 2.20 can be evaluated using the error function. The

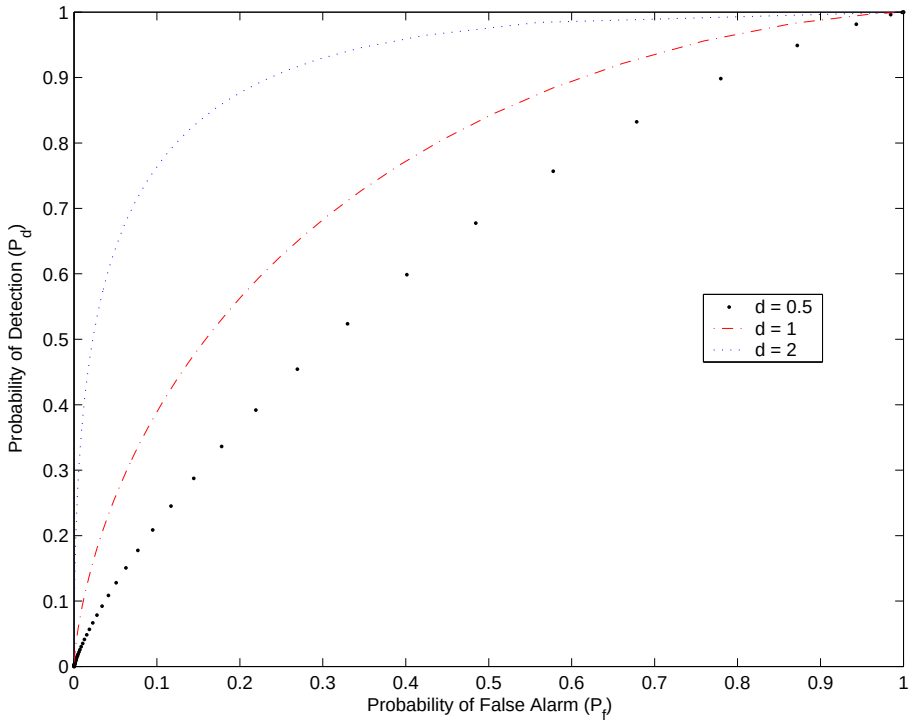


Figure 2.5: ROC for the detection of a constant dc signal

resulting ROC is given by Fig. 2.5 for different values of d.

2.3 Network Protocol Design

A protocol defines the format and the order of messages exchanged between two or more communicating entities, as well as the actions taken on the transmission and/or receipt of a message or other event [32]. A large number of protocols are in use on the internet, proprietary networks and multimedia systems. In general, protocols aim to achieve desired goals unique to the particular computing infrastructure or environment in which it is intended to operate e.g. protocols for ad-hoc wireless networks, sensor networks, etc.

According to [29], a protocol has five basic elements:

1. The service to be provided by the protocol.
2. The assumptions about the environment in which the protocol is to be executed.
3. The vocabulary of messages used to implement the protocol.
4. The format of the message in the vocabulary.
5. The procedure rules guarding the consistency of message exchanges.

In the context of sensor networks, the challenges of limited energy supply, bandwidth considerations, small memory availability, lower computation capability, density of the deployed network all need to be taken into account while designing a suitable network protocol. The five elements of a protocol completely specify its functionality and provides information about its internal operation. Most protocol design in sensor networks seeks to optimize the constraints listed above.

In traditional computer networks, a layered approach is used. The layered approach enables modular interaction between the different layers. This implies that different layers can be modified without affecting other layers. The Internet is an example of a network based on conventional network protocol design principles. Table 2.2 lists typical protocols in a Network and the layers at which they operate.

The TCP and UDP protocols at the transport layer, Table 2.2, can be used to illustrate trade-offs that can be implemented in protocols to reflect the type of service that such protocols are designed to deliver. As transport layer protocols,

Table 2.2: OSI layers and typical Network Protocols

OSI Layer	Network Protocols
Application	HTTP, SMTP, SNMP, FTP, Telnet
Presentation	SMTP, SNMP, FTP
Session	NetBIOS, DLC
Transport	TCP, UDP
Network	IP
Data Link	802.11
Physical	Ethernet, Token Ring, FDDI

they are responsible for moving data from the session layer to the network layer. Although both TCP and UDP have the same objectives, they differ in the way they operate. For example, the TCP protocol is a connection-oriented protocol designed to provide guaranteed delivery of data. TCP offers reliable data transport through built-in mechanisms like acknowledgements, retransmissions and timeouts. UDP is a connectionless protocol which does not guarantee data delivery rather, a best-effort transmission is attempted.

In order to adapt these networking principles to sensor networks it is often necessary to reduce the level of overhead involved in the execution of a protocol. Due to the peculiar structure of sensor networks the OSI stack has been adapted to ensure efficiency and functionality as in Figure 1.2.

The functions of the different layers of the sensor protocol stack are as discussed below.

- **Physical layer**

  This is the first layer of the OSI networking model. It performs services requested by the data link layer. This layer consists of physical hardware,

radio transceiver and other interfaces. The main functions at this layer are the establishment and termination of a connection. Modulation schemes are implemented on this layer.

- **Data link layer**

  The data link layer is level two of the OSI model. It connects the network layer to the physical layer. Service requests from the network layer are received by this layer while it relies on the physical layer for services. This layer provides error detection and correction in the transmitted data and ensures that data is transmitted seamlessly between the Network layer and the Physical layer.

- **Network layer**

  The network layer is level three of the OSI model. It responds to service requests from the transport layer and issues service requests to the data link layer. This layer is responsible for addressing and naming and subsequent translation of those addresses when in the receive mode. In addition, this layer determines the route of data from source to destination. It handles functions such as routing and managing packet congestion.

- **Transport layer**

  It responds to service requests from the application layer and issues service requests to the network layer. The transport layer provides transparent transfer of data between hosts. It is responsible for end-to-end error recovery and flow control. The purpose of the Transport layer is to provide transparent data transfer between end users, thus relieving the application layer

from any concern with providing reliable and cost-effective data transfer.

- **Application layer**

  This layer interfaces directly to application processes and performs common application services for them. An example of common application services is data processing for detection classification or tracking. Since sensor networks are mostly application driven, services tend to run on the application layer.

Research in sensor networks has led to the development of customized protocols for the different layers of the OSI stack. Due to the constraints encountered in a sensor network the protocols are designed to accomodate these challenges while offering acceptable service. Table 2.3 shows examples of protocols used in sensor networks and the OSI layer at which they operate [8].

Table 2.3: OSI layers and typical Sensor Network Protocols

OSI Layer	**Sensor Network Protocols**
Application	SQTL [50], SMP [8]
Transport	–
Network	LEACH [24], SPIN [23], Directed Diffusion [31], Gossiping [22]
Data Link	SMACS and EAR [55], Hybrid TDMA/FDMA [51]
Physical	Ultrawideband (UWB)[8]

Chapter 3

State of the Art: Efficient Approaches in Sensor Networks

In this chapter, an overview of relevant algorithms and techniques proposed in literature for coping with various challenges in sensor networks is provided. In particular, examples of works that seek to minimize energy consumption and efficiently utilize communication bandwidth are presented. In order to fully capture the state of the art in energy management and bandwidth efficient techniques in sensor networks this chapter is divided into sections. The first section discusses energy-efficient techniques that aim to prolong the operating lifetime of a sensor network while the second section discusses methods that have been proposed to manage bandwidth utilization. The last section deals with the concepts of coverage and connectivity which have been identified as essential requirements in many applications of sensor networks.

3.1 Energy-Efficient Techniques

As noted in Chapter 1, energy consumption in a WSN can be divided into three domains; communication, sensing and computation. Thus, energy efficiency generally implies optimal apportionment of the energy requirements in each of the three consumption domains. Communication has been identified as the chief energy consumer in a sensor network. Many techniques have been investigated with the aim of minimizing energy consumption in sensor network. These approaches either attempt to modify the topology of the sensor network in order to reduce the energy utilized during communication of data (or information) or to reduce the amount of data transmitted. Other approaches include the idea of operating the sensor nodes in modes that optimize the duty cyle of the nodes.

3.1.1 Clustering

A sensor network typically spans a wide area and sensed data needs to be communicate to remote locations. A topological rearrangement may bring certain nodes close together and achieve significant network lifetime extension [24]. [24] discusses a clustering approach called LEACH (Low Energy Adaptive Clustering Hierarchy) that organizes the network into clusters. An analysis of the lifetime of networks using this clustering approach shows that network lifetime can be significantly increased compared with networks not using clustering. [35] modifies the approach used in LEACH to achieve further extension in network lifetime. Other topological approaches have been proposed in literature.

3.1.2 Power-Saving Modes of Operation

Another proposed approach for controlling energy consumption in wireless sensor networks is operating the sensor nodes in power saving modes of operation. Power saving modes are modes in which the components of a sensor node are either turned on or off depending on the operation that need to be performed [53]. Table 3.1 shows these modes of operation. It is based on the principle of switching on or off the transceiver, processor or sensor on a sensor node. Significant energy saving can be achieved when a carefully designed switching policy is implemented. As noted in [8] switching between modes does not necessarily save power . Power savings are only achieved if the time the node spends in a certain mode exceeds a certain threshold. [53] shows the energy consumption for each of the states based on the μAMPS sensor platform. The values are as shown in Table 3.2.

Table 3.1: Power Saving Modes of Operation [53]

Sleep State	Processor	Memory	Sensor	Radio
s_0	Active	Active	On	Tx,Rx
s_1	Idle	Sleep	On	Rx
s_2	Sleep	Sleep	On	Rx
s_3	Sleep	Sleep	On	Off
s_4	Sleep	Sleep	Off	Off

Table 3.2: Power consumed in different sleep states [53]

Sleep State	Power(mW)
s_0	1,040
s_1	400
s_2	270
s_3	200
s_4	10

3.1.3 Power Control

It has been established that the major source energy depletion in a sensor node is communication. Therefore, it is desirable to control the transmit power used in data transmission. The process of controlling the power with which data is transmitted based on transmit distance is termed power control [53]. Power Control has been used in many wireless systems including bluetooth [3] and in CDMA systems [47]. In general, only a discrete set of power levels are allowable.

3.2 Bandwidth-Efficient Techniques

Recent research activity has also focused on the efficient use of bandwidth. Envisioned applications of sensor networks that take advantage of the small size of sensor nodes involve large numbers of sensor nodes. In a dense network, bandwidth is likely stringent for most conventional networking protocols. Therefore, a useful algorithm for use in these networks will need to consider bandwidth efficiency.

3.2.1 Mobile Agent Computing Paradigm

Traditional computing models use a centralized processing center which serves as destination of data/information in a distributed network. However, this approach may put a lot of burden on the central node. If the central node is more powerful than other nodes then it may have to be placed at a careful location to ensure that data from other nodes can easily reach it. A recently proposed paradigm is the Mobile agent computing paradigm in which a mobile agent is sent from the processing center and migrates between the nodes to perform data processing at each local node. It then retrieves a result which it carries from node to node. As it travels between nodes it fuses the classification/detection result with the fusion result obtained from previous nodes until it achieves a specified confidence level. It has been shown in [62] that the mobile agent approach achieves bandwidth efficiency compared with the client/server computing model because the agent is usually of small size. A summary of the advantages of the Mobile agent technique is given below [62]:

- **Network Bandwidth Requirement** is reduced since the size of the mobile agent is small. This may be important in situations where the communication is through low bandwidth wireless connections.

- **Stability and Reliability:** Mobile agents are robust to node failures because agent transfers take place when the link is alive.

3.2.2 Clustering

The clustering technique has been shown to provide good bandwidth management [24]. The clustering technique exploits the proximity of neighbors to organize sensor nodes into smaller units. In addition to saving energy used in communication between nodes, the density of a cluster is typically much less than that of the entire network thus resolving bandwidth issues like collisions and network congestion. [56] found through simulations that the performance of a network drops when the load offered from the sensor nodes exceeded the capacity of the network. They identified the need to explore innovative congestion avoidance schemes that provide adequate performance for the particular network application. Some suggestions given include reducing the data reporting rate per sensor, turning sensors off and fusing information in a manner that optimizes the communication operations.

3.2.3 Resource Selection

Selective use of resources or data in a sensor network is a broadly used approach to cope with the challenges in a sensor network including the management of bandwidth [49], [54], [53]. A densely deployed sensor network possesses a high degree of redundancy [19], the number of sensor nodes in the network may be much more than the number required to monitor a given event, target or phenomenon. Due to the fact that sensor nodes may need to operate over a long period of time whether or not there is a target introduces redundant sensing and communication. Indeed, some applications may require continuous sensing while others only need sporadic sensing. Similarly, communications can also be continuous or intermittent. These

redundancies can be reduced while achieving a satisfactory level of service from the sensor network. Effective strategies have been proposed that control these excessive and possibly unnecessary use of resources.

3.3 Coverage and Connectivity

As discussed in Chapter 1, a sensor network is often used to detect, track, localize or classify a target [33]. Under these circumstances, there is a need for the network to sufficiently surround the phenomenon or target and be able to communicate data from one end of the network to the other. Hence, the emergence of research in coverage and connectivity. Coverage is used here to refer to the ability of a group of sensors to provide a resultant sensing field that is continuous throughout the region. Connectivity however, refers to the ability of nodes to communicate data from one end of the network to the other. Researchers have defined coverage and connectivity from different perspectives and using different metrics. [9] provides an elaborate analysis of these definitions and approaches. Coverage and connectivity are related concepts in that they have to do with the ability of the sensor network to establish a field of influence that is completely connected without discontinuities.

3.3.1 Coverage

Coverage has been viewed as an important factor in judging the QoS of a sensor network [39]. Let the number of sensor nodes covering a point in a field be denoted by K. Then the covering of that point can be said to be K-covering. In order to

achieve an acceptable QoS for a particular application, an average level of covering is required. Coverings above this required value can be regarded as redundant. For densely deployed networks, redundant sensing is predominant. In order to analyze the coverage in different network configurations, many methods have been used including that of the Voronoi tesselation and Delaunay triangulation [39],[34]. Circle covering techniques [61] have also been used to analyze coverage [54].

3.3.2 Connectivity

The subject of connectivity has attracted a lot of research in recent years. It is a crucial characteristic of a sensor network due to the need for data to be communicated between nodes, routing of information through the network and other similar services. [30] discusses an efficient technique for achieving connectivity in sensor networks. [24], [35] also consider the need for connectivity. Among the requirements of a useful sensor network is the ability to faithful transmit its data to a user. A loss of connectivity typically implies that data from the network may be lost in transit. This can adversely affect decisions reached about the target or phenomenon being monitored.

3.4 Related Work

In [54], a node subset selection algorithm was discussed based on the idea that in a densely deployed sensor network, exclusive sets of sensors could be used to monitor a given area. Using this method, a different subset is activated each time in a manner that looks like a handing over between the different subsets. However,

the main objective of the paper was to provide a technique for determining this set of sensors offline. It was mentioned in their paper that a deployment strategy suitable for the developed approach was under investigation. [54] considers two methods used to solve for the subsets namely; simulated annealing and a heuristic algorithm. The heuristic algorithm surpassed the simulated annealing method in performance.

The subject of data subset selection is closely related to node subset selection in that data is retrieved selectively from specific nodes. Sestok et al.[49] considers the problem of selecting data in the network in order to manage bandwidth and extend network lifetime. They proposed a simplified randomization scheme in which each sensor decides whether or not to send data in a particular time slot based on a preset probability value. In their method, all nodes are active and must make a decision in every time slot. In addition, [49] analyzed the impact of data selection on detection from a signal processing perspective. The approach they proposed although practical and simple may not significantly control the energy constraint in sensor networks. It does not consider the need to capture data from all over the network in a manner that coverage is achieved by the chosen subset. If the randomization technique is used for node selection, it is equivalent to a random selection of nodes. There is also the possibility that selected sensors are far apart which may impact the performance of routing and data dissemination algorithms. This can lead to loss of network connectivity.

In this thesis, a technique is proposed that makes no assumption about the topology of the deployed network. Rather a framework is developed that seeks to

exploit the redundancy that is present in a dense network in order to select a spatially spread out subset of the sensors. After predetermined durations this subset is changed. It is shown that the subsets, while not necessarily exclusive are determined autonomously by the network and consistently achieves coverage and connectivity. The algorithm explicitly uses information about node energy and exhibits robustness to fault while extending network lifetime.

Chapter 4

Spatial Selection

The problem of subset selection is primarily a mathematical problem. Selecting a representative subset from a large dataset can be computationally intensive, leading to the development of efficient techniques. Areas such as mathematical regression seek a subset of points that best fit a set of data points. These points may be viewed as representative of a given universal set of points subject to given optimality criteria [7]. Typical optimality criteria include the Mean Squared Error(MSE) and Minimum Mean Squared Error (MMSE) [7],[17],[37]. However, most of these techniques demand complex iterations and may be time consuming, thus the need to develop simplified techniques for use in sensor networks.

The randomization technique discussed by [49] randomly selects a sensor node based on the outcome of a probability variable. Randomization has been viewed as a practical approach to most subset selection problems but it often results in the selection of skewed subsets [7]. Skewness of a data set may be interpreted as a one-sided view of a target or phenomenon in sensor networks. More perspectives of the

phenomenon can be obtained through a more robust selection technique. In the context of sensor networks where challenges such as limited energy and insufficient bandwidth exist, timeliness and efficiency are desirable. In this chapter, a technique that leverages the autonomous and unattended operation of a sensor network is developed as an approach for a sensor network to manage energy and bandwidth. Selection refers to the process of determining which sensor nodes are retained in the working state while others go the sleep state. The working state is characterized by working transceivers and sensors while in the sleep state the transceivers are turned off. For the purposes of this thesis node states such as those proposed in [53] (also in Table 3.1) are used. The sensor nodes in the working state are regarded as selected nodes. The subsets selected are such that nodes are elected from all over the network. Therefore, the guidelines for designing a robust node selection protocol include low latency, simplicity and minimal energy consumption. The time taken for an algorithm to converge is generally proportional to the amount of energy consumed by the algorithm. A node selection algorithm needs to execute in short time in order to minimize energy consumption. To achieve simplicity, the use of replies and acknowledgements commonly used in networking have been avoided.

4.1 Development of the Spatial Selection Algorithm

Given the unique nature of sensor networks: the ability to operate autonomously, capability to interact with neighboring sensor nodes and collaboratively derive intelligence about the environment, a robust node selection algorithm will neces-

sarily take into account the opinions of neighbors when selecting a representative subset. It should be noted that the primary aim of sensor networks is to detect a target or event in the environment in which it is deployed. As mentioned in Chapter 2, sensor nodes have a small form factor and the ability to carry out computations and multimodal sensing. These coupled with the need to deploy sensor nodes in a distributed fashion provides the opportunity for nodes to interact with each other. However, these sensor nodes may be deployed in large numbers. In general, node deployments are not deterministic either due to the inaccessibility of the monitored area or in hazardous situations, e.g. in millitary operations where nodes may be deployed by dropping them off an airplane. The assumed scenario for the node selection algorithm is that of a densely deployed network of randomly located nodes. A typical configuration of such a dense network is as shown in Figure 4.1. It can be observed that in order to choose a subset of these nodes without external interaction, each node may decide to elect itself randomly. This is similar to the randomization technique. If the sensor network is viewed as a geographic region with spatially distributed points, the problem of selecting a representative subset of the sensors becomes that of developing an appropriate sampling plan. "Representative" is used to refer to the property of the selected subset to remain connected while providing sensing coverage. Geographic sampling plans are discussed in [10],[48]. Geographic sampling plans are used for the estimation of populations e.g. plant populations, insect population densities etc. However, in the sensor network scenario, the aim is to choose a subset of the nodes using sampling plans. In this situation, the sytematic sampling plan using quadrats is applicable.

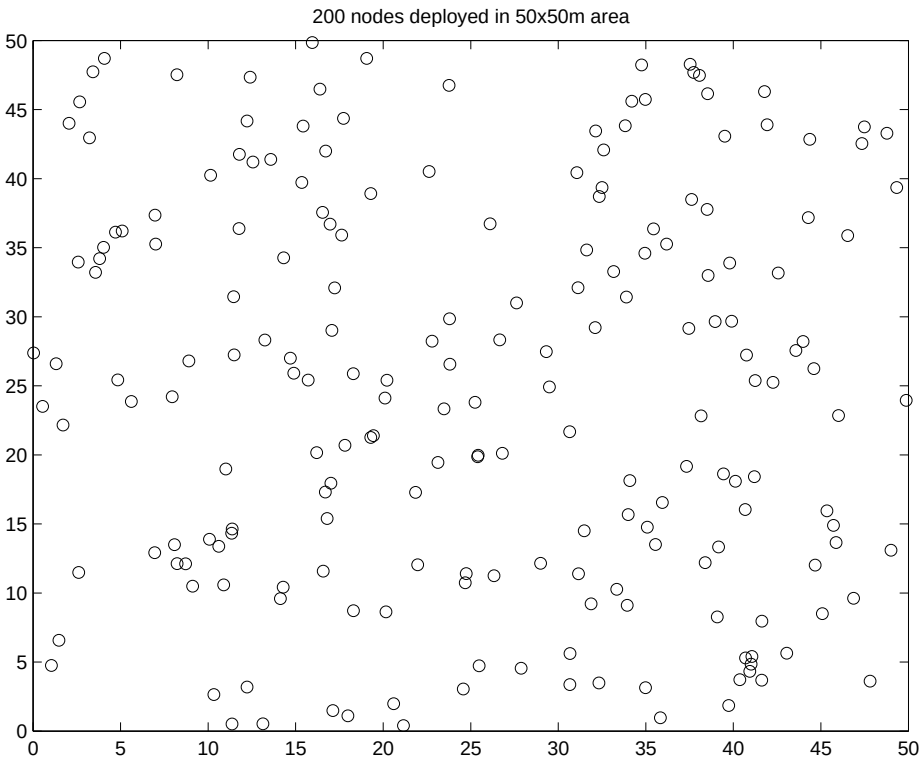


Figure 4.1: Random deployment of sensor nodes in a dense network

The systematic sampling plan using quadrats randomly places a quadrat in the geographic region while the population within each quadrat is noted. The final population figure for a region is estimated by assuming that the number of quadrat throws is representative of the entire population. In a modified procedure, the quadrats can be intentionally placed to form a grid that cover the whole area. A related procedure uses traverses to create a grid over the field.

The remaining task is to decompose the systematic schemes described above into an autonomous procedure executed cooperatively in a manner suitable for sensor networks. This will require the selection of a limited number of nodes within each sampling quadrat (rectangular region) or mutually perpendicular traverses.

In the simplified scheme shown above, each quadrat has a specific dimension denoted by d, each square of side d encloses a set of sensors which may be assumed homogeneous in capacity and function.

In order to establish a useful metric for the selection of these sensor nodes in the prescribed square sections, we note that the sensor network is dense and has a significant amount of redundancy. Although redundancy may be desired in order to obtain a robust sampling of the sensed phenomenon, it can be exploited to choose a sufficient number of working nodes. We attempt to quantify the level of redundancy in the network by determining the degree to which sensor readings in close proximity resemble each other. This may be done using the concept of spatial dependence as discussed in the Chapter 2. In order to estimate spatial dependence we use the semi-variogram. The semi-variogram gives an estimate of the dependence between sensors in a densely deployed sensor network in the form of a parameter termed the "Range" or "correlation length". When the correlation length is determined it can be used as the selection parameter. For example, to define the radius of the basic circle or the length of the quadrat to be used in sampling the network. Figure 4.2 can be redrawn to become Figure 4.3 using circular discs as the sampling element. This redefines the systematic sampling approach with squares as a sampling with overlapping discs or circles.

It is worth noting here that the circle is an attractive sampling element because in order to achieve coverage and redundancy the discs only need to overlap in order to cover all the points in the network as in Figure 4.3. Although this circle covering issue is related to the circle covering problem [61], we do not approach the problem using this method. The overlapping arrangement of the discs ensures

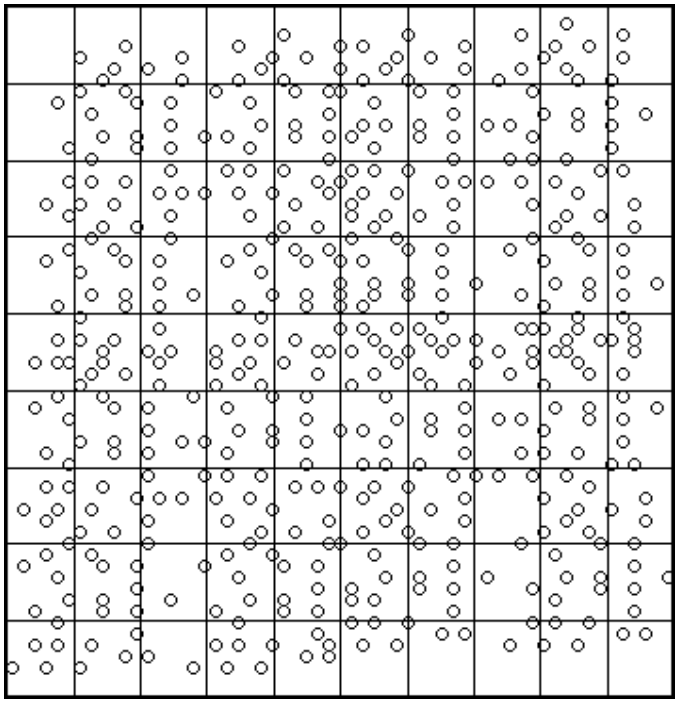


Figure 4.2: Systematic sampling of a sensor field

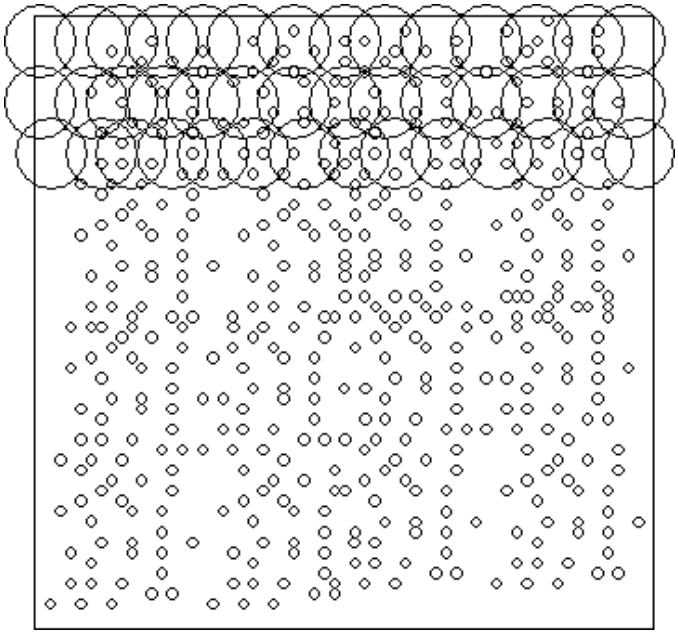


Figure 4.3: Systematic sampling with discs

that some redundancy is retained in the final selection.

Thus, to select sensors in a sensor field we propose the determination of the "Range of Influence" as discussed in Section 2.1 (called *Correlation Distance* henceforth) of the target signature in the sensor field and using this value as a sampling parameter of the sensor field. Specifically, the correlation distance will be used as the radius of the sampling discs. The determination of the correlation distance is described next.

4.1.1 Determination of Correlation Distance

Range or correlation distance is generally estimated by taking samples of data from the target, spatial process or event of interest. However, in the given scenario, the goal is to determine the correlation distance in the absence of real data. A model is generated for the signals emitted by a phenomenon and data captured is simulated using imaginary sensors placed at different points in the sensor field. Then, the empirical semi-variogram [14] is determined for the sensor field. As noted in Chapter 2, the empirical semi-variogram requires that the points used in the estimation of the semi-variogram be placed equidistant to each other. This simulation approach is flexible and allows arbitrary placement of sensor nodes for the purpose of determining the signal energy captured at a given time instant.

As a specific example, the semi-variogram is determined for a sensor field with acoustic sensors, deployed to detect a moving target such as a vehicle passing through a field. Let the target be modeled as an isotropic radiating source which generates a continuous spatial random process over the monitored region. The signals emitted by the target can be represented by an inverse law decay [33],

$$E_{sensed} = E_{source}/d^{\alpha} \tag{4.1}$$

where E_{sensed} is the value sensed at a node, E_{source} is the amplitude of the signal radiated by the target, d is the euclidean distance between the target and the node. In the case of acoustic signals, $\alpha = 2$. Similar signal models have been assumed in [15],[18].

The noise in the given area is assumed to be additive, white and Gaussian (AWGN). The effect of noise on the signals captured can be ignored since the semi-variogram of independent Gaussian noise is a horizontal line with no sill as in Figure 4.4. Note that because the target is moving, the model stated above is applied for a given snapshot of time.

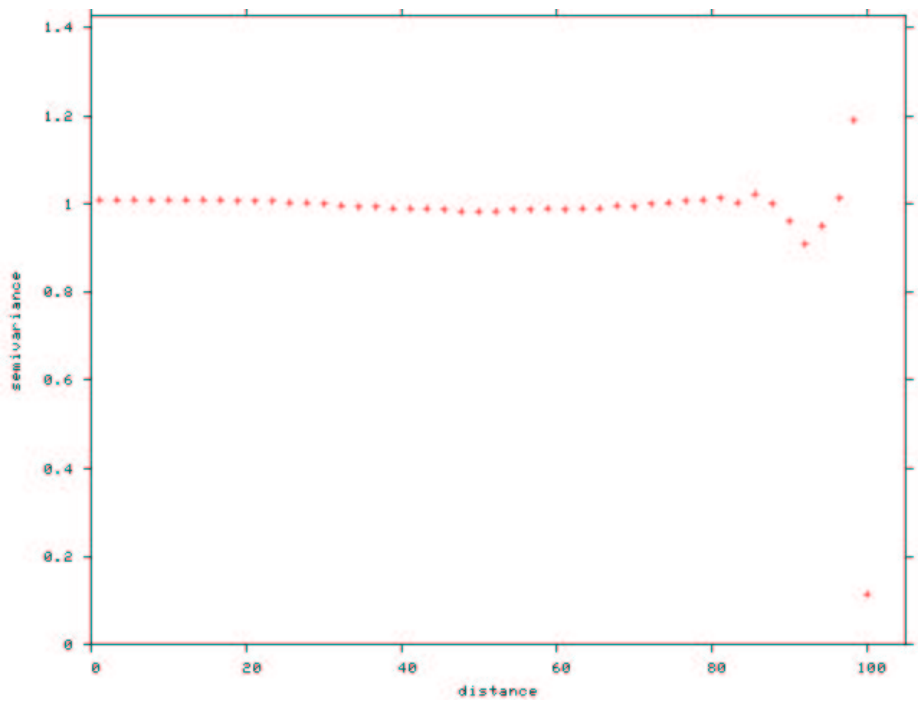


Figure 4.4: Semi-variogram for independent Gaussian noise

It may be possible to determine the correlation length of the sensor field on the fly while the target is travelling but it will require collaboration between sensor nodes and can involve overhead which may lead to energy wastage. Therefore, the approach proposed uses pre-deployment estimation using a simulated sensor field as discussed above. Semi-variogram estimation can be carried out using the techniques discussed in Chapter 2.

Using the acoustic model, Eq. 4.1, the resulting power-law semi-variogram (also known as the generalized semi-variogram) is as shown in Figure 4.5. The variogram was obtained for 10,000 simulated sensor nodes in a 50m x 50m field but varying the density does not change the shape. This type of variogram is characterized by infinite variance and does not have a definite correlation distance since it does not have a sill. However, we retain this approach as a valid method for determining the correlation distance in sensor networks. Note that the value of the correlation distance is independent of the amplitude of the received signal $E_{received}$. Thus, in the simulation carried out above, the signals captured are generated using a normalized signal model. In order to overcome the lack of a specific correlation distance, under the given model, the selection algorithm is designed using an arbitrary value for the selection distance. We suggest this value should be significantly less than the sensing radius of the sensors in the network or the maximum transmission distance of the sensor nodes (whichever is smaller). This intuitively ensures coverage and connectivity. In general, a small value can achieve these goals.

While the discussion above is based on an acoustic signal model, other signal models are possible. [15] discusses a space-time representation for signal modeling.

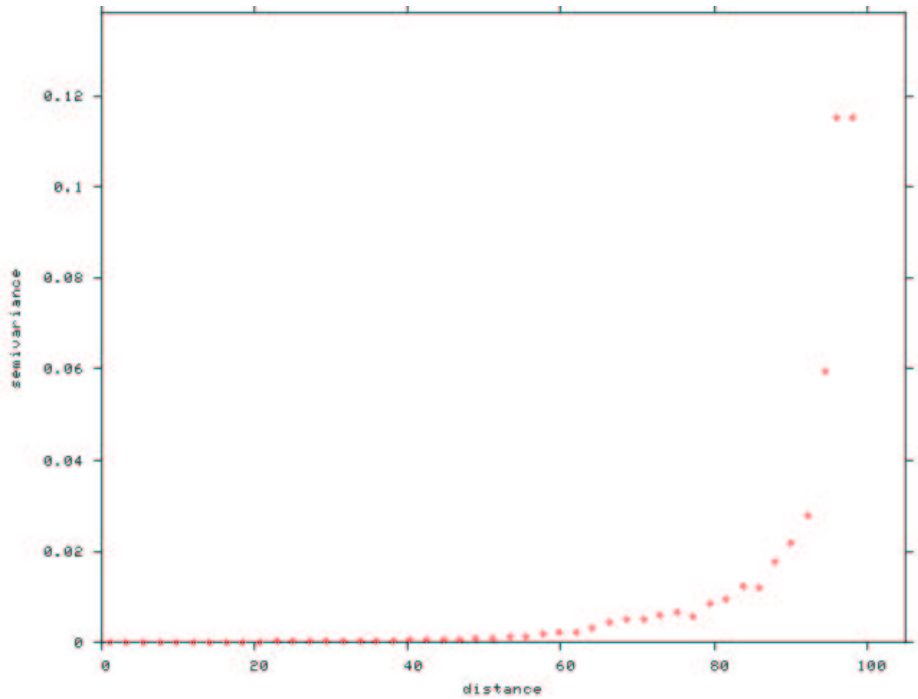


Figure 4.5: Power-law semi-variogram for model in Eq. 4.1. Produced using GSTAT [4]

A probabilistic signal model may also be used in cases of completely stochastic signals or a mixture of deterministic and random signals.

4.2 Details of the Selection Algorithm

Under this algorithm, a node operates in three states namely ACTIVE, SLEEP, NEGOTIATE. The state diagram, Figure 4.6 depicts the transition between these states. The ACTIVE state refers to the condition in which the node can sense, receive and transmit data. In the SLEEP state, the transceiver is turned off. The states here should not be confused with the Power saving modes of operation discussed in Chapter 3.

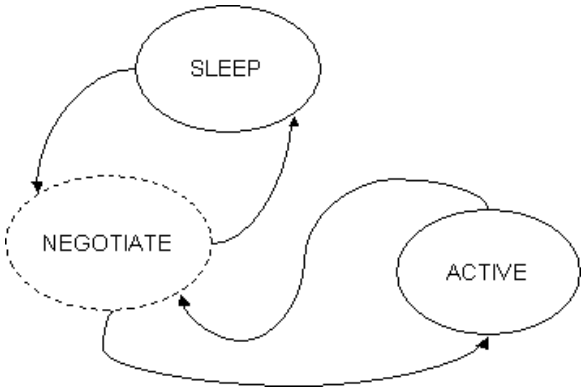


Figure 4.6: State Machine for Selection Algorithm

In the discussion of the selection algorithm, the correlation distance is designated as $r_{negotiate}$ because it serves as a negotiation distance for proximal nodes. All nodes go into the NEGOTIATE state to execute the algorithm. The algorithm proceeds as follows: Each node sends out a 32-bit packet containing its remaining energy to a distance corresponding to the correlation length ($r_{negotiate}$) of the network. All nodes within this distance receive the packet and compare the value received in the packet with their own remaining energy, if the value received is less than their own energy, they discard the received packet and continue to receive packets sent by other neighboring nodes, otherwise if the energy value received is greater than or equal to their own energy, they go to the SLEEP state. The pseudo-code for the algorithm is as shown in Figure 4.7.

```
begin
    r_negotiate ← constant;
    t_sleep ← constant;
    foreach Node n_i do
        n_i.state = NEGOTIATE;
        set transmitPower ∝ r_negotiate;
        energyPacket.send();
    end
    foreach Node n_j neighbor of Node n_i do
        energyPacket.receive();
        if Node n_j.energy ≤ Node n_i.energy then
            n_j.state = SLEEP
        else
            n_j.state = ACTIVE
        end
    end
end
```

Figure 4.7: Spatial selection algorithm

Note that some nodes may be isolated and negotiation packets sent from other nodes may not be able to reach them. In such cases, the approach is to include a timeout (a few microseconds or less), depending on the size of the network, after which if no packets are received the node goes to the ACTIVE state. The choice to put such a node in ACTIVE state is to ensure that data is recorded in the area covered by that node. This way, data gathering algorithms to be run including routing algorithms and data transmissions to a base station (typically far from the deployment field) that are likely to require higher transmission ranges will be

able to collect data from such node to the base station. It is also possible that over time, factors like wind may bring such nodes closer to other nodes giving the node a chance to go to SLEEP state in subsequent rounds of negotiation.

In the algorithm presented, nodes use power control to determine the transmission power level that corresponds to the chosen $r_{negotiate}$ value. This ensures that a small amount of power is used in negotiation. Note that the power control levels are discrete and the choice of $r_{negotiate}$ must take into consideration the possible levels. The sleep time (t_{sleep}) equals the time that working nodes stay in the ACTIVE state. All nodes may be initially deployed either in the SLEEP or ACTIVE state. The NEGOTIATE state is a transient state in which although the node is 'ACTIVE', it only participates in packet exchange for node selection and operates with parameters set for the selection algorithm. If deployed in the SLEEP state, all nodes will need to wake up after a preset amount of time. If deployed in the ACTIVE state, all nodes will have to wait a certain small amount of time (known by all nodes) before the algorithm is initiated. This will ensure 'soft' synchronization. This wait time also applies at execution of the algorithm to ensure that all nodes are awake and take part in the negotiation at the same time. Such wait times may be of the order of nanoseconds.

4.3 Implementation Approach

The implementation of the algorithm will be in two phases; a pre-deployment phase in which the selection distance (correlation distance) is determined and a post-deployment phase in which the algorithm is executed at the nodes. Three

application scenarios can be identified for this algorithm; data selection, node selection and sensor selection. Data selection refers to the process of "picking" data from a network after a phenomenon has been sensed. Node selection chooses appropriate nodes that achieve coverage and connectivity in a sensor network while sensor selection identifies sensors, either of the same modality or different, that can be used in given time slots.

4.3.1 Pre-deployment Phase

In this phase, the value of the correlation distance is determined through simulation using a suitable model for the phenomenon to be monitored, for example the variogram approach discussed previously. The nodes to be deployed are "hard-coded" with the $r_{negotiate}$ value and the sleep time (t_{sleep}). As outlined above, the value of $r_{negotiate}$ will be used to set the transmit power during NEGOTIATION.

4.3.2 Post-deployment Phase

Nodes may be deployed in either the SLEEP or ACTIVE state. When the algorithm is initiated, all nodes go into the NEGOTIATE state and begin sending energy packets. Nodes within the $r_{negotiate}$ circumference of neighboring nodes receive the packets and make a decision whether to go to the SLEEP or ACTIVE state. The algorithm can also be triggered by the occurence of an event such as the appearance of a target, by a remote request for data or other topological re-organization of the network e.g. cluster-head changes.

Chapter 5

Performance Evaluation

In an attempt to evaluate the performance of the proposed algorithm, a simulation was carried out. The simulation consists of creating scenarios that are similar to those encountered in a deployed network. The simulation experiments are discussed in this chapter. The results obtained are interpreted and analyzed and the potential advantages of this algorithm are shown. The parameters used in the setup of the simulation are outlined along with any assumptions made.

5.1 Simulation Parameters

In order to evaluate the performance of the algorithm, the first order radio model used in the LEACH and PEGASIS simulations [24], [35] is assumed. The transceiver consists of a transmit and receive circuitry which consume identical energy while extra power is consumed by the transmit amplifier. Under this simplified model, the radio dissipates $E_{elec} = 50$ nJ/bit for both the transmitter and receiver circuitry and $\epsilon_{amp} = 100$pJ/bit/m^2 for the transmitter amplifier to achieve an ac-

ceptable signal to noise ratio (E_b/N_0). An inverse square law energy decay is assumed for the transmitted signals. Thus, to transmit a k-bit message up to a distance d using this radio model, the following relationships hold

$$\begin{aligned} E_{Tx}(k,d) &= E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \\ &= E_{elec} * k + \epsilon_{amp} * k * d^2 \end{aligned} \tag{5.1}$$

To receive this message, the radio expends:

$$\begin{aligned} E_{Rx}(k) &= E_{Rx-elec}(k) \\ &= E_{elec} * k \end{aligned} \tag{5.2}$$

In the above, $E_{Tx}(k,d)$ and $E_{Rx}(k)$ correspond to the energy consumed by the transmitter electronics when transmitting a k-bit packet to a distance d and for receiving it respectively. Figure 5.1 shows the simplified first order radio model. Other parameters used in the simulations are listed in Table 5.1.

To establish a traffic pattern for the network, we assume a continuous constant bit rate (CBR) transmission of 16Kbps for each sensor node between rounds of selection. This traffic model is used solely for the purpose of showing the energy decay characteristics of different networks.

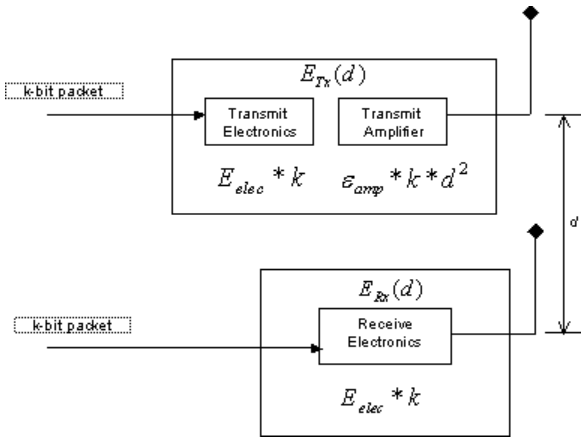


Figure 5.1: Simplified First Order Radio Model [24]

Table 5.1: Simulation Parameters used for performance evaluation

Network Area	50m by 50m
Node Placement	Random (uniform distribution)
Number of Nodes	200
Initial Node Energy	35 - 36 Joule (random)
Sensing Radius	12m

5.2 Simulation Experiments

5.2.1 Effect of the variation in Range parameter

In this simulation, the effects of choosing different values for $r_{negotiate}$ on the number of sensors selected and the spatial distribution of the selected subset are investigated. Selected sensors for $r_{negotiate}$ values of 3m and 5m are shown in Fig. 5.2 and Fig. 5.3 respectively. In these figures, the filled circles represent nodes in ACTIVE state while the unfilled circles represent nodes in SLEEP state. Figure 5.4 plots the number of selected sensors for different values of $r_{negotiate}$.

As shown in Figure 5.4, the number of selected sensors reduces with increasing

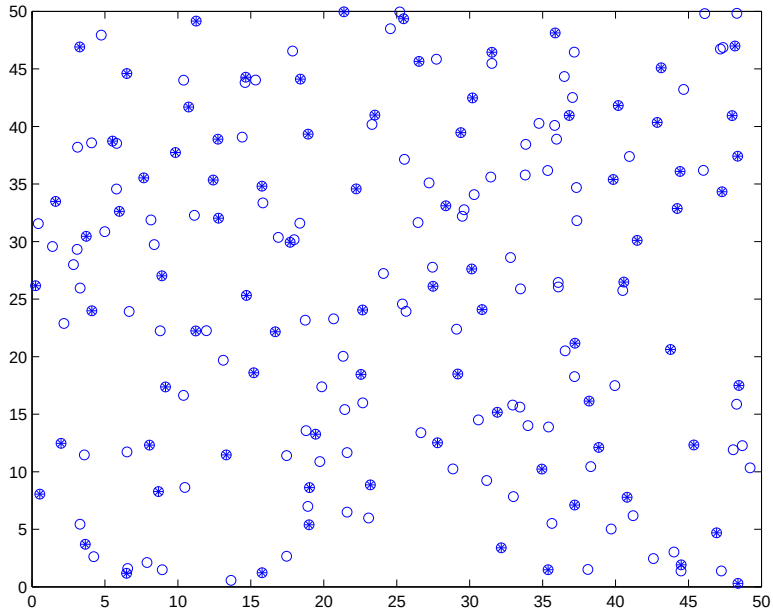


Figure 5.2: Node Selection using $r_{negotiate}$=3m

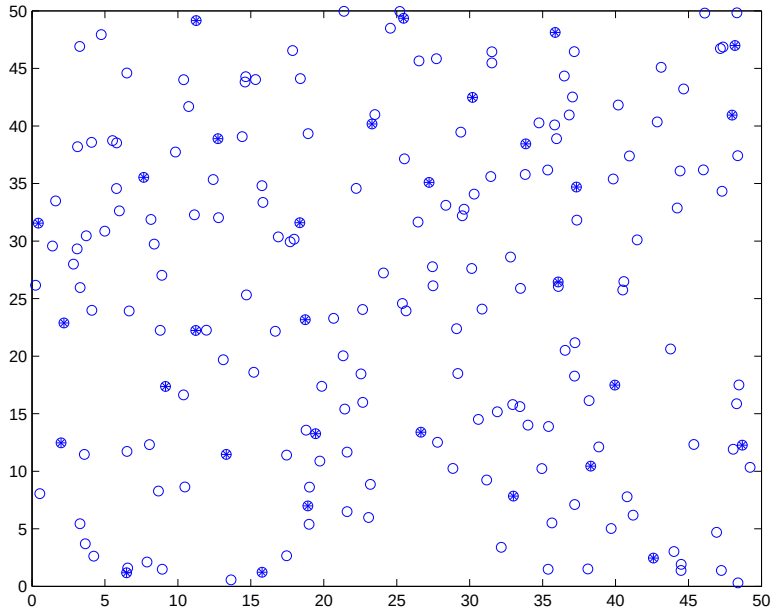


Figure 5.3: Selection using $r_{negotiate}$=5m

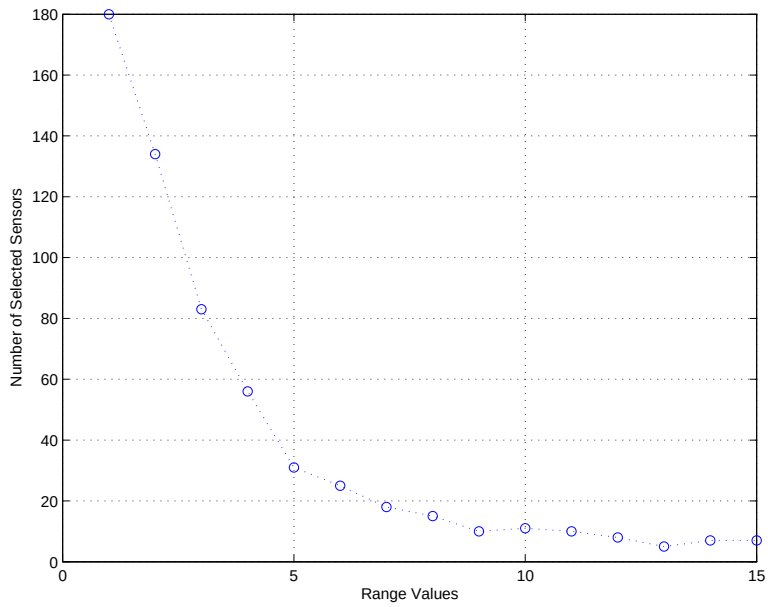


Figure 5.4: Plot of number of nodes selected for different $r_{negotiate}$ values.

$r_{negotiate}$ and the selection method ensures that sensors are selected from all over the field.

5.2.2 Effect of variation in Deployment Density

In this case, the number of nodes deployed in the given area is varied while $r_{negotiate}$ and t_{sleep} are held constant. Figure 5.5 depicts the scenario.

Figure 5.5 shows that for a given $r_{negotiate}$ and t_{sleep} more energy savings can be achieved for more densely deployed networks. This suggests that the choice of $r_{negotiate}$ and t_{sleep} for denser networks can be higher than that in sparser networks for a given network operating lifetime.

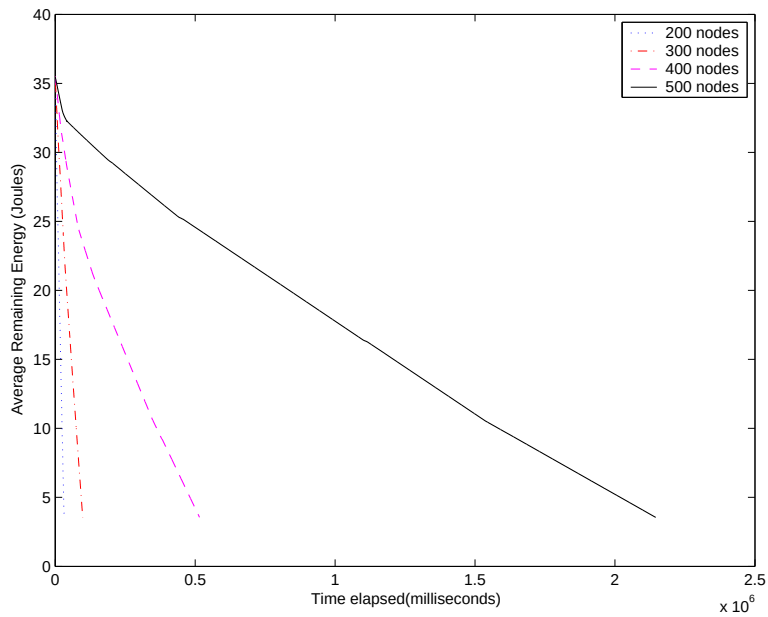


Figure 5.5: Energy consumption for different node densities using $r_{negotiate}$=3m and t_{sleep}= 5 seconds.

5.2.3 Effect of varying t_{sleep}

A variation in sleep time may also affect the energy consumption of the network. This is due to the fact that active sensor nodes may experience varying number of events (sensing and data transmission) during this time. Figure 5.6 shows that for a given $r_{negotiate}$ value a variation in sleep time has significant effects on the lifetime of the network. Variation of t_{sleep} from 3 sec to 10 sec shows that the smaller the sleep time, the longer the lifetime of the network. As shown, the network with no node selection has a short lifetime with steep energy decay.

5.2.4 Selection Energy Consumption with $r_{negotiate}$

Due to the requirement that the selection algorithm consumes a small amount of energy, we investigate the worst-case energy consumption of the algorithm with

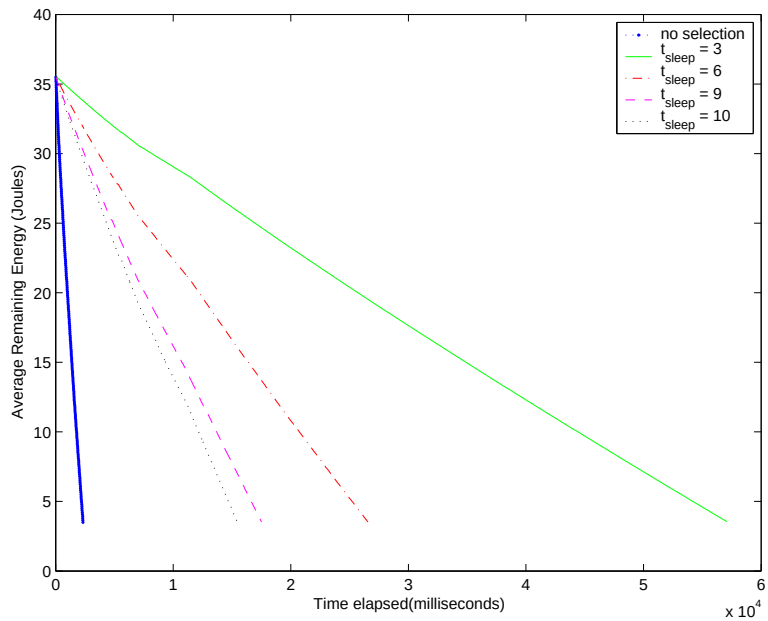


Figure 5.6: Effect of varying the sleep time $r_{negotiate}$= 3m.

variation in $r_{negotiate}$. The number of deployed sensor nodes is maintained at 200 nodes.

Table 5.2 shows that one round of selection consumes a small percentage of available energy with the energy consumption for $r_{negotiate}$ = 6m being only about 16-millionth percent of the available network energy.

5.2.5 Selection Energy Consumption with Network Density

We estimate the worst-case energy consumption of the selection algorithm for different network densities. $r_{negotiate}$ was chosen to be 3m while the number of nodes deployed was varied from 200 to 1000 nodes.

As shown in Table 5.3, the energy consumption for one round of selection only increases by approximately 5.6 mJ when the number of deployed nodes increased

Table 5.2: Total Energy Consumption for one round of selection with $r_{negotiate}$ (200 nodes)

$r_{negotiate}$	Energy(J)	% Energy (x10^{-6})
1	3.8783x10^{-4}	5.4654
2	5.5935x10^{-4}	7.8740
3	7.3216x10^{-4}	10.3037
4	9.2544x10^{-4}	13.0372
5	0.001080	15.2090
6	0.001159	16.3089
7	0.001243	17.4919
8	0.001387	19.5290
9	0.001586	22.3324
10	0.001664	23.4345
11	0.001746	24.5970
12	0.001647	23.2011
13	0.002095	29.5074
14	0.001986	27.9743
15	0.001586	22.3289

Table 5.3: Energy Consumption for one round of selection with different Node Densities

# of Nodes	Energy (x10^{-4}(J))	% Energy (x10^{-5})
200	7.3376	1.0331
250	10.4879	1.1826
300	13.2224	1.2408
400	19.7472	1.3902
500	26.1112	1.4713
1000	63.0880	1.7771

by a factor of 5. This further illustrates that the selection algorithm will typically consume a small amount of energy.

5.2.6 Sensing Redundancy

It is often desirable to have a certain amount of redundancy in a sensor network in order to achieve robust sensing of the target or phenomenon. The algorithm presented exploits the redundancy present in a densely deployed sensor network to extend network lifetime. However, varying amounts of redundancy are retained depending on the value of $r_{negotiate}$ chosen. The levels of redundancy left after node selection are analyzed here through simulation.

First the coverage of the network is assumed to be the ability of the network to sense the whole field at least once (1-Coverage). Other coverage levels may be needed depending on the type of application. The 1-Coverage used in this simulation is only for illustration purposes. The coverage of the network is displayed using coverage maps constructed as follows.

For simplicity, each sensor is assumed to have a circular sensing area with a radius r. A pixel value representing a section of the region of interest is incremented by 1 for each sensor whose sensing radius reaches that point. This method is used in computer vision where it is known as *accumulation*. Thus, a pixel value of 4 implies that the point under consideration is covered by 4 sensors.

In order to estimate sensing redundancy, each coverage map is characterized by three parameters.

- **Absolute Redundancy:** This is the number of pixels in the coverage map that are covered by more sensors than the coverage value required by the

application. In this simulation, only one sensor is required to cover each point. A point covered by 2 sensors has an absolute redundancy of 1. This is sensitive to the resolution of the coverage map (image) used.

- **Relative Redundancy:** Relative redundancy is the absolute redundancy of a given coverage map divided by the absolute redundancy of a redundantly covered rectangular region (equal to the size of the image). The value of relative redundancy shows the percentage of the rectangular region that is covered by the selected subset [61],[54]. It does not give information about the contiguity of the coverage. This is not sensitive to the resolution of the coverage map.

- **Coverage Contiguity:** Shows whether the sensor subset creates a contiguous coverage region or leaves holes in the covered area. Coverage is achieved if the covered region is contiguous even if there are uncovered regions around the edge of the coverage map. Otherwise coverage is not achieved.

A relative redundancy of 1 implies that the chosen subset covers the rectangular region completely. Whereas a relative redundancy of 0.9 shows that 90% of the rectangular region is covered but does not show if the coverage is connected. Coverage helps determine if the coverage offered by the chosen subset is connected without any breaks. Table 5.4 shows the redundancy parameters for different values of $r_{negotiate}$.

The resolution of the coverage map used in this case is 256x256 which represents the field of size 50m x 50m.

Table 5.4: Redundancy parameters for sensing Coverage

$r_{negotiate}$	Absolute Redundancy	Relative Redundancy	Coverage Contiguity
1	65536	1	Yes
2	65536	1	Yes
3	65536	1	Yes
4	65535	0.9999	Yes
5	64245	0.9803	Yes
6	62038	0.9466	Yes
7	52799	0.8056	Yes
8	48028	0.7328	No
9	26575	0.4055	No
10	41355	0.6310	Yes
11	27873	0.4253	No
12	16268	0.2482	No
13	2842	0.0434	No
14	7836	0.1196	No
15	13348	0.2037	No

5.2.7 Uniformity of Coverage

The spatial selection algorithm designed in this work has the potential of providing uniform sensing of the environment. The uniformity of the resulting coverage after node subset selection is analyzed in this section. The analysis uses the coverage maps developed in Section 5.2.6. The uniformity of coverage is characterized by the mean value of the coverage, the variance of the coverage and a coordinate pair called the *spatial uniformity* (S.U.) and defined as follows:

$$S.U. = \sum_{i=1}^{N}(p_i * [x, y]) / \sum_{i=1}^{N}(p_i) \tag{5.3}$$

where p_i is the i-th pixel value, $[x, y]$ is the vector coordinate of the pixel and

N is the number of pixels. The spatial uniformity can be viewed as the centroid of the distribution of coverage values. The uniformity based on Eq. 5.3 is interpreted relative to the centroid of a spatially flat coverage level thresholded at the desired minimum coverage (in this case 1-Coverage). The position of the spatial uniformity relative to the center of the coverage map gives an idea of the flatness of the spatial distribution of coverage values. Table 5.5 shows the values of mean, variance and spatial uniformity of the resulting node selection for different values of $r_{negotiate}$.

The parameters were obtained for an image of size 256x256 with a centroid of (128.5,128.5). The image represents a field of size 50m x 50m.

Table 5.5: Parameters for uniformity of Coverage

$r_{negotiate}$	Mean	Variance	S.U.(x,y)
1	27.8617	76.5582	125.283,127.74
2	20.802	42.3801	126.293,128.903
3	12.712	12.3849	124.165,133.553
4	8.69939	6.99783	125.307,125.628
5	4.71587	2.60757	121.881,127.023
6	3.65266	1.59517	127.284,125.781
7	2.71509	1.73647	134.004,138.24
8	2.22528	1.04906	126.032,118.003
9	1.39973	0.378039	132.843,116.553
10	1.6996	0.703535	121.154,135.159
11	1.36336	0.549246	119.408,142.814
12	1.12001	0.39474	124.192,128.827
13	0.602402	0.326421	116.891,136.338
14	0.993652	0.246504	132.47,120.651
15	0.982742	0.487317	118.194,112.944

The progressive reduction in mean of coverage with increasing $r_{negotiate}$ is not surprising since the number of selected sensors is also decreasing. The same argument holds for the variance. The spatial uniformity values are close to the center coordinates of the chosen image (128.5, 128.5) which confirms that the distribution is fairly even over the image. However, more information can be derived from this simulation if it is viewed together with Table 5.4, since a good spatial uniformity does not necessarily imply coverage.

To further illustrate the potential of the algorithm to provide uniform coverage the spatial uniformity values are plotted. The true centroid of a flat distribution of coverage values is shown as a circle in Figure 5.7. Coverage maps for this simulation are shown in Appendix A. The results obtained regarding coverage in this section apply similarly to connectivity if a circular transmission region is assumed for each sensor node.

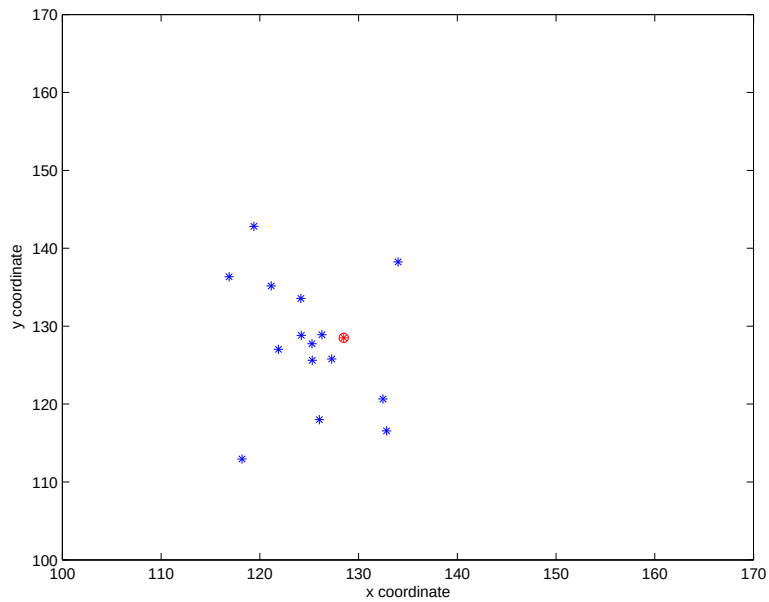


Figure 5.7: Cluster of spatial uniformity values for different $r_{negotiate}$

5.2.8 Fault Tolerance

Most sensor network environments are inhospitable and node failures are common. In order to evaluate the robustness of the spatial selection algorithm to node failures, a random percentage of the nodes are assumed faulty. The energy, coverage and spatial distribution characteristics of the sensor network are discussed for a given value of $r_{negotiate}$. Note that faulty nodes are effectively removed from the network configuration and do not take part in the selection process. Simulation results show that the trend is similar to that of a network with no faulty nodes except for the generally lower level of coverage. For illustration purposes $r_{negotiate}$ is chosen as 3m.

As shown by Figures 5.8 and 5.9, the spatial selection algorithm still ensures that the selected subset maintains spatial spread despite node failures. Table 5.6 shows the number of nodes in ACTIVE state and the percentage of failed nodes.

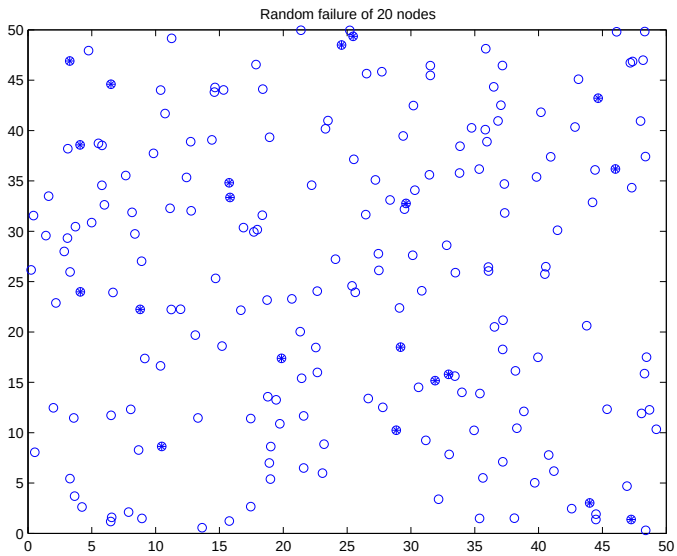


Figure 5.8: Positions of 20 randomly failed nodes

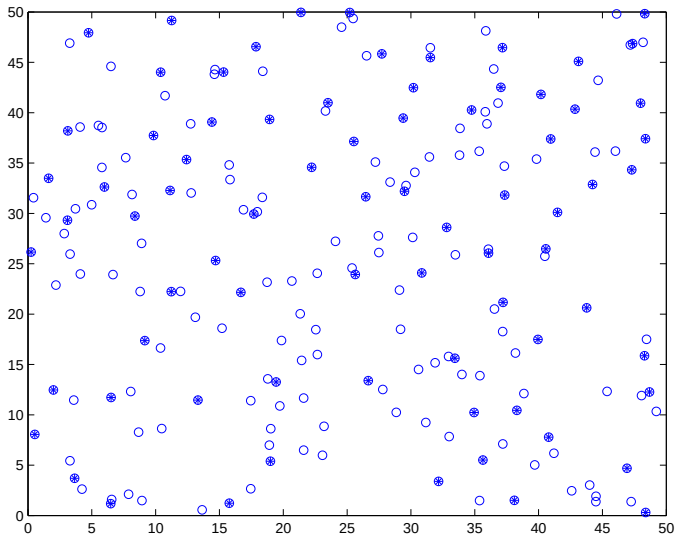


Figure 5.9: Positions of selected sensors when 10% nodes failed

Table 5.6: Number of selected nodes with Node Failures ($r_{negotiate} = 3m$)

%Faulty Nodes	Number of Selected Nodes
5	81
10	76
15	77
20	73
25	61
30	62
35	55
40	51
45	46
50	40

Next the sensing redundancy provided by the network is analyzed as shown in Table 5.7. It is seen that the network continues to achieve coverage when up to 50 percent of the nodes have failed. The level at which coverage is lost will depend on the value of $r_{negotiate}$ chosen and the original density of the network. The size of the coverage map is 256 x 256.

To evaluate the uniformity of the coverage provided by the selected subset, each point was accumulated and the resulting map analyzed as discussed in Sec. 5.2.6. Table 5.8 shows the results obtained for uniformity of coverage with different percentages of faulty nodes. The distribution of the spatial uniformity values is close enough to (128.5,128.5). This shows that the algorithm has the potential to provide uniform coverage despite node failures.

Table 5.7: Redundancy parameters for sensing Coverage ($r_{negotiate} = 3m$)

%Faulty Nodes	Absolute Redundancy	Relative Redundancy	Coverage Contiguity
5	65536	1	Yes
10	65536	1	Yes
15	65451	0.998703	Yes
20	65536	1	Yes
25	65450	0.9987	Yes
30	65450	0.9987	Yes
35	65450	0.9987	Yes
40	65170	0.9944	Yes
45	64981	0.9915	Yes
50	63830	0.9739	Yes

Table 5.8: Spatial parameters for uniformity of Coverage

%Faulty Nodes	Mean	Variance	S.U.(x,y)
5	12.2305	10.8605	131.531,131.189
10	11.4974	14.736	132.492,141.123
15	11.7976	15.113	134.416,135.057
20	11.318	10.5615	129.736,134.103
25	9.2429	6.4707	133.709,132.819
30	9.6371	7.8139	130.975,129.541
35	8.4780	7.0389	133.755, 130.684
40	7.6728	5.9639	132.346, 132.289
45	6.9538	5.4385	134.851, 133.967
50	6.37288	6.7542	122.332,134.523

5.3 Detection Performance

In this section, the detection performance of detectors based on the selection of data in sensor networks are considered. This is motivated by the fact that most uses of sensor networks require that a signal processing task be carried out. The comments provided here are based on the analysis of [49] for a similar data selection scenario. We comment on the impact of selecting a subset of sensed data on detection performance with an illustrative example.

In general, detection performance is depicted using the Receiver Operating Characteristic (ROC). We determine the ROC curves for an example binary detection problem to illustrate the impact of the size of data used in detection on detection performance. This example uses the general gaussian problem discussed in [58]. Under a binary hypothesis such as that discussed in Section 2.2, let

$$H_0 : x_i = n_i \tag{5.4}$$

$$H_1 : x_i = s_i + n_i \tag{5.5}$$

Assume that a detector receives an N-sample vector obtained from N-sensors denoted under each hypothesis of the binary detection problem as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}$$

with mean vector $\mu_{\mathbf{1}}$ under $\mathbf{H}_1$ and $\mu_{\mathbf{0}}$ under $\mathbf{H}_0$. The covariance matrix under each hypothesis can be defined respectively as $\mathbf{K}_1$ and $\mathbf{K}_0$. Let the inverse of each $\mathbf{K}$ be respresented by $\mathbf{Q}$ and $\mathbf{I}$ be an identity matrix. Then the following relationships hold for $\mathbf{H}_1$

$$E[\mathbf{x}|\mathbf{H_1}] = \mu_{\mathbf{1}} \tag{5.6}$$

$$\mathbf{K}_1 = E[(\mathbf{x} - \mu_1)(\mathbf{x}^T - \mu_1^T)|H_1 \tag{5.7}$$

$$\mathbf{Q}_1 = \mathbf{K}_1^{-1} \tag{5.8}$$

$$\mathbf{Q}_1\mathbf{K}_1 = \mathbf{K}_1\mathbf{Q}_1 = \mathbf{I}, \tag{5.9}$$

Combining Eqs. (5.6),(5.7),(5.8) we can obtain the probability density of the observations under H_0

$$p_{\mathbf{x}|H_1}(\mathbf{x}|H_1) = [(2\pi)^{N/2}|\mathbf{K}_1|^{1/2}]^{-1}exp[-\frac{1}{2}(\mathbf{x}^T - \mu_1^T)\mathbf{Q}_1(\mathbf{x} - \mu_1)] \tag{5.10}$$

Through a similar set of definitions the probability density under H_0 is

$$p_{\mathbf{x}|H_0}(\mathbf{x}|H_0) = [(2\pi)^{N/2}|\mathbf{K}_0|^{1/2}]^{-1}exp[-\frac{1}{2}(\mathbf{x}^T - \mu_0^T)\mathbf{Q}_0(\mathbf{x} - \mu_0)] \tag{5.11}$$

Then the logarithm of the likelihood ratio test can be written as [58]:

$$\frac{1}{2}(\mathbf{x}^T - \mu_0^T)\mathbf{Q}_0(\mathbf{x} - \mu_0) - \frac{1}{2}(\mathbf{x}^T - \mu_1) \gtrless \gamma* \tag{5.12}$$

where γ^* represents $\ln\eta + \frac{1}{2}\ln|\mathbf{K}_1| - \frac{1}{2}\ln|\mathbf{K}_0|$

The result given in Eq. 5.12 is the difference between two *quadratic forms*. It can conveniently be assumed that the observed vector has equal mean under both hypothesis but possibly different variances, then Eq. 5.12 becomes

$$1/2(\mathbf{R^T} - \mu^{\mathbf{T}})(\mathbf{Q_0} - \mathbf{Q_1})(\mathbf{R} - \mu) \gtrless \ln\eta + \mathbf{1/2}\ln\frac{|\mathbf{K_1}|}{|\mathbf{K_0}|} = \gamma^*. \tag{5.13}$$

Let the difference in covariance matrix be $\Delta\mathbf{Q} = \mathbf{Q}_0 - \mathbf{Q}_1$. The likelihood ratio test can be written as

$$l(\mathbf{x}) = \mathbf{x}^T \Delta\mathbf{Q}\mathbf{x} \gtrless 2\gamma^* = \gamma' \tag{5.14}$$

Consider a case when the covariance matrix under $\mathbf{H}_0$ is diagonal and the variances are equal, but the signal components are correlated.

$$\mathbf{K}_s = \sigma_s^2 \mathbf{I} \tag{5.15}$$

$$\mathbf{H} = (\sigma_n^2 \mathbf{I} + \sigma_s^2 \mathbf{I})^{-1} \sigma_s^2 \mathbf{I} \tag{5.16}$$

The likelihood ratio becomes

$$l(\mathbf{x}) = \frac{1}{\sigma_n^2} \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2} \mathbf{x}^T \mathbf{x} = \frac{1}{\sigma_n^2} \frac{\sigma_s^2}{\sigma_n^2 + \sigma_s^2} \sum_{i=1}^{N} x_i^2. \tag{5.17}$$

absorbing the constant into the right hand side leads to

$$l(\mathbf{x}) = \sum_{i=1}^{N} x_i^2 \gtrless \gamma'' \tag{5.18}$$

In order to calculate the performance of the test, let the number of data samples N be an even integer. Then the probability density of l can be written as

$$p_{l|H_0}(L|H_0) = \frac{L^{N/2-1} e^{-L/2\sigma_n^2}}{2^{N/2} \sigma_n^N \Gamma(N/2)}, \; L \geq 0, \tag{5.19}$$

$$= 0, \; L < 0, \tag{5.20}$$

and similarly under H_1

$$p_{l|H_1}(L|H_1) = \frac{L^{N/2-1}e^{-L/2\sigma_1^2}}{2^{N/2}\sigma_1^N\Gamma(N/2)},\ L \geq 0, \tag{5.21}$$

$$= 0,\ L < 0, \tag{5.22}$$

where $\sigma_1^2 = \sigma_s^2 + \sigma_n^2$.

Thus, the expressions for P_D and P_F are,

$$P_D = \int_{\gamma''}^{\infty} [2^{N/2}\sigma_1^N\Gamma(N/2)]^{-1}L^{N/2-1}e^{-L/2\sigma_1^2}dL \tag{5.23}$$

and

$$P_F = \int_{\gamma''}^{\infty} [2^{N/2}\sigma_n^N\Gamma(N/2)]^{-1}L^{N/2-1}e^{-L/2\sigma_n^2}dL \tag{5.24}$$

To plot the ROC there is a need to evaluate the two integrals in Eqs. 5.23 and 5.24. Substituting $M = N/2 - 1$ and $\gamma''' = \gamma''/2\sigma_n^2$ then P_F can be written as

$$P_F = 1 - \int_0^{\gamma'''} \frac{x^M}{M!}exp(-t)dt. \tag{5.25}$$

Eq. 5.25 can be solved using the incomplete Gamma function [58]. However, a second approach integrates the expression M times as shown below.

$$P_F = exp(-\gamma''')\sum_{k=0}^{M}\frac{(\gamma''')^k}{k!}. \tag{5.26}$$

a similar expression can be obtained for P_D using $\gamma^{iv} = \gamma''/2\sigma_1^2$.

$$P_D = exp(-\gamma^{iv}) \sum_{k=0}^{M} \frac{(\gamma^{iv})^k}{k!}. \tag{5.27}$$

Figure 5.10 shows some typical ROC curves for different numbers of selected data subset (owing to different $r_{negotiate}$). Note that the ROC performance is better for larger values of N. This can be attributed to the fact that the detector has access to more raw data. Hence, in choosing the value of $r_{negotiate}$ the quality of detection should be considered. [49] presents a similar analysis to that presented in this section for designing detectors in the presence of randomization.

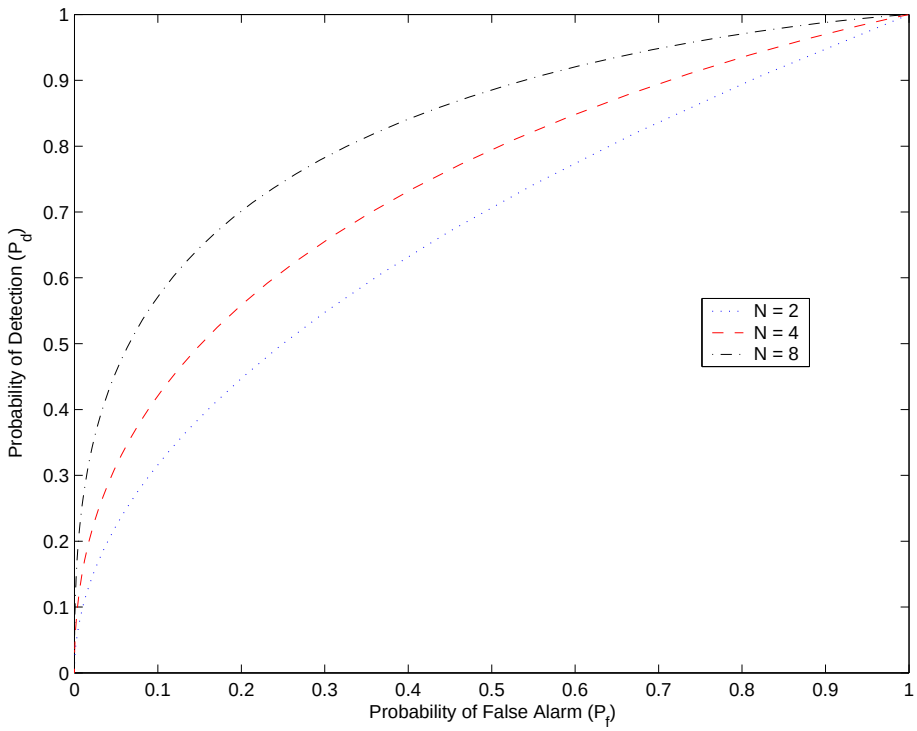


Figure 5.10: ROC curves showing detection performance for different values of N and $\sigma_s^2/\sigma_n^2 = 1$

Using spatial selection in a dense network, the signals captured may be more representative of the target since the selected sensors are distributed all over the deployed area rather than from a skewed selection.

The analysis provided above does not account for the spatial uniformity of the data samples or the position of the target relative to the samples. Spatial orientation of the data may have an impact on predicting detection performance.

5.4 Simulation Notes

The simulator used to derive the results presented in this thesis was developed using JAVA. It consists of three classes; the Main, the Message and the Node classes. The Main class implements interactions between the Node and Message classes while the Node class implements the characteristics of the node including the first order energy model. The Message class constructs the 32-bit energy packet that is exchanged between nodes.

Chapter 6

Conclusions and Future Work

In this thesis, a spatial selection technique is proposed as a robust strategy to cope with resource challenges in wireless sensor networks. We have shown that the algorithm consistently selects sensor nodes from all over a densely deployed network in a manner that coverage and connectivity are achieved. The selection algorithm was shown to extend the lifetime of a sensor network. In addition, it has a potential to manage bandwidth.

Another advantage of this algorithm is that it is independent of network topology which makes it applicable in both clustered and unclustered networks. The use of this algorithm does not affect other networking protocols proposed for use in sensor networks because the chosen sensor subset can be seen as a newly deployed network on which the algorithms will be run. For example, after deploying a network, nodes may be selected before a clustering algorithm is executed or vice versa. The proposed spatial algorithm is also flexible since it provides free parameters ($r_{negotiate}$ and t_{sleep}) which can be chosen to obtain desirable lifetime

extension, bandwidth fidelity or quality of detection.

As shown by the simulations, the slope of energy decay in the sensor network using spatial selection is smaller for a given value of $r_{negotiate}$ than when no selection is performed. This illustrates that energy utilization in the network is distributed uniformly and shows that the QoS of the network can be maintained for longer.

Although, the variogram method did not result in a precise value for the correlation distance in our signal model example, it has been presented to demonstrate an approach for determining the correlation distance of the network. Investigation into other valid signal models and their correlation properties may lead to definite correlation distance values. Based on the results obtained in this work about spatial selection, the correlation distance may be useful as a metric for node deployment in situations where it is convenient to deterministically place the sensors in a given pattern. This can lead to cost benefits in network design such as the analyses provided in [11].

The method we presented may be extended to cases in which different sensors on a node are selected at different times such as selecting imaging sensors at a given time while acoustic sensors are chosen in a different time slot. Overall, the algorithm outlined in this work seems a simple technique for selecting both node and sensor subsets while utilizing little energy. A careful combination of the parameters offered by the algorithm can lead to a network with desired operating lifetime, coverage and connectivity properties.

Future work may investigate the impact of spatially distributed data selection on the detection accuracy of a target or phenomenon. Practical implementation of

the algorithm in a dense network is needed to validate the claims made about the benefits it offers. In this work, issues such as uneven terrains were not considered but the algorithm can be extended. Extension of this work to consider multiple sensing modalities and different sensor types can widen the application scope of the algorithm. Although the algorithm is well suited for standing sensor networks in which the nodes in the network are always sensing events, it may also be triggered by events like the appearance of a target.

Bibliography

Bibliography

[1] http://robotics.eecs.berkeley.edu/~pister/SmartDust/in2010.

[2] http://www.alertsystems.org.

[3] http://www.bluetooth.org.

[4] http://www.gstat.org.

[5] J. Agre and L. Clare. An integrated architecture for cooperative sensing networks. *IEEE Computer Magazine*, May 2000.

[6] J. R. Agrea, L. P. Clarea, G. J. Pottieb, and N. P. Romanova. Development platform for self-organizing wireless sensor networks. volume 3713, pages 257–268, 1999.

[7] Miller A.J. *Subset Selection in Regression*. Chapman and Hall, New York, 1990.

[8] I.F. Akyildiz, Weilian Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. In *IEEE Communications Magazine*, August 2002.

[9] Douglas M. Blough and Paolo Santi. Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc networks. July.

[10] James E. Burt and Gerald M. Barber. *Elementary Statistics for Geographers*. The Guildford Press, 1996.

[11] Krishendu Chakrabarty, S.Sitharama Iyengar, Hairong Qi, and Eungchun Cho. Grid coverage for surveillance and target location in distributed sensor networks. *IEEE Transactions on Computers*, December 2002.

[12] S. Cho and A. Chandrakasan. Energy-efficient protocols for low duty cycle wireless microsensor networks. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000.

[13] T. Collier and C. E. Taylor. Self-organization in sensor networks. *Journal of Parallel and Distributed Computing*, 2003.

[14] Noel A.C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, 1991.

[15] Ashwin D'Costa, Vinod Ramachandran, and Akbar M. Sayeed. Distributed classification of gaussian space-time sources in wireless sensor networks.

[16] Marco F. Duarte and Yu Hen Hu. Vehicle classification in distributed sensor networks.

[17] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, 2000.

[18] Robert E. Van Dyck. Detection performance in self-organized wireless sensor networks. In *IEEE International Symposium on Information Theory*, June - July 2002.

[19] D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world with wireless sensor networks. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2001.

[20] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In *MobiCom*, 1999.

[21] Matheron G. Principles of geostatistics. *Economic Geology*, 1963.

[22] S. Hedetniemi and A. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 1988.

[23] W. Heinzelmann, J. Kulik, and H. Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th ACM/IEEE MobiCom Conference (MobiCom '99)*, August 1999.

[24] W.R. Heinzelmann, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, January 2000.

[25] A. Hero, H. Messer, J. Goldberg, D.J. Thomson, M.G. Amin, G. Giannakis, A. Swami, J.K. Tugnait, A.L. Nehorai, A.and Swindlehurst, J.F. Cardoso, Lang Tong, and J. Krolik. Highlights of statistical signal and array processing. *IEEE Signal Processing Magazine*, September 1998.

[26] Alfred Hero. Signal detection and classification. *CRC Handbook on Signal Processing*, 1997.

[27] C. Herring and S. Kaplan. Component-based software systems for smart environments. *IEEE Personal Communications*, October 2000.

[28] Jason Hill, Robert Szewczyk, Alec Woo, Seth Hollar, David Culler, and Kristofer Pister. System architecture directions for networked sensors.

[29] Gerard J. Holzmann. *Design and Validation of Computer Protocols*. Prentice Hall, 1991.

[30] Metin Inanc, Malik Magdon-Ismail, and Bulent Yener. Power optimal connectivity and coverage in wireless sensor networks. In *IEEE International Conference on Communications*, April 2002.

[31] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCOM '00)*, August 2000.

[32] James F. Kurose and Keith W. Ross. *Computer Networking; A Top-Down Approach Featuring the Internet*. Addison Wesley, 2002.

[33] D. Li, K. Wong, Y. Hu, and A. Sayeed. Detection, classification, tracking of targets in micro-sensor networks. *IEEE Signal Processing Magazine*, March 2002.

[34] Xiang-Yang Li, Peng-Jun Wan, and Frieder O. Coverage in wireless ad-hoc sensor networks. In *IEEE International Conference on Communications*, April 2002.

[35] S. Lindsey and C.S. Raghavendra. Pegasis: Power-efficient gathering in sensor information systems. In *Aerospace Conference Proceedings*, March 2002.

[36] Hawkins D. M. and Cressie N. Robust kriging - a proposal. *Journal of the International Association for Mathematical Geology*, 1984.

[37] Nafie M., Tewfik A., and Ali M. Deterministic and iterative solutions to subset selection problems. *IEEE Transactions in Signal Processing*, July 2002.

[38] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. Wireless sensor networks for habitat monitoring.

[39] Seaphan Meguerdichian, Farinaz Koushanfar, Miodrag Potkonjak, and Mani B. Srivastava. Coverage problems in wireless ad-hoc sensor networks. *IEEE INFOCOM*, 2001.

[40] Edzer J. Pebesema. Gstat user's manual. Online at www.gstat.org.

[41] E.M. Petriu, N.D. Georganas, D.C. Petriu, D. Makrakis, and V.Z. Groza. Sensor-based information appliances. *IEEE Instrumentation and Measurement Magazine*, December 2000.

[42] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.

[43] G.J. Pottie and W.J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, May 2000.

[44] J.M Rabaey, J. Ammer, J.L. da Silva Jr., and D. Patel. Picoradio: ad-hoc wireless networking of ubiquitous low-energy sensor/monitor nodes. *Proceedings of the IEEE Computer Society Annual Workshop on VLSI (WVLSI'00), Orlando, Florida*, April 2000.

[45] J.M Rabaey, J. Ammer, J.L. da Silva Jr., D. Patel, and S. Roundy. Picoradio supports ad hoc ultra-low power wireless networking. *IEEE Computer Magazine*, 2000.

[46] Vijay Raghunathan, Curt Schurgers, Sung Park, and Mani B. Srivastava. Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, March 2002.

[47] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[48] Brian D. Ripley. *Spatial Statistics*. John Wiley & Sons, 1981.

[49] C.K. Sestok, M.R. Said, and A.V. Oppenheim. Randomized data selection in detection with applications to distributed signal processing. In *Proceedings of the IEEE*, August 2003.

[50] C. Shen, C. Srisathapornphat, and C. Jaikaeo. Sensor information networking architecture and applications. *IEEE Personal Communications*, August 2001.

[51] E. Shih, B.H. Calhoun, S. Cho, and A. Chandrakesan. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. *Proceedings of the ACM MobiCom'01*, July 2001.

[52] E. Shih, S. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. *Proceedings of ACM MobiCom'01, Rome, Italy*, July 2001.

[53] Amit Sinha and Anantha Chandrakasan. Dynamic power management in wireless sensor networks. *Design and Test of Computers, IEEE*, March-April 2001.

[54] Sasa Slijepcevic and Miodrag Potkonjak. Power efficient organization of wireless sensor networks. *IEEE International Conference on Communications(ICC)*, June 2001.

[55] K. Sohrabi, , J. Gao, V. Ailawadhi, and G. J. Pottie. Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, October 2000.

[56] Sameer Tilak, Nael B. Abu-Ghazaleh, and Wendi Heinzelman. Infrastructure tradeoffs for sensor networks. *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, September 2002.

[57] C. K. Toh. *Ad Hoc Mobile Wireless Networks: Protocols and Systems*. Prentice Hall, 2002.

[58] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Detection, Estimation and Linear Modulation Theory.* John Wiley & Sons Inc., New York, 2001.

[59] X. Wang, H. Qi, and S. S. Iyengar. Collaborative multi-modality target classification in distributed sensor networks. In *Information Fusion*, pages 285–290, Jul 2002.

[60] B. Warneke, B. Liebowitz, and K.S.J. Pister. Smart dust: communicating with a cubic-millimeter computer. *IEEE Computer Magazine*, January 2001.

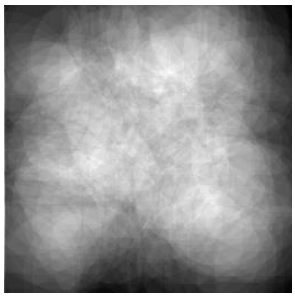
[61] R. Williams. *The Geometrical Foundation of Natural Structure: The Source Book of Design.* Dover, New York, 1979.

[62] Yingyue Xu and Hairong Qi. Performance evaluation of computing paradigms in mobile ad hoc sensor networks. *The 9th International Conference on Parallel and Distributed Systems (ICPADS), Taiwan*, December 2002.

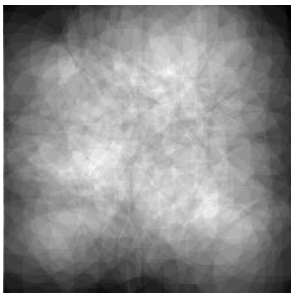
Appendix

Appendix A

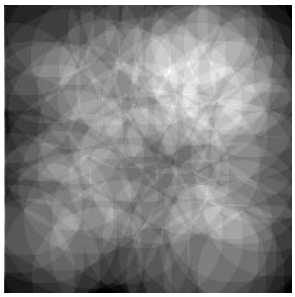
A.1 Coverage Maps - No Faulty Nodes



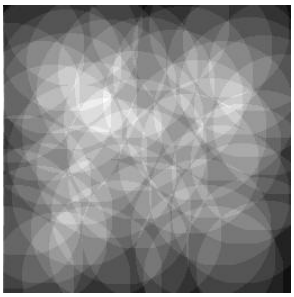
(a) Coverage map for $r_{negotiate} = 1$



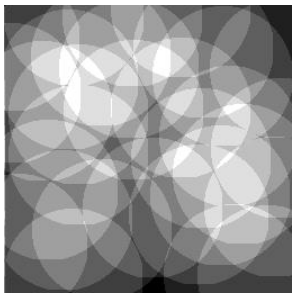
(b) Coverage map for $r_{negotiate} = 2$



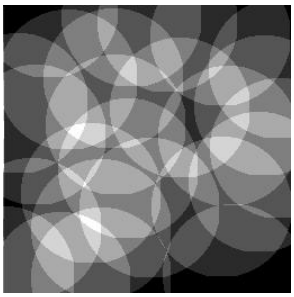
(c) Coverage map for $r_{negotiate} = 3$



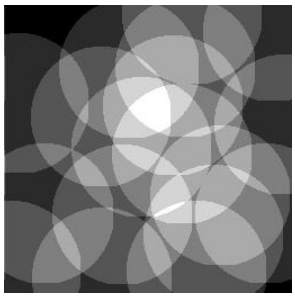
(d) Coverage map for $r_{negotiate} = 4$



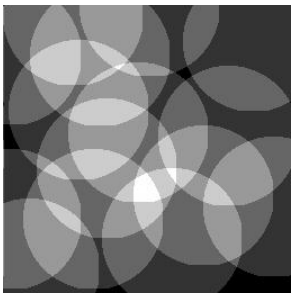
(e) Coverage map for $r_{negotiate} = 5$



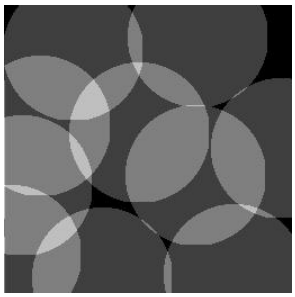
(f) Coverage map for $r_{negotiate} = 6$



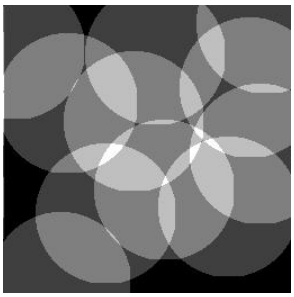
(g) Coverage map for $r_{negotiate} = 7$



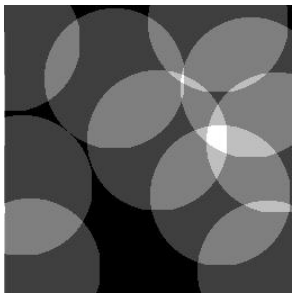
(h) Coverage map for $r_{negotiate} = 8$



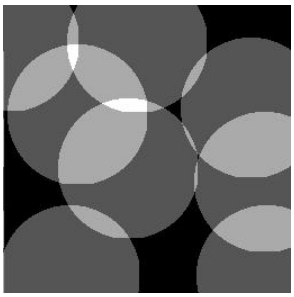
(i) Coverage map for $r_{negotiate} = 9$



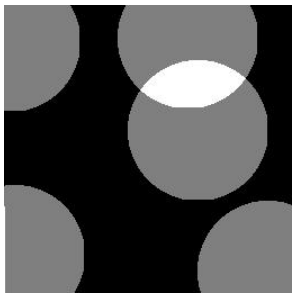
(j) Coverage map for $r_{negotiate} = 10$



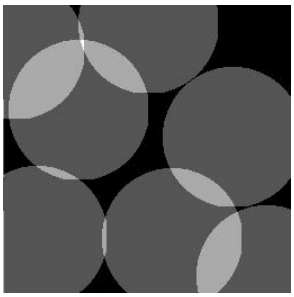
(k) Coverage map for $r_{negotiate} = 11$



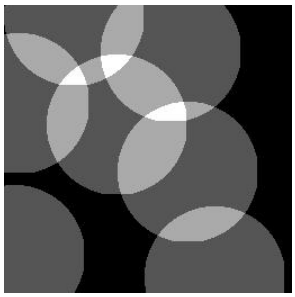
(l) Coverage map for $r_{negotiate} = 12$



(m) Coverage map for $r_{negotiate} = 13$

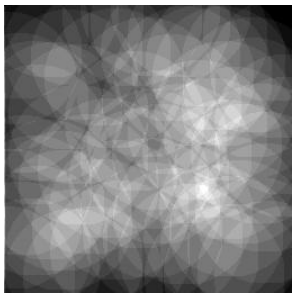


(n) Coverage map for $r_{negotiate} = 14$

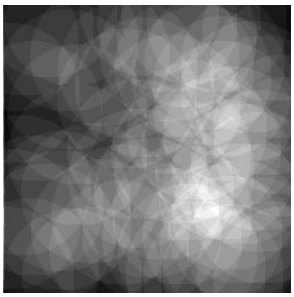


(o) Coverage map for $r_{negotiate} = 15$

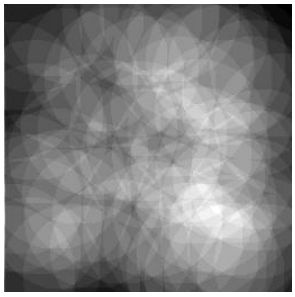
A.2 Coverage Maps - With Faulty Nodes



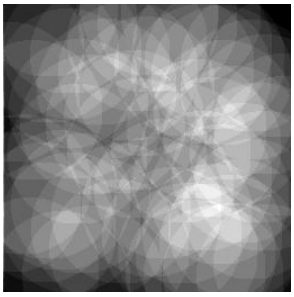
(a) Coverage map for 5% Faulty Nodes



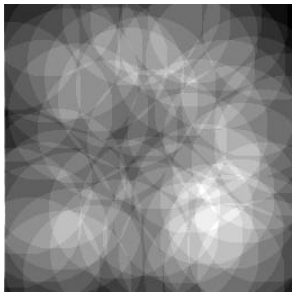
(b) Coverage map for 10% Faulty Nodes



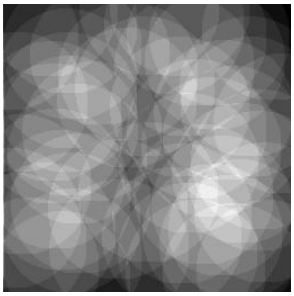
(c) Coverage map for 15% Faulty Nodes



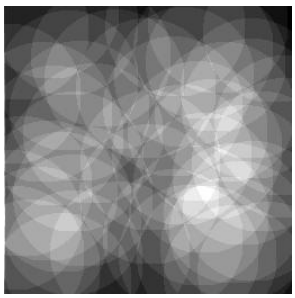
(d) Coverage map for 20% Faulty Nodes



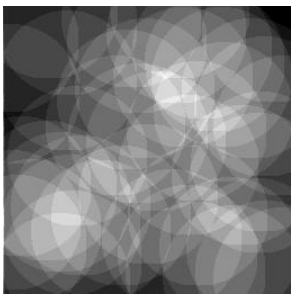
(e) Coverage map for 25% Faulty Nodes



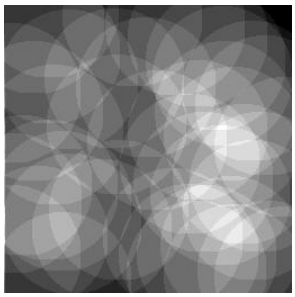
(f) Coverage map for 30% Faulty Nodes



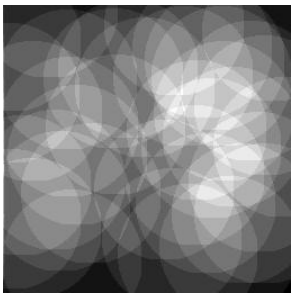
(g) Coverage map for 35% Faulty Nodes



(h) Coverage map for 40% Faulty Nodes



(i) Coverage map for 45% Faulty Nodes



(j) Coverage map for 50% Faulty Nodes

Vita

Olawoye Ayorinde Oyeyele was born in Ilorin, Nigeria in 1975. His major research interests are in the areas of signal and image processing, digital design, and networking. He graduated from Obafemi Awolowo University, Ile-Ife, Nigeria in February 1999 with a Bachelor of Science degree in Electronics and Electrical Engineering. His undergraduate thesis was about the design and implementation of a secure communication system. The thesis involved the design and realization of a Frequency Modulation (FM)/Amplitude Modulation (AM) receiver synchronized to a transmitter at three different frequencies. The transmitter sends AM control signals to the receivers to inform them of a new FM transmit frequency. This way interference and jamming can be avoided in large public arena communication systems.

Prior to graduate study, Olawoye worked for Accenture, a Big 5 Management Consulting firm delivering technology solutions strategically to help clients achieve their corporate goals. At Accenture, he was involved in the development of an Information and Communications Technology (ICT) strategy, the deployment of SAP, an Enterprise Resource Planning (ERP) application and the assessment of different Human Resource Management Software (HRMS) for a client.

In Fall 2002, he came to the University of Tennessee, Knoxville as a graduate student in Electrical and Computer Engineering. He worked in the Advanced Imaging and Collaborative Information Processing (AICIP) laboratory of Dr. Hairong Qi with a research topic in efficient techniques for Wireless Sensor Networks (WSN). He developed a node selection protocol for robustly selecting a representative subset of sensor nodes in massively deployed dense wireless sen-

sor networks. The algorithm developed is detailed in his Master's thesis entitled "A Robust Node Selection Strategy for Lifetime Extension in Wireless Sensor Networks". Olawoye Oyeyele's publications include

Olawoye Oyeyele, Hairong Qi. "A Robust Node Selection Strategy for Lifetime Extension in Wireless Sensor Networks" submitted to the First IEEE Sensor and Ad Hoc Networking Conference (SECON) 2004.