EVENT DISTORTION BASED CLUSTERING ALGORITHM FOR ENERGY HARVESTING WIRELESS SENSOR NETWORKS

A thesis submitted to
The Graduated School of Engineering and Science
Izmir Institute of Technology
In Partial Fulfillment of Requirements for a degree of

MASTER OF ENGINEERING

In Electrical - Electronics Engineering

By
ALİ MUDHEHER RAGHIB KAFI AL-QAMAJI

April 2017
IZMIR

We approve the thesis of **ALİ Mudheher Raghib Kafi AL-QAMAJI**

Examining Committee Members:

Assoc. Prof. Dr. Barış ATAKAN
Department of Electrical - Electronics Engineering
Izmir Institute of Technology

Assoc. Prof. Dr. Radosveta SOKULLU
Department of Electrical and Electronics Engineering
Ege University

Assist. Prof. Dr. Berna ÖZBEK
Department of Electrical - Electronics Engineering
Izmir Institute of Technology

21 April 2017

Assoc. Prof. Dr. Barış ATAKAN
Supervisor, Department of Electrical - Electronics Engineering,
İzmir Institute of Technology

Prof. Dr. Enver TATLICIOĞLU
Head of the Department of
Electrical - Electronics Engineering

Prof. Dr. Aysun SOFUOĞLU
Dean of the Graduate School
of Engineering and Sciences

ACKNOWLEDGMENTS

First of all, I owe a huge deal of thanks to my supervisor Assoc. Prof. Dr. Barış ATAKAN for his excellent guidance, tolerance and patience in every level of my thesis study. Many thanks also for making reasonable amount of time whenever I asked for his support. I would like to thank committee members for giving me contribution which helped me a lot in revising the thesis.

I would like to express my sincere thanks to Assist. Prof. Dr. Berna ÖZBEK and Assoc. Prof. Dr. Mustafa A. ALTINKAYA for their supports, helps and favors at first stages of my master studies. Special thanks to the colleagues and friends from IYTE, namely Serpil Yılmaz, Başak Esin Guzel, and Oktay Karakus for their helpful suggestions and comments during my courses.

Last but definitely not least, I thank my mother and my father a lot for their true and invaluable support. Very special thanks to my brother Ahmed, and my sisters Aseel and Enas who motivated and advised me through some difficult times.

ABSTRACT

EVENT DISTORTION BASED CLUSTERING ALGORITHM FOR ENERGY HARVESTING WIRELESS SENSOR NETWORKS

Wireless Sensor Network (WSN) is a set of inexpensive densely deployed wireless sensor nodes with limited functionalities and scarcity in energies, whose observations are forwarded or relayed by intermediate nodes to the Base Station (BS). In the networks with densely deployed nodes, the observations are likely to be highly correlated in the space domain. This type of correlation is referred as spatial correlation, which produces unfavorable redundant readings causing energy wasting.

In this thesis, the main task is to reduce these nodes that have redundant readings by using a clustering algorithm called Event Based Clustering (EDC) algorithm. The clustering algorithm is based on exploiting the spatial correlation that is used to cluster the sensor nodes. Exploiting spatial correlation is proposed by using Vector Quantization (VQ) with respect to the distortion constraints. Furthermore, this algorithm is applied for energy harvesting sensor nodes. Also, the inessential sensor nodes that have correlated readings are reduced for improving the Energy-Efficacy (EE) with acceptable level of event signal reconstruction distortion at the sink node.

After applying the EDC algorithm, the communication model is changed from single-hop model to two-hop (clustered-network) model. Hence, a theoretical framework of distortion function, i.e., accuracy level, for both single-hop and two-hop communication models is derived. Then, single-hop and two-hop communication models are compared in terms of achieved distortion level, number of alive nodes, and energy consumption for different sizes of event area. Finally, the effects of various harvested energy level on the clustered-network is studied with respect to the same terms.

ÖZET

ENERJİ TOPLAYAN KABLOSUZ SENSÖR AĞLARI İÇİN OLAY BOZULMASINA DAYALI KÜMELENME ALGORİTMASI

Kablosuz Sensör Ağı (KSA), işlevsellikleri sınırlı, enerjileri az, pahalı olmayan, yoğun bir şekilde konuşlandırılmış kablosuz sensör düğümleri setidir. Sensör düğümlerinin gözlemleri, ara düğümler tarafından baz istasyonuna iletilir. Yoğun bir şekilde konuşlandırılmış düğümlerden oluşan ağlarda gözlemler büyük oranda ilintilidir. Uzaysal ilinti olarak adlandırılan bu tür ilintiler, gereksiz algılayıcı okumalarına neden olur ve dolayısıyla enerji israfına neden olur.

Bu tez çalışmasında enerji toplayan sensör ağları için Olay Bozulmasına Dayalı Kümeleme (OBK) algoritması adı verilen bir kümeleme algoritması verilmektedir. OBK algoritması sensörler arasındaki uzaysal ilintiyi kullanarak gereksiz okumalara sebep olan düğümleri azaltmak sureti ile enerji tüketimini azaltmayı amaçlamaktadır. Uzaysal ilintinin kullanılması bozulma kısıtlamalarına göre Vektör Nicemleme (VN) ile gerçeklenmektedir. Sensörlerin baz istasyonuna yaptıkları iletişim için tek hop ve iki hop olmak üzere iki ayrı iletişim modeli düşünülerek bu modellerin her biri için iki ayrı olay bozulma fonksiyonu türetilmektedir.

Baz istasyonu OBK algoritması ile türetilen bu bozulma fonksiyonlarını kullanarak ilintili okumalara sahip sensör düğümlerinin veri iletimini kontrol edebilmektedir. Bu kontrol sayesinde hangi sensörün iletim yapacağına karar verilerek belli bir olay bozulma kısıtlaması karşılanırken enerji-etkin iletişim de sağlanabilmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Wireless Sensor Network

Wireless Sensor Network (WSN) is a set of sensor nodes that are connected with each other through wireless and noisy channel. Each sensor node requires to sense for physical phenomena (e.g., field source, fixed and random point sources). Then, each sensor node forwards its observations to the Base Station (BS) to be processed or viewed by the end user as in Fig 1.1. However, that's not the whole story, there are many challenges which need to be addressed by the developers and engineers in WSNs such as fault tolerance, scalability, production costs, hardware constraints, communication and power consumption [1].

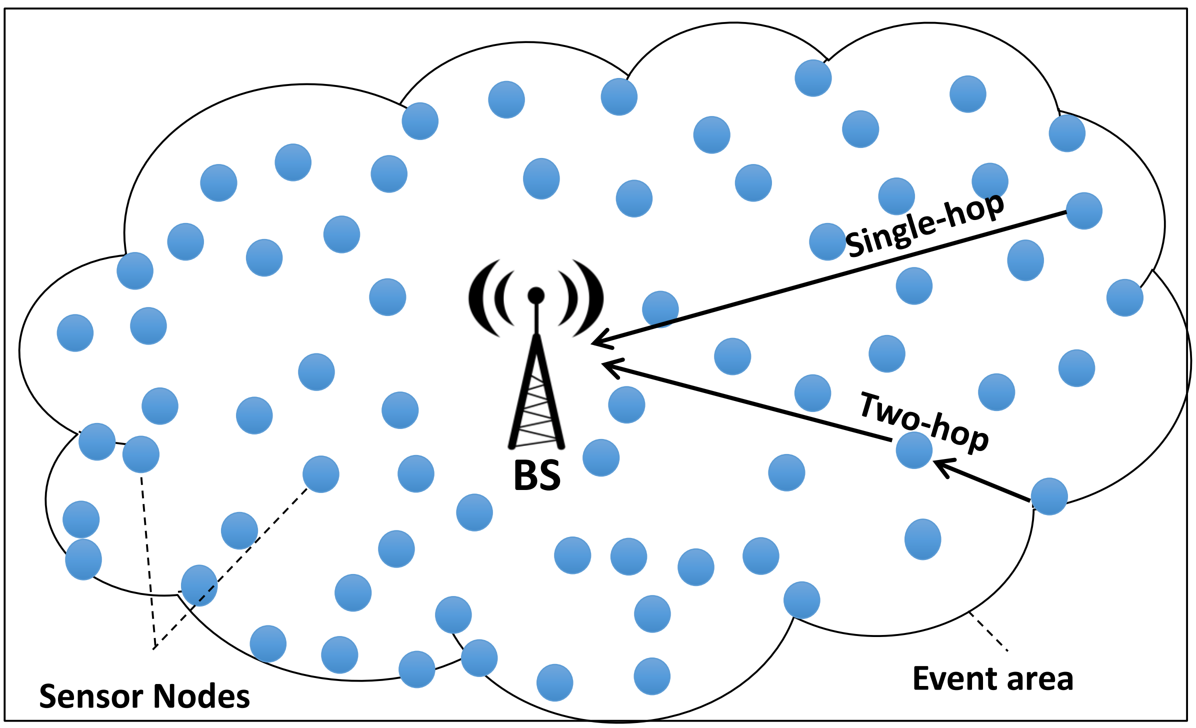


Figure 1.1 Wireless Sensor Network

In WSNs, nodes' deployment can be random or systematic (fixed topologies) determined by the designer or application [1]. Random deployment has an advantage of being less expensive than fixed deployment. It enables WSNs to be applied in hostile and unreachable environment such as battlefield and under water. As a result, the random deployment is more general case but there is a coverage problem (i.e., it's not guaranteed that sensors will cover the whole event area) [1]. Hence, this can be solved by increasing nodes density in an event area, i.e., increasing the number of sensors. Since the communication among sensor nodes is one of the main source of energy consumption, communication protocols are devised for reducing energy consumption.

In WSNs, it is possible to exploit spatial correlation in sensor readings to improve energy efficiency. In densely deployment case, nearby sensor nodes get similar observations [3]. By keeping the data distortion below a predefined threshold, sensor nodes can exploit spatial correlation to reduce amount of data forwarded to the sink node. This reduction will improve Energy-Efficiency (EE) and preserve the distortion level, i.e., without changing the accuracy of received data at the sink node.

One of the main challenges of communication protocols and applications in WSNs is energy saving [1]. The battery-powered sensor nodes need to be charged regularly. Reaching operation is expensive, on addition, some event areas are unreachable. The developers addressed this issue by equipping energy harvesting unit to each sensor node [2]. For harvesting sensor networks, there are various sources are available to be harvested, e.g., Solar, Vibration (motion), Winds, Thermal and Radio frequency (RF) Signals, etc. Each of these harvesting sources provides a different level of harvesting energy as shown in Table 1.1. In this thesis, RF-harvesting will be used for its inexpensive implementation and availability.

1.2.Types and Applications

In various scenarios sensor nodes are required due to their capability to observe a different type of event sources, e.g., humidity, temperature, and lighting, etc. Based on the applications, different types of WSNs are needed:

- **Terrestrial WSN**: Typically, it consists of hundreds to larger number of low-cost wireless sensor nodes deployed in event area, either in an ad hoc (unstructured) or in a pre-planned (structured) manner. In ad hoc deployment, sensor nodes can be randomly placed into the target area by being dropped from a plane. For example, deploying a WSN on an active volcano [17]. Furthermore, some event areas may remain uncovered (coverage issue), thus nodes density is required and thus, results into difficulties in maintenance (e.g., recharging, communication, etc.). In pre-planned deployment, all or some sensor nodes are deployed based on fixed planned architecture, for example grid and optimal [25] placement models. The advantage of pre-planned deployment is that fewer nodes can be deployed, which makes the maintenance less cost than ad hoc deployment case.
- **Undergrounded WSN** [26] [27] [16]: set of sensor nodes are buried underground either in a cave or mine and are used to monitor underground conditions. Furthermore, sink nodes are placed above ground as relays, for taking readings from underground sensor nodes to the base station. With respect to maintenance, equipment and deployment, underground WSN are more expensive than terrestrial WSN, where suitable equipment parts are needed to ensure reliable communication through soil, rocks, water, and other mineral contents to reduce the signal losses and attenuation level. Before sensor nodes deploying for underground WSN, it requires careful planning of energy and cost considerations. Underground sensors are powered with a limited battery power and it is difficult to recharge or replace their batteries.

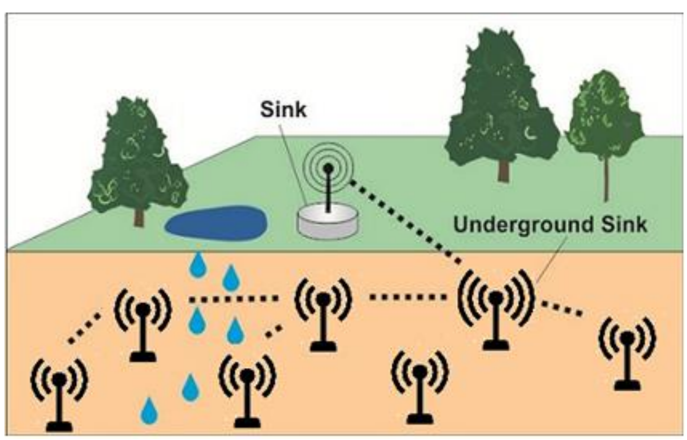


Figure 1.2 Undergrounded Wireless Sensor Network

- **Underwater WSN** [28] [29]: A set of number of sensor nodes or vehicles are placed underwater that provides great opportunity to discover oceans as to improve the studies in environmental issues (Fig 1.3), e.g., marine life in the oceans climate, variations of climate and population of coral reefs, etc. As compared to terrestrial WSNs, less nodes are needed but more expensive to be sank underwater. Acoustic waves are used for wireless communications between underwater sensor nodes. Generally, the main challenges in underwater acoustic communication are bandwidth limitation, propagation delay, and signal fading problems.

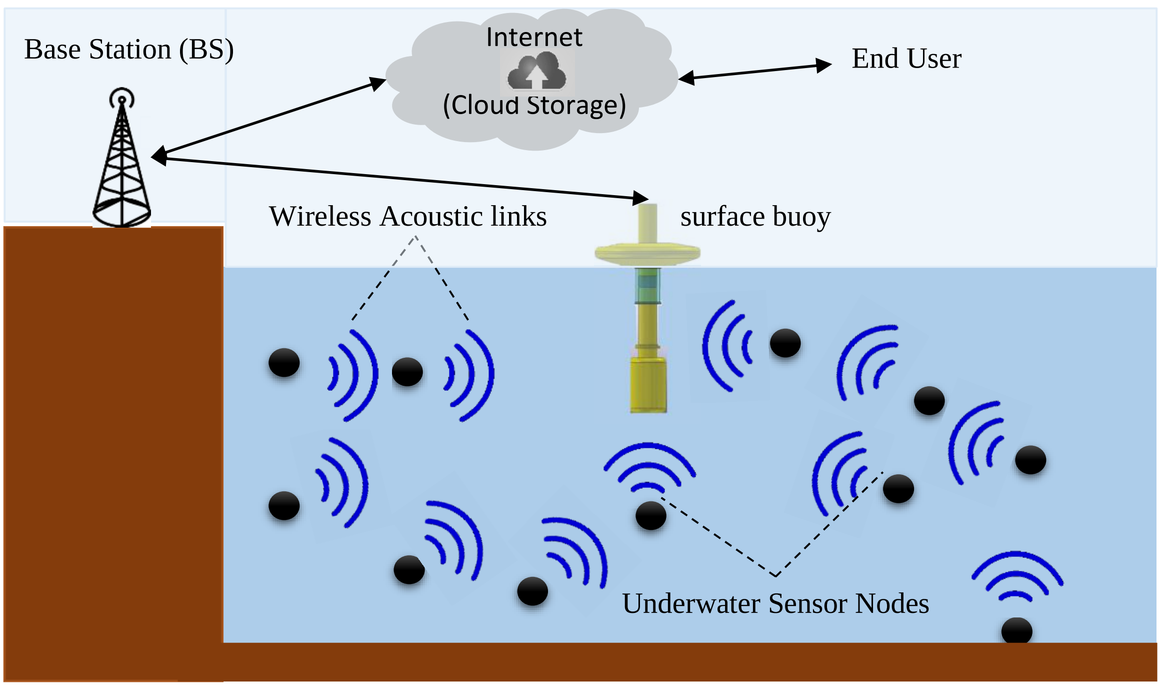


Figure 1.3 Underwater Wireless Sensor Network

- **Multi-media WSN** [30] [18]: A set of inexpensive sensor nodes are equipped with cameras and microphones for monitoring multimedia events, such as videos, audios, and images (Fig 1.4). These sensors are placed in structured manner to ensure the coverage. Furthermore, multi-media WSN requires high bandwidth to deliver the media content, e.g., video stream, that requires high data rate. Hence, it causes high energy consumption rate. Besides, it is important that a certain level of Quality of Service (QoS) is achieved for delivering data reliably.

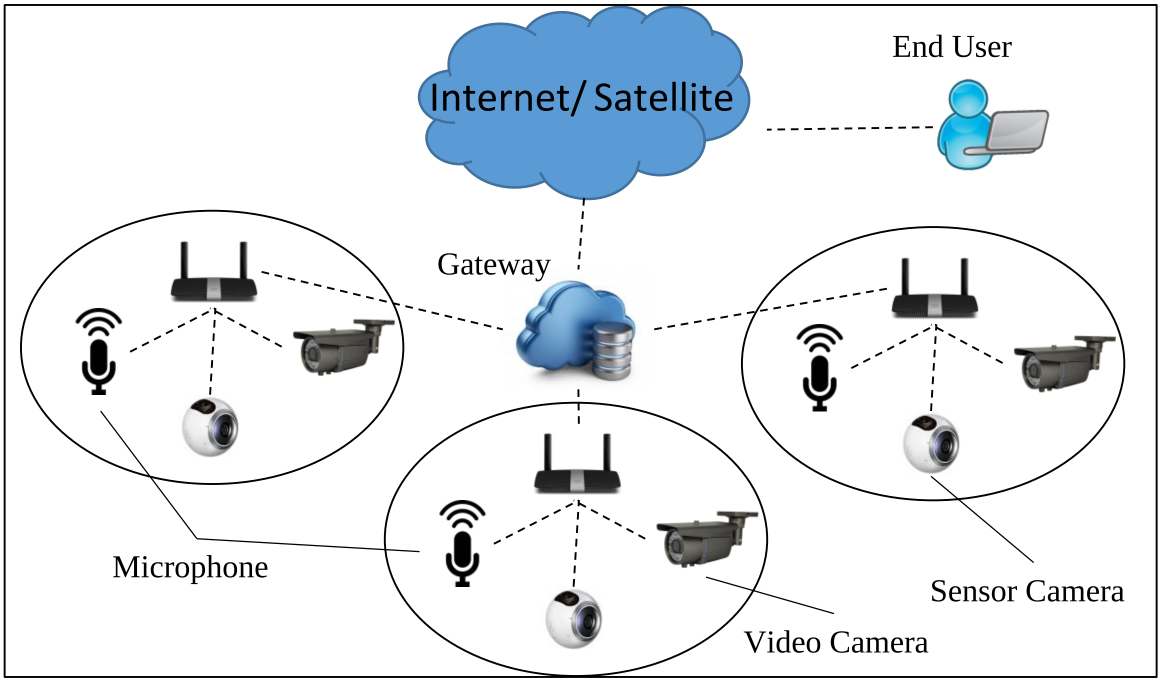


Figure 1.4 Multimedia Wireless Sensor Network

- **Mobile WSN** [15] [19]: A set of dynamic sensor nodes are capable to sense, compute, and communicate like static nodes, i.e., sensor nodes with fixed locations. However, there are two main differences with static WSN. Firstly, dynamic sensor nodes are capable of repositioning and organize themselves. Since they are able to move from one place to another, the higher degree of coverage and connectivity can be achieved. Mobile sensor nodes can deliver gathered information to nearby mobile nodes. Secondly, there is a difference in data distribution. In a static WSN, data can be distributed using fixed routing, while mobile sensor nodes are using a dynamic routing. As a result, the issues in mobile sensor nodes include nodes deployment, localization, self-organization, navigation, control, coverage, energy, maintenance, and data processing, etc.

- Mainly WSNs applications (Fig 1.5) are bounded in civil applications (e.g., alarm, medical, and home etc.) and military applications [1]: -
  - **Environmental applications** [21]: These sensor networks have many different applications in the environment. They can be used for the wildlife tracking

applications, earth quack monitoring, soil observations, atmosphere context, irrigation monitoring, natural disasters and almost agricultural issues can be addressed by sensor nodes.

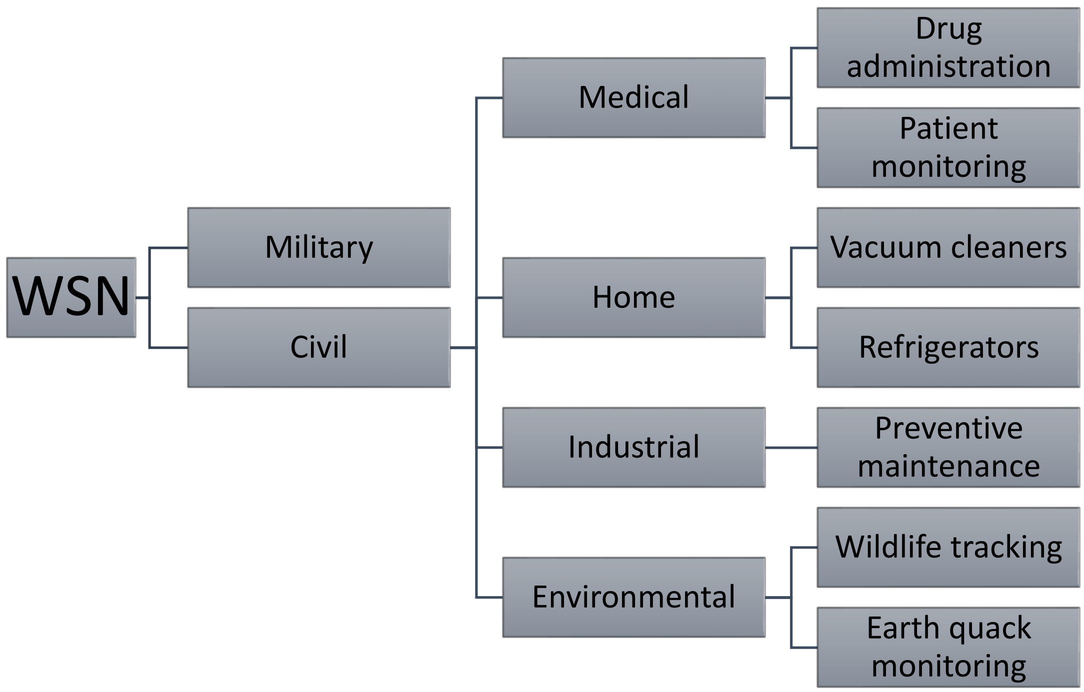


Figure 1.5 Applications of Wireless Sensor Network

- **Medical applications**: The advances in integrated biomedical devices and smart sensor nodes, have attracted researchers and engineers to exploit sensor networks in biomedical field. Some of the medical applications for WSNs are provided for diagnostics, monitoring of disabled, patient, drug administration in hospitals, movements of insects or other animals, doctors inside hospital, and emergency response [31].
- **Military applications** [23]: Rapid deployment and self-organizing aspects that are made by WSNs are very useful for sensing and observing friendly or hostile movements in military operations (Fig 1.6). Furthermore, sensor nodes are capable to detect chemical, nuclear, biological attacks, reconnaissance of opposing forces

and terrain in C4ISR systems (i.e., Command, control, communications, computers, intelligence, surveillance, and reconnaissance), and battlefield surveillance, etc.

- **Home applications** [20]: As the technology is developing, smart sensor nodes can be equipped in home devices, such as vacuum cleaners, micro-wave ovens, refrigerators, and VCRs. These nodes allow users to control home devices locally and remotely by the capability of these nodes to be connected with each other and with external networks by the internet or satellite.

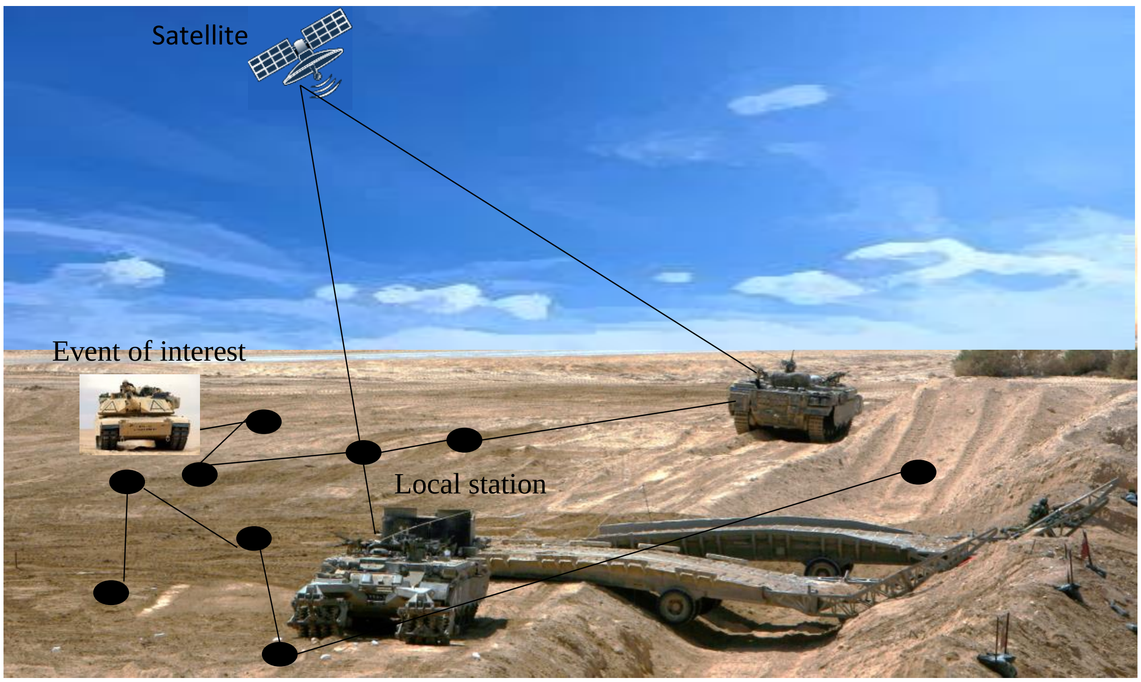


Figure 1.6 Military Applications for Wireless Sensor Networks

- **Industrial applications** [22]: Wired or wireless sensor networks can be used for industrial fields, such as industrial sensing and control applications, building automation, monitoring material fatigue, smart structures with embedded sensor nodes and access control etc. Sensor based monitoring systems and manual monitoring systems are usable in industrial area for preventive maintenance. Wired sensors in such a systems require high cost for upgrading, deployment and require personal supervision.

1.3.Sensor Node Components

The main components [1] of any typical sensor node have been illustrated in Fig 1.7. They are: processer, transceiver, memory, power source and sensing unit. All components of a sensor node have different functions as defined below:

- Processer (micro-controller) unit: It's a low power consumption unit responsible for processing data (e.g., modulation and signal processing, etc.) and controlling other units to perform specific tasks at any given time.

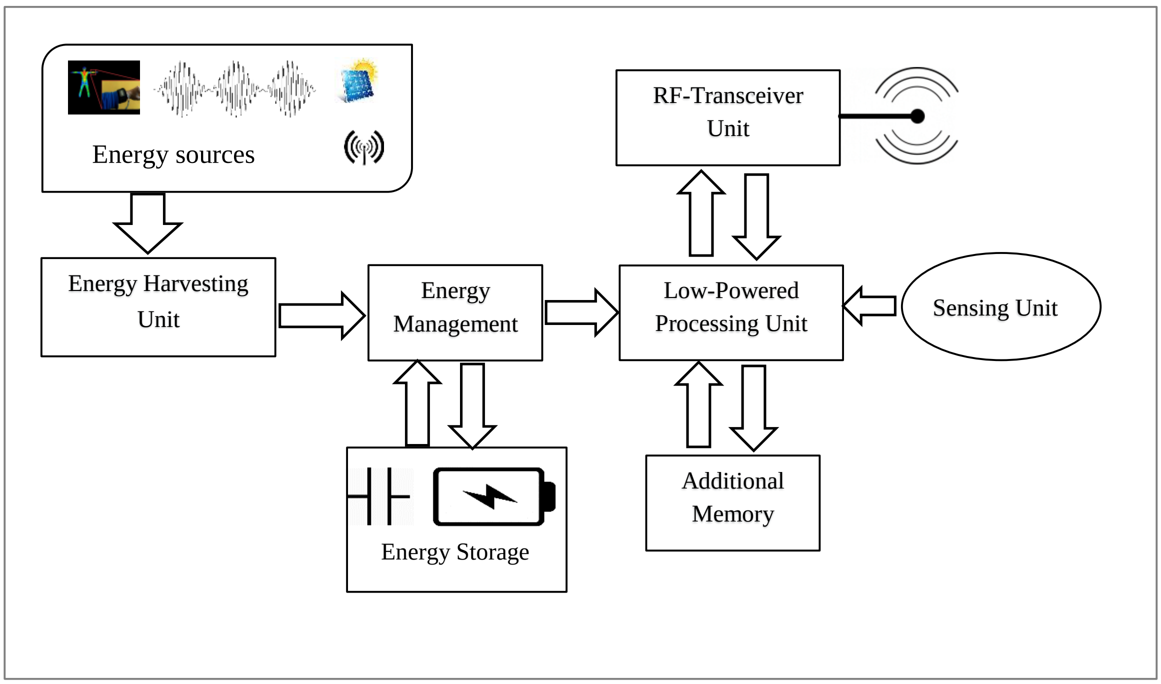


Figure 1.7 Components of Sensor Node

- Transceiver unit: Wireless sensors use radio frequency media for transmission and receiving, by allocating specific frequencies with four different states; transmit, receive, idle, and sleep. This unit is defined as the most unit consuming power than other units because of its antenna.

- Memory unit: Mainly it's used for saving data before transmitting and after receiving.
- Power unit: It can be capacitor or battery (rechargeable or non-rechargeable). It is responsible for powering other units.
- Sensing unit: This unit is used for tracking or observing the physical or environment events.
- External unit: It can be another sensor node, transceiver unit, power unit or harvesting unit.

Table 1.1. Examples of Energy Harvesting Sources [2]

Source	Environmental location	Source Power	Harvested Power
Light	Indoor	$0.1\ mW/cm^2$	$10\ \mu W/cm^2$
	Outdoor	$100\ mW/cm^2$	$10\ mW/cm^2$
Vibration/Motion	Human	$0.5\ m\ at\ 1\ Hz$	
		$1\ m/s^2\ at\ 50\ Hz$	$4\ \mu W/cm^2$
	Machine	$1\ m\ at\ 5\ Hz$	
		$10\ m/s^2\ at\ 1\ kHz$	$100\ \mu W/cm^2$
Radio Frequency	GSM BSS	$0.3\ \mu W/cm^2$	$0.1\ \mu W/cm^2$
Thermal	Human	$20\ mW/cm^2$	$30\ \mu W/cm^2$
	Machine	$100\ mW/cm^2$	$1-10\ mW/cm^2$

1.4.Harvesting Technology

Energy gathering from surrounding environment is one of perspective solutions [2] for powering sensor nodes. There are different kinds of sources that can be harvested as shown in Table 1.1. RF-harvesting will be used in this thesis for its inexpensive implementation and this type of energy source is more reliable. RF-harvesting is done by equipping a harvesting circuit to each sensor node and this type of network is called Energy Harvesting Wireless Sensor Network (EH-WSN). RF-harvesting converting electromagnetic energy from radio frequencies to the electrical form (DC) to be saved in the storage device. The main components of harvesting unit are: antenna, impedance matching, voltage multiplier and storage capacitor.

1.5. Aim of The Study

For dense nodes' deployment, sensors have redundant readings, and hence there is a wasting of energy and information. Consequently, the main problem is, how the number of nodes which have similar readings can be reduced in order to select the specific nodes that must transmit their data to the Base Station (BS) with respect to the coverage, reliability (distortion), and Energy-Efficiency (EE).

The algorithm proposed in this thesis, addresses these problems, by exploiting spatial correlation in densely deployment case to cluster sensor nodes and reduce their energy consumption. Furthermore, the presented theoretical framework for point-to-point (single-hop) communication model, to test distortion level of all sensor nodes after nodes reduction, and for two-hop communication model to present final accuracy level of the proposed clustering algorithm with respect to power constraints.

1.6. Outline

The remaining part of this thesis is organized as follow. Chapter 2 covers related works. In chapter 3, the Event Distortion Based Clustering (EDC) algorithm is presented with theoretical frameworks for both single-hop and two-hop communication models, besides, the energy model for the resulting network (clustered-network) is discussed. EDC analysis and simulation results are presented in chapter 4 and the conclusion in the chapter 5.

CHAPTER 2

BACKGROUND AND RELATED WORKS

2.1.Overview

The advantages of WSNs are more precise nowadays and in the field of communications they cannot be neglected anymore. WSNs can provide varieties of research works with wide range of applications in military, industry, security and environmental fields [1] as discussed in previous chapter. But the drawback of WSN is its limited energy constrains as it might be deployed in areas where it is not feasible for its batteries to be replaced or recharged.

This chapter provides a brief description of the existing works related to this thesis' study. The overview of the clustering algorithms for WSNs is described in Section 2.2. In addition, the exploiting of spatial correlation is defined in Section 2.3. Finally, the summery is discussed in section 2.4.

2.2.Clustering for Wireless Sensor Networks

Reliability, Energy-Efficiency (EE), channel capacity, accuracy, and communication are crucial issues that have to be considered along with the WSN. In sensor networks, distant nodes depend on a number of intermediate nodes to forward their observations about physical phenomena. Consequently, the intermediate sensor nodes run out of their energy, and hence the reliability level is lost at the sink node. These issues can be addressed by clustering the sensor nodes. In clustering, the sensor nodes are divided into different clusters. Each cluster is managed by a node that is referred as Cluster Head (CH) and other nodes that are referred as Cluster Members (CMs). CMs don't communicate directly with the sink node. They have to pass their readings to the CH. Then, the CH aggregates their readings and transmis their data to the BS or sink node. Furthermore, The CHs are responsible for coordinating with both inter-cluster and intra-cluster communications[1] as shown in Fig 2.1.

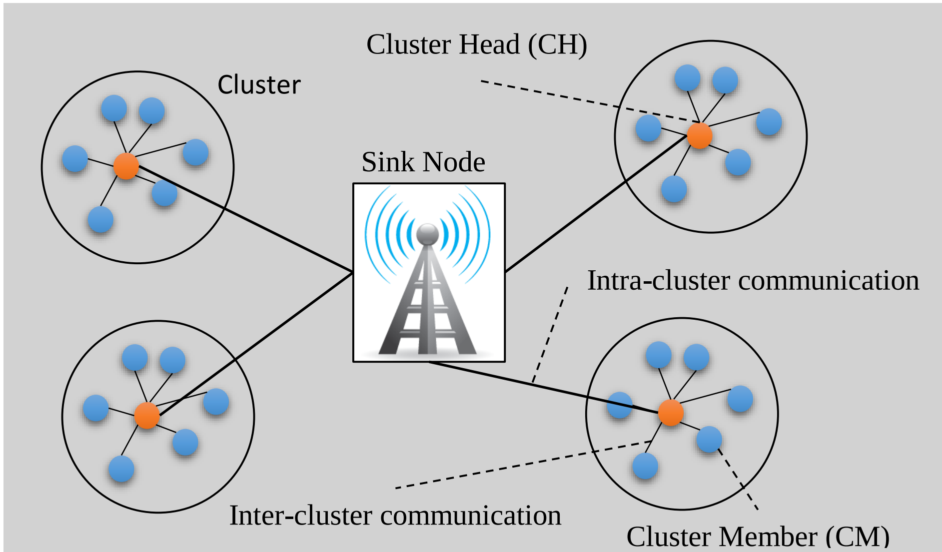


Figure 2.1 Clustering for Wireless Sensor Network

The crucial results from clustering the sensor nodes are prolonged network lifetime, scalability for WSNs, and reduced communication overhead, etc. Hence, the clustering algorithms can be exploited in many applications, such as earthquake monitoring and applications that support data aggregation, such as microclimate and habitat monitoring.

As discussed in previous chapter, in a dense deployment case, nearby sensor nodes have correlated sensing readings (due to spatial correlation) as in Fig 2.2. Hence, it may not be essential for every sensor node to transmit its data to the sink; instead, a smaller number of sensor nodes referred to as active nodes, might be sufficient to transfer event data to the sink node with respect to the predefined distortion level, while the nodes that are prevented from observing are called inactive nodes. One prevalent way [7] for conserving energy is clustering the network into several clusters for example Low-Energy Adaptive Clustering Hierarchy (LEACH) [8] and an Energy Efficient Clustering Scheme (EECS) [9].

[1] The communications that generated inside the cluster (e.g., between CM and CH) are referred as inter-cluster communications, while the communications between two CHs that located at different clusters are referred as intra-cluster communications.

Furthermore, there are several clustering algorithms that exploits spatial correlation: GAG [10], and EAST [11], etc.

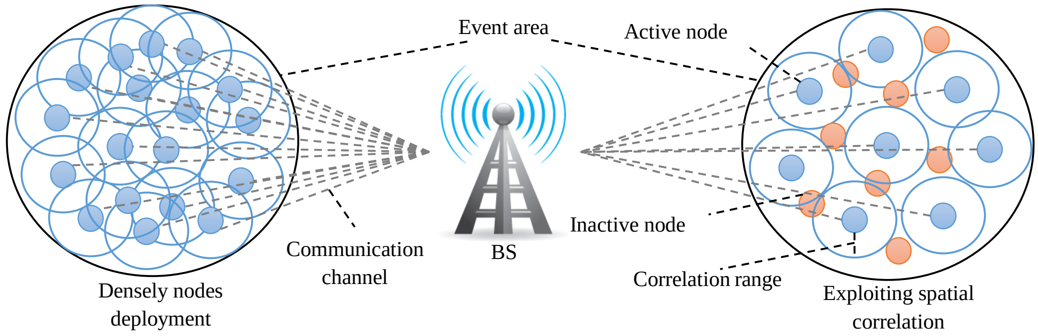


Figure 2.2 Exploiting Spatial Correlation

In [10], the authors have presented a clustering algorithm to reduce the number of transmissions towards the sink node to reduce overhead packets and to enhance total EE of sensor nodes. The process has been done by grouping the sensor nodes which have similar readings into set of clusters, the size of each cluster defined by correlation range. At each iteration, only one sensor node from each cluster is required to send its own data to the sink node, based on fault-tolerance (accuracy level) with 10% error threshold. Finally, a reduction of 70 % of overhead packets have been observed.

In [11], the author followed the general idea of [10], but in [11], the difference between event sources has been investigated. For events that change at small range, the correlated region can be decreased to keep the accuracy level, i.e., the event needs to be informed by nearby nodes. For events that do not change at small range, the correlated region can be increased to improve the EE. The size of the correlation region can be resized by the sink node.

To qualify the existing algorithms in WSNs, distortion level and energy conservation should be considered as the main issues. In WSNs, the distortion represents the similarity level between original and estimated readings. Different distortion functions have been derived in [3] and [4] in Minimum Mean Squared Sense (MMSE).

In [3], the author presented an algorithm called Iterative Node Selection (INS) for selecting the representative nodes by exploiting spatial correlation through Vector Quantization (VQ). The communication model has been considered according to the point-to-point (single-hop) case, i.e., each sensor node transmits its reading directly to the sink node, as shown in Fig 2.2. In addition, the author hasn’t assumed any channel noise in the communication links. Furthermore, the spatial correlation model between sensors observations is modeled by exponential model.

In [4], the same communication model has been used, but the sensor readings are similar and without observation noise. Also, there is one noise in the end of network (sink node). Hence, the distortion function is derived with respect to power constraints.

2.3. Summary

The proposed work is inspired by these above works to minimize data redundancy for preserving node energy and prolonging the network life time. In this thesis, a new clustering algorithm is proposed by exploiting spatial correlation through Vector Quantization (VQ). In contrast to the previous studies (e.g., [3]), our focus is that there is a channel noise in each communication link, besides there are observation noises for readings of sensor nodes. In such a case, the distortion function is derived with respect to the power constraints. Then, by clustering the sensor nodes through VQ, a different communication model is formed. Hence, another distortion function is derived for clustered-network, i.e., two-hop communication model, which is unlike previous studies that used fault-tolerance to define the accuracy level.

CHAPTER 3

EVENT DISTORTION-BASED CLUSTERING

3.1. Overview

In WSNs, clustering is one of the important techniques that is applied for enhancing Energy-Efficiency (EE) and prolonging the lifetime of the network. In the preceding chapter, various clustering algorithms which can exploit spatial correlation among sensor nodes and different distortion functions to examine the reliability level of sensor network are discussed. Those previous works have been devised for wireless sensor networks in which sensors are assumed to be battery-powered operated nodes. However, the clustering of battery-free sensor nodes in energy harvesting wireless sensor networks poses different challenges, such as lack of a permanent power source to sense or enable communication process whenever it is needed. In order to address such challenges, in this chapter, a clustering algorithm called Event Distortion Based Clustering (EDC) algorithm is introduced for energy harvesting wireless sensor networks.

In this chapter, the problem statement is first given in Section 3.2. Then, the network model of WSN is introduced in Section 3.3. In Section 3.4, the EDC algorithm is introduced and finally, the energy model that is used for the performance evaluations of the EDC algorithm is given in Section 3.5.

3.2.Problem Statement

In this thesis, the network is considered as a set of N homogeneous wireless sensor nodes $\{n\}_N = [n_1, n_2, \dots \dots \dots, n_N]$. The sensor nodes are assumed to be randomly and densely deployed in an event area, and hence the sensor readings tend to be spatially correlated and thus, redundant. Hence, due to this redundancy which may inject excessive traffic load on the network and the lack of permanent energy source in the nodes, it is important to decide which sensor nodes should transmit their readings in order to provide an acceptable level of the event signal reconstruction distortion at the sink node. This issue is addressed by the EDC algorithm by exploiting spatial correlation through the Vector Quantization (VQ) in the following chapter. Next, the network model for WSN is considered in this thesis is introduced.

3.3. Network Model

The sensor nodes are deployed in the event area for observing the physical event, which generates an event signal, S, through a point source. Each sensor node has a fixed location and it capable to send its data through single-hop (point-to-point) and multi-hop communication models as illustrated in Fig 3.1. The sink node is assumed to be at the center of the network. Furthermore, the following assumptions are considered in the thesis:

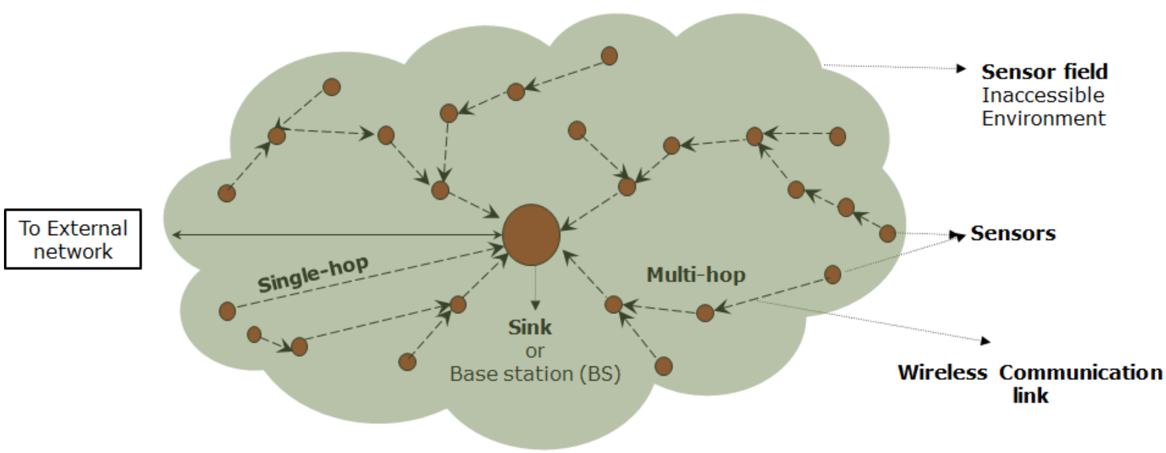


Figure 3.1 Wireless Sensor Network Considered in This Thesis

- All sensor nodes are densely and randomly deployed in the event area.
- All sensors are capable of adjusting their transmission power levels.
- All sensor nodes are homogeneous and have the same capabilities.
- Each sensor is equipped with a harvesting unit to harvest energy from ambient electromagnetic radiation in radio frequencies.

- Each sensor node transmits its data to the other sensor node or to the sink node without encoding the data.

3.4. Event Distortion Based Clustering (EDC) Algorithm

Here, the EDC algorithm is introduced. It is an off-line algorithm which runs in the sink node to cluster the sensor nodes so as to determine which sensor nodes should send their readings to the sink node for an acceptable level of event signal reconstruction distortion. The distortion function is first derived for single-hop model. Then, the clustering of the sensor node is explained through the EDC algorithm. Finally, based on the clustered network, the distortion function is used to evaluate the performance of the EDC algorithm.

3.4.1. Distortion with Point-To-Point Communication

In this section, a scenario in which all of N sensor nodes send their readings to the sink node through single-hop links are considered. It is also assumed that all the sensor readings can reliably reach the sink node without any loss. The model for the information gathered by N sensors in an event area is illustrated in Fig. 3.2. The sink is concerned in estimating the event source (i.e., point source at location, n_0), S, according to the observations of the sensor nodes, n_i, in the event area. Each node, n_i, observes, $X_i[n]$, the noisy version of the event information, $S_i[n]$, which is spatially correlated to the event source, S. In order to transfer these observations to the sink node, the sensor nodes don't encode their readings. The uncoded information are then sent to the sink node by single-hop communications through channel noise, $W_i[n]$. The sink node estimates these information, $Y_i[n]$, to get the estimate, $\hat{S}$, of the event source S.. Each observed sample, $X_i[n]$, of sensor, n_i, at time n is defined as:

$$X_i[n] = S_i[n] + N_i[n] \tag{3.1}$$

where the location of node n_i denoted by the subscript i, i.e., (x_i, y_i).

$S_i[n]$ is the realization of the space-time process $s(t,x,y)$ at time $t^2 = t_n$ and $(x,y) = (x_i, y_i)$, and $N_i[n]$ is the observation noise. $\{N_i[n]\}_n$ is a sequence of i.i.d Gaussian random variables of zero mean and variance σ_N^2. Furthermore, the noises at different sensors are independent of each other, that is, $N_i[n]$ and $N_j[n]$ are independent for $i \neq j$ and $\forall n$.

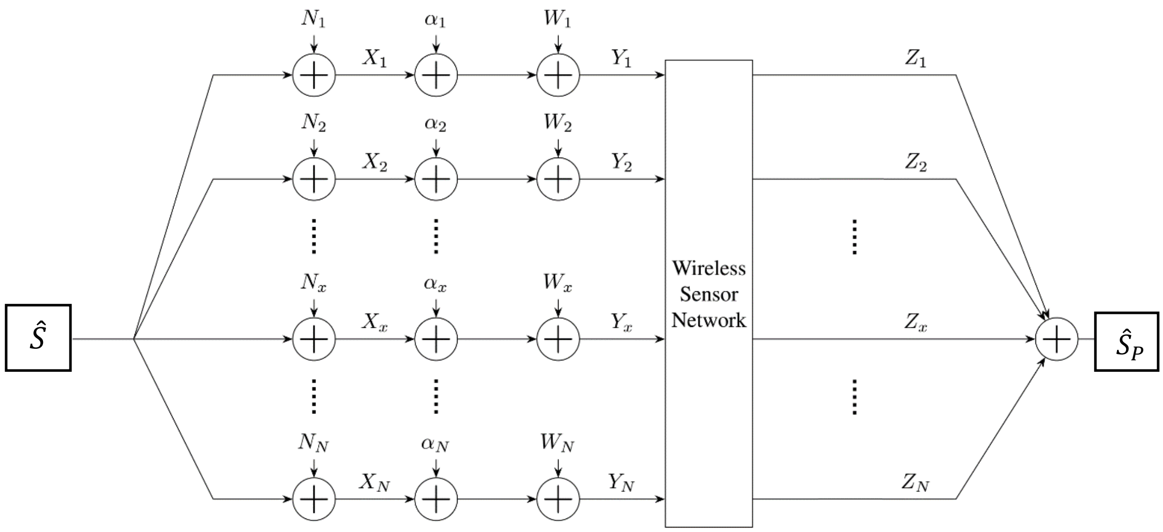


Figure 3.2 Point-to-Point communication Model

By dropping the time index n, (3.1) can be restated as:

$$X_i = S_i + N_i \tag{3.2}$$

where the observation noise, N_i, of each sensor node, n_i, is modeled as i.i.d. Gaussian random variable of zero mean and variance σ_N^2. Furthermore, the event information, S_i, at node n_i is modeled as Joint Gaussian Random Variables (JGRVs) as:

$$E[S_i] = 0, \quad var[S_i] = \sigma_S^2, \qquad i = 1,2,\dots,N. \tag{3.3}$$

The physical event information is assumed to have an exponential autocorrelation function [32], [33] the correlation coefficient , $\rho(S_i, S_j)$, between the observations of sensor nodes, n_i and n_j, can be given as follows:

---

[2] Note that, a discrete-time model is used, since each node samples the physical phenomenon synchronously after the initial wake-up.

$$\rho(S_i, S_j) = \frac{E[S_i S_j]}{\sigma_S^2} = e^{\left(d(n_i,n_j)/\theta_1\right)^{\theta_2}}. \tag{3.4}$$

Here, $\theta_2 = 1$ is used in exponential model, and θ_1 controls the correlation range between the observations of sensor nodes. Furthermore, $d(n_i, n_j) = ||n_i - n_j||$ denotes the distance between nodes n_i and n_j. This correlation coefficient, $\rho(S_i, S_j)$, is assumed to be non-negative and decreases monotonically with the distance, $d(n_i, n_j)$, with limiting values of 1 at $d = 0$ and of 0 at $d = \infty$.

The event source, S, is a JGRV and it is of special interest. Hence, the $\rho(S, S_i)$ and $d(n_0, n_i)$, can be denoted as the correlation coefficient and the distance, between the event source, S, at n_0 and sensor node, n_i, respectively. As follows:

$$\rho(S, S_i) = \frac{E[S\ S_i]}{\sigma_S^2} = e^{\left(d(n_0,n_i)/\theta_1\right)^{\theta_2}}. \tag{3.5}$$

where $d(n_0, n_i) = ||n_0 - n_i||$ denotes the distance between event source at position n_0, S, and node, n_i. As each sensor node, n_i, observes an event information, X_i. This information is then encoded to be forwarded to the sink node. It has been investigated that for sensor networks, the uncoded transmission performs better than the coding approach with Gaussian source [4]. Hence, uncoded transmission is adopted for the sensor readings, in this analysis. Consequently, each node, n_i, sends to the sink a scaled version of the observed sample, X_i, to meet its power constraint, $\hat{P}$, i.e., transmitted power for data from sensor, n_i, to the sink node, which it can be defined by calculating the energy consumption for transmitting number of bits over specific interval, it will be defined in Section 3.5. It is given as follows:

$$\sum_{i=1}^{N} E[(\alpha\, X_i)^2] \leq N\hat{P}. \tag{3.6}$$

By taking the upper bound, then (3.6) can be simplified with respect to (2) as:

$$E\left[\left(\alpha\,(S_i + N_i)\right)^2\right] = \hat{P}. \tag{3.7}$$

Since each sensor, n_i, is capable to adjust its power level based on the destination, then $(\hat{P} = P_i)$[3]. Then, the scaler, α_i, can be defined after taking the expectation operator as:

$$\alpha_i = \sqrt{\frac{P_i}{\sigma_S^2 + \sigma_N^2}}. \tag{3.8}$$

where σ_S^2 and σ_N^2 are the variances of the event information, S_i, and the observation noise, N_i, respectively. Then, the received signal at sink node, Y_i, can be defined as

$$Y_i = \alpha_i X_i + W_i. \tag{3.9}$$

where $\{W\}_i$ is a channel noise between node, n_i, and sink node, which its modeled as a set of i.i.d Gaussian random variables with zero mean and variance σ_W^2.

The sink needs to calculate the estimation of each event information, S_i. Since uncoded transmission is used, it is well known that the Minimum Mean Square Error (MMSE) estimation is the optimum decoding technique [34]. Hence, the estimation, Z_i, of the event information, S_i, is simply the MMSE estimation of Y_i. Then, Z_i can be defined according to the linear transformation of Y_i as:

$$Z_i = a\, Y_i. \tag{3.10}$$

where a is a constant. For optimal case, the received sample, Z_i, in sink node is equal to the observation, S_i, of node, n_i, as given below:

$$E[(Z_i - S_i)^2] = 0. \tag{3.11}$$

In order to find an optimal a, the derivative is taken with respect to a and the expectation is applied after (3.8) is substituted in (3.9), as follows:

$$E\left[\frac{d}{da}\left(\left(a^2\, {Y_i}^2\right) - 2(a\, Y_i\, S_i) + \left({S_i}^2\right)\right)\right] = 0. \tag{3.12}$$

Then,

[3] The transmitted powers can be calculated based on sensor locations at the sink node, and hence they are assumed to be fixed.

$$E\big[\big(2\ a\ {Y_i}^2\big) - 2(Y_i\ S_i)\big] = 0. \tag{3.13}$$

Finally, a can be written as:

$$a = \frac{E[Y_i S_i]}{E\big[{Y_i}^2\big]}. \tag{3.14}$$

After finding a, Z_i can be easily expressed by substituting (3.2), (3.9) and (3.14) is substituted in (3.10), as follows;

$$Z_i = \frac{E[Y_i S_i]}{E[{Y_i}^2]} Y_i = \frac{E[(\alpha_i\ X_i + W_i\ )\ S_i]}{E[(\alpha_i\ X_i + W_i)^2]} Y_i. \tag{3.15}$$

Since there is no correlation between S_i and W_i, it can be written as;

$$Z_i = \frac{E[(\alpha_i\ X_i\ S_i)]}{E[(\alpha_i\ X_i)^2] + E[(W_i)^2]} Y_i = \frac{\alpha_i\ E[X_i\ S_i]}{var(\alpha_i\ X_i) + var(W_i)} Y_i. \tag{3.16}$$

where the means of X_i and W_i are zeros. Consequently, Z_i can be simplified as;

$$Z_i = \frac{\alpha_i\ E[{S_i}^2]}{{\alpha_i}^2(\sigma_S^2 + \sigma_N^2) + \sigma_W^2}\ Y_i. \tag{3.17}$$

Finally, Z_i can be stated as;

$$Z_i = \frac{\alpha_i\ \sigma_S^2}{P_i + \sigma_W^2}\ Y_i. \tag{3.18}$$

The Z_i is decoded using (MMSE) estimator in the sink node. Hence, the estimation, $\hat{S}_P$, of event source, S, can be computed by taking the average of all the received event information, Z_i. Then, the estimated version, $\hat{S}_P$, can be formed as,

$$\hat{S}_P = \frac{1}{N}\sum_{i=1}^{N} Z_i. \tag{3.19}$$

Finally, the distortion for single-hop communication model, D_{point}, can be presented with respect to the estimated value, $\hat{S}_P$, as:

$$D_{point} = E\left[\left(S - \hat{S}_P\right)^2\right] = E[S^2] - 2\,E\left[S\,\hat{S}\right] + E\left[\hat{S}^2\right]. \tag{3.20}$$

The first term, $E[S^2]$, is equal to σ_S^2. The second term, $E\left[S\,\hat{S}_P\right]$, is calculated by substituting (3.19), (3.18), (3.9), and (3.2) as follows:

$$E\left[S\,\hat{S}_P\right] = E\left[S\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\alpha_i\,\sigma_S^2}{P_i + \sigma_W^2}\,\left(\alpha_i\,(S_i + N_i) + W_i\right)\right)\right]. \tag{3.21}$$

Note that, there is no correlation between (S and N_i), and (S and W_i). Hence, it can be solved by:

$$E\left[S\,\hat{S}_P\right] = \frac{1}{N}\sum_{i=1}^{N}\frac{{\alpha_i}^2\,\sigma_S^4}{P_i + \sigma_W^2}\,\rho(S, S_i). \tag{3.22}$$

Since mean of $\hat{S}$ is zero, the third term, $E\left[\hat{S}_P^2\right]$, is equal to the variance of the estimated version, $\hat{S}$, of S. i.e.,

$$E\left[\hat{S}_P^2\right] = var\left(\hat{S}\right). \tag{3.23}$$

By using (3.17),

$$E\left[\hat{S}_P^2\right] = var\left(\frac{1}{N}\sum_{i=1}^{N} Z_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} var(Z_i) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[Z_i Z_j\right]. \tag{3.24}$$

Substituting (3.19), (3.18), (3.9), and (3.2) in (3.24), then (3.24) can be extended as shown:

$$\begin{aligned} E\left[\hat{S}_P^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{\alpha_i\,\sigma_S^2}{P_i + \sigma_W^2}\right)^2\,var(\alpha_i\,X_i + W_i) + \frac{\sigma_S^4}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\alpha_i}{P_i + \sigma_W^2}\,\frac{\alpha_j}{P_j + \sigma_W^2} \\ \times E\left[\left((\alpha_i(S_i + N_i) + W_i\right)\,\left(\alpha_j(S_j + N_j) + W_j\right)\right]. \end{aligned} \tag{3.25}$$

Then, it can be reduced after taking the expectation as shown:

$$E\left[\hat{S}_P^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\alpha_i^2\,\sigma_S^4}{P_i + \sigma_W^2} + \frac{\sigma_S^4}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{{\alpha_i}^2}{P_i + \sigma_W^2}\,\frac{{\alpha_j}^2}{P_j + \sigma_W^2}\,\rho\left(S_i, S_j\right). \tag{3.26}$$

Using (3.22) and (3.26) in (3.20), the result can be finalized as:

$$D_{point} = \sigma_S^2 - \frac{\sigma_S^4}{N^2} \sum_{i=1}^{N} \frac{\alpha_i^2}{P_i + \sigma_W^2} \left(2\, N\, \rho(S, S_i) - 1 - \sum_{j=1}^{N} \frac{{\alpha_j}^2\, \rho\left(S_i, S_j\right)}{P_j + \sigma_W^2} \right). \qquad (3.27)$$

where $\rho(S, S_i)$ and $\rho\left(S_i, S_j\right)$ are correlation coefficients between node, n_i, and source event, S, and between node, n_i, and node, n_j, respectively. This distortion function represents the accuracy level of original event S at sink node for single-hop communication with respect to observation and channel noises.

3.4.2. Clustering Formation

According to the results that obtained in the preceding section, the EDC algorithm is introduced. The EDC algorithm is firstly required to select a smaller set of sensor nodes, rather than all sensor nodes. These selected sensor nodes are defined through exploiting spatial correlation such that the acceptable level of event signal reconstruction distortion can be maintained at the sink node. The EDC algorithm uses VQ as a way to exploit spatial correlation. Secondly, the sensor nodes are clustered by the sink node based on the selected and unselected nodes that are defined by VQ. An overview about the VQ design problem is given next.

VQ is a lossy data compression method [12], that compresses the set of pixels for images, or the set of bits in signals (e.g., speech signals). The VQ maps k-dimensional source vectors (i.e., pixels or bits) into the finite set of vectors called codewords. The set of all codeword vectors is called codebook. Associated with each codeword, a nearest region is called Voronoi region. However, this method reduces the quality of images or signals, because VQ is nothing more than an approximation.

Generally, the samples of speech signal might be temporally correlated, and the pixels of specific image might be spatially correlated. Therefore, the selection of correlated points based on a distortion constraint has been investigated by exploiting VQ methods [12]. Hence, these methods have been used by properly addressing the dense deployment issue in WSN [3].

The VQ design can be described as follows [12]: Given a vector source with its statistical properties known[4], given a distortion constraint, and given the number of codewords, the VQ algorithm tries to find a codebook and partitions (i.e., voronoi regions) which result in the smallest possible average distortion. More specifically, the VQ algorithm aims to represent all possible codewords in a code space by a subset of codewords, within the distortion constraint (i.e., maximum acceptable distortion). Hence, the spatial correlation can be exploited using VQ, where all the sensor nodes in an event area are needed to be represented with a smaller number. If two dimensional source vector is selected, the code space in the VQ method is applied to the network topology with the sensor nodes as the codeword spaces. The VQ algorithm, when it is applied to the sensor nodes selection issue, the codebook and the partitions can be determined. Since the sensor nodes are densely deployed, it is possible to refer the closest nodes to the fixed locations of codebook as Representative Nodes, while the other remaining nodes are referred as Unrepresentative Nodes.

The EDC algorithm starts with selecting the length of codebook as one, to be an input to the VQ, as well as all N sensor nodes as a source vector. The outputs are one representative node and set of unrepresentative nodes. Then, the EDC algorithm iteratively increases the number of representative nodes (i.e., by increasing the length of codebook). For each increment, the unrepresentative nodes are decreased. The EDC algorithm continues to increase the number of representative nodes until the point distortion, defined by (27), of the unrepresentative nodes achieves distortion constraint.

The INS algorithm [3] successfully selects the representative nodes but it does not exploit the unselected nodes. However, the EDC algorithm consideres only the unrepresentative nodes readings (instead of representative nodes) to be transmitted to the sink node. Moreover, the EDC algorithm uses the representative nodes in relaying the observations of unrepresentative sensor nodes to the sink node. In the EDC algorithm, only the readings of unrepresentative sensor nodes are considered at the sink node. The majority of the unrepresentative nodes and the advantage of their coverage made the distortion more likely to be maintained at the sink node, rather than representative sensor nodes.

[4] The statistical property is referred as a type of source vector distribution, or deployment approach in the WSN sense. In WSN, nodes deployment can be done through Uniform, Poisson, or Gaussian distribution.

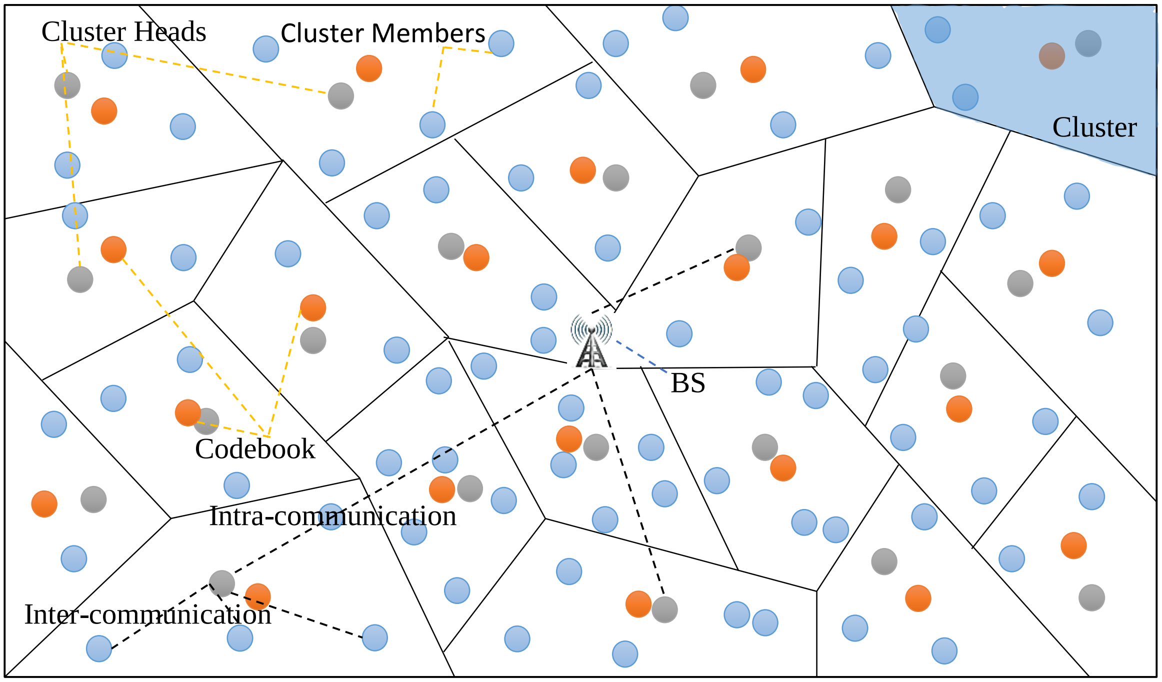


Figure 3.3 Clustered-Network by EDC Algorithm

The EDC algorithm then forms a sample topology, as shown in Fig 3.3. The Voronoi diagrams are referred as clusters, administered by their Cluster Heads (CHs), (i.e., c representative nodes). All Cluster Members (CMs), (i.e., all r unrepresentative nodes) are located inside clusters that are assigned by K-Nearest Neighbor (KNN) [24]. Each cluster k contains set of m_k[5] cluster members, which are defined through KNN with respect to their CHs. Furthermore, The distortion constraint represents the maximum acceptable distortion level in single-hop case for unrepresentative nodes, $D_{Point}(r)$ with respect to distortion threshold, $D_{thresold}$, which it defined with respect to the application requirement.

[5] Each cluster contains different numbers of CMs, i.e., $m_k \neq m_l$, that are calculated in the sink node based on KNN with respect to voronoi regions, and hence they assumed to be fixed not random.

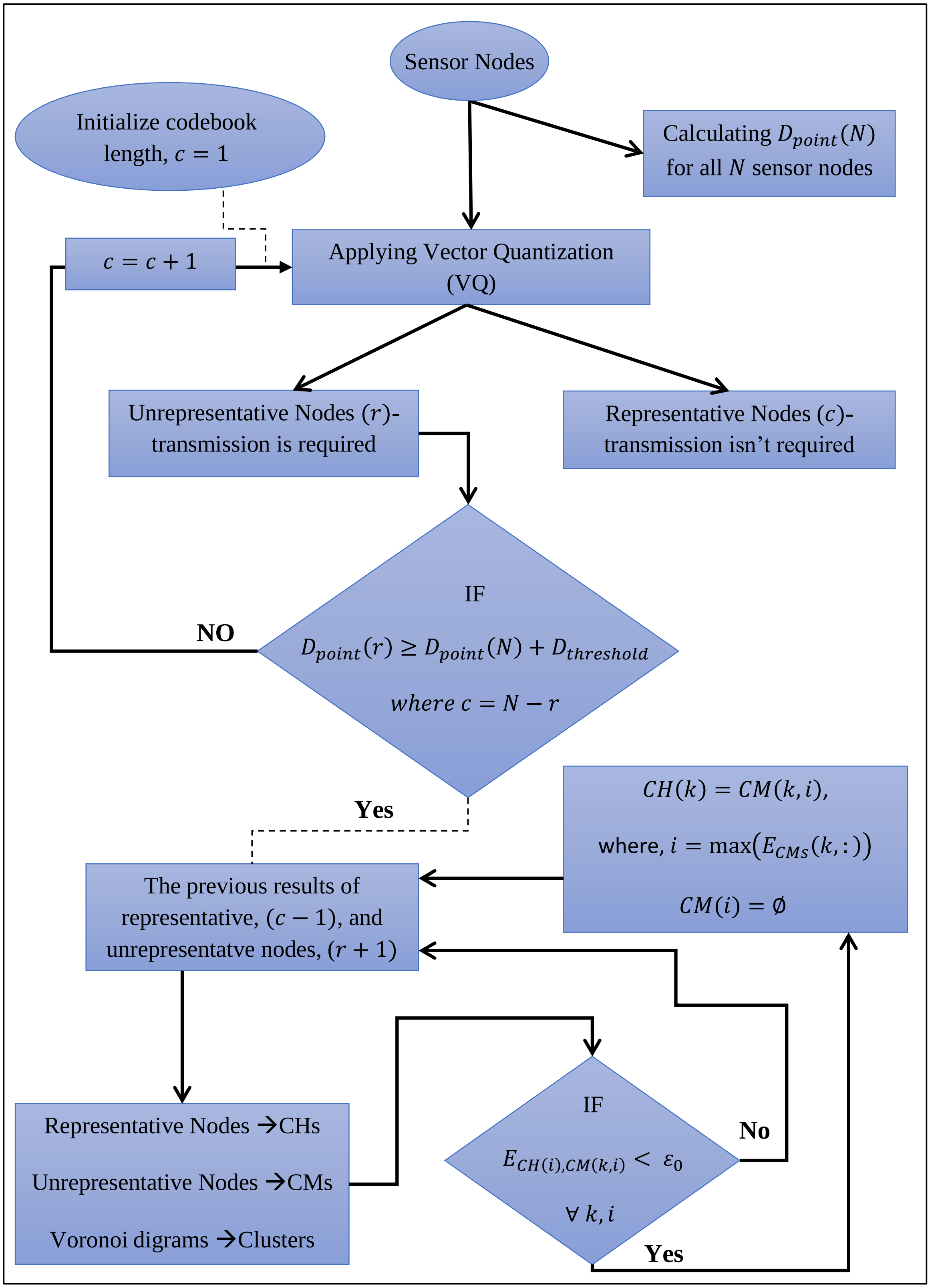


Figure 3.4 Event Distortion Based Clustering (EDC) Algorithm

After defining a clustered-network, the EDC algorithm starts to check the energy level of all sensor nodes (CH & CM) and changes their duties based on their remaining energies. Since each sensor node is assumed to transmit its information about its remaining energy level to the sink node.

In the case that the energy level of CH (or CM) is under a predefined threshold, ε_0, it turns to inactive state (i.e., sensor node is prevented from observing, aggregating and communicating etc.). Furthermore, if the energy level of cluster head k at cluster k, $E_{CH}(k)$, is less than ε_0, the CM with highest energy level is elected to play CH role until CH recover its energy. Moreover, recovering the energy is assumed to be done continuously through energy harvesting from ambient electromagnetic radiation in radio frequencies.

During the lifetime of the network, the network topology can change due to failure or battery drain of sensor nodes. However, since the distortion depends on the physical phenomenon, such a change should not affect the distortion achieved at the sink, unless the number of CMs decreases significantly (i.e., when the energy level of any cluster member i at cluster k, $E_{CM}(k,i)$, is less than ε_0). In such a case, the event information cannot be captured in the sink node at the desired distortion level.

3.4.3. Distortion with Two-hop Communication

The presented EDC algorithm uses two-hop communication model for data transfer as shown in Fig 3.3, unlike single-hop model that has been defined in Fig 3.2. The model for the information gathered by all Cluster Members (CMs) in the event area is shown in Fig 3.5. The sink node estimates the event source, S, i.e., point source at location n_0, according to the observations of only CMs, $S_{k,i}$. Each cluster member i at cluster k, $CM_{k,i}$, observes the noisy version, $X_{k,i}$, of the event information, $S_{k,i}$, which is spatially correlated with the event source, S. In order to transfer these observations to the sink node, the CMs don't need encode their readings. The uncoded information are then sent to the cluster head k at cluster k, CH_k, by single-hop communications through channel noise, $W_{k,i}$. Then, each CH_k receives and aggregates the samples from its CMs, $Y_{k,i}$, after decoding them. Also, CH_k does not need to encodes its data, and hence it only requires to receive the observations from m_k cluster members and then it aggregates their readings. Then, the aggregated information, S_k, is relayed to the sink node through another channel noise, g_k. The sink is then estimates from the information of CH_k, Y_k, to get the estimation, $\hat{S}_R$, of the event source, S. Hereafter, for ease of illustration, the time index is dropped. Hence, each observed sample, $X_{k,i}$, of $CM_{k,i}$ is defined as:

$$X_{k,i} = S_{k,i} + N_{k,i}. \tag{3.28}$$

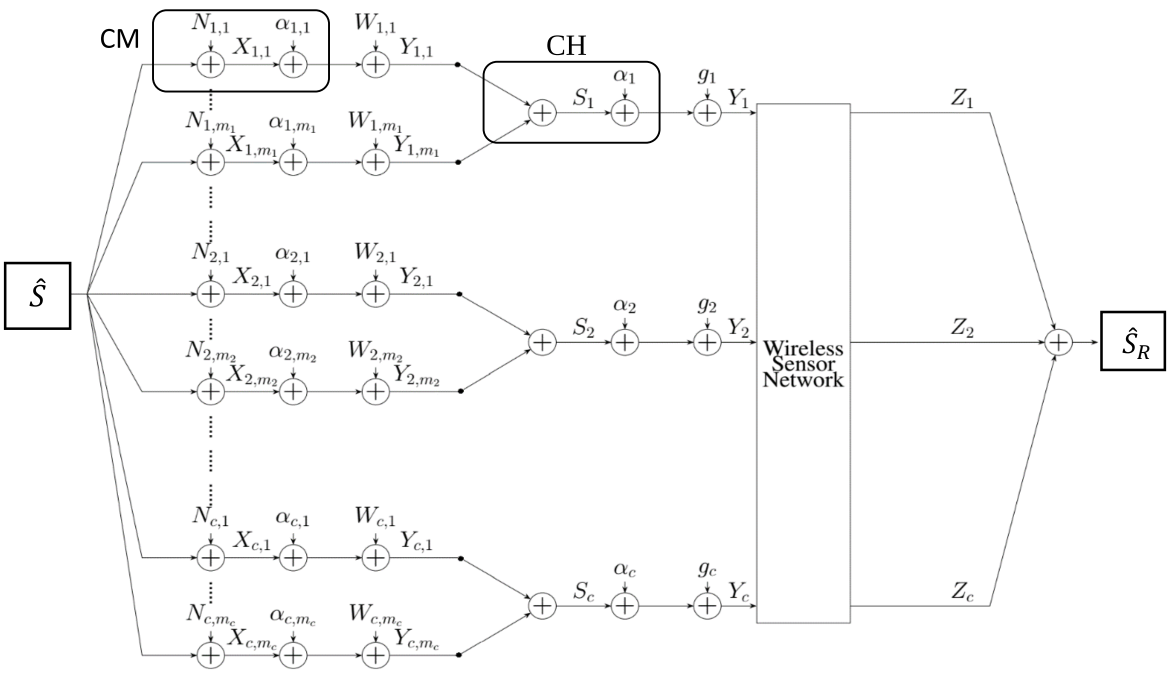


Figure. 3.5 Two-hop Communication Model

where $\{N\}_{k,i}$ is a set of i.i.d Gaussian random variables with zero mean and variance σ_N^2. The event information, $S_{k,i}$, at each $CM_{k,i}$ is modeled according to the Joint Gaussian Random Variables (JGRVs) as follows:

$$E[S_{k,i}] = 0, \qquad var[S_{k,i}] = \sigma_S^2, \quad i = 1,2, \dots, m_k. \tag{3.29}$$

where these properties are the same for different CMs and even in different clusters. The CMs that are located in the same cluster might be correlated, as well as in different clusters. Since, the physical event has an exponential autocorrelation function, then, the correlation coefficient , $\rho(S_{k,i}, S_{k,j})$, between the observations of cluster member i and j at cluster k, and

$\rho(S_{k,i}, S_{l,j})$, between the observations of cluster members i and j at clusters k and l, respectively can be given as follows:

$$\rho\left(S_{k,i}, S_{k,j}\right) = \frac{E\left[S_{k,i}\, S_{k,j}\right]}{\sigma_S^2} = e^{\left(d\left(CM_{k,i}, CM_{k,j}\right)/\theta_1\right)^{\theta_2}} \tag{3.30}$$

$$\rho\left(S_{k,i}, S_{l,j}\right) = \frac{E\left[S_{k,i}\, S_{l,j}\right]}{\sigma_S^2} = e^{\left(d\left(CM_{k,i}, CM_{l,j}\right)/\theta_1\right)^{\theta_2}}. \tag{3.31}$$

Similarly, the event source, S, is modeled as JGRV. The correlation coefficient, $\rho(S, S_{k,i})$, and the distance, $d(n_0, CM_{k,i})$, between the event source, S, and $CM_{k,i}$, respectively, are given bellow:

$$\rho\left(S, S_{k,i}\right) = \frac{E\left[S\, S_{k,i}\right]}{\sigma_S^2} = e^{\left(d\left(n_0, CM_{k,i}\right)/\theta_1\right)^{\theta_2}}. \tag{3.32}$$

Since uncoded transmission has been adopted in this analysis for single-hop communication with Gaussian source, then the received sample $Y_{k,i}$ at CH can be defined as:

$$Y_{k,i} = \alpha_{k,i} X_{k,i} + W_{k,i}. \tag{3.33}$$

where $\{W\}_{k,i}$ is a channel noise (between $CM_{k,i}$ and CH_k), which is defined as a set of i.i.d Gaussian random variables with zero mean and variance σ_W^2. And the scalar α is defined with respect to its power constraints, $\hat{P}$, as:

$$\sum_{i=1}^{m_k} E[\alpha\, X_{k,i}]^2 \leq m_k\, \hat{P}. \tag{3.34}$$

Since each $CM_{k,i}$ is capable to adjust its transmitted power level, then $\hat{P} \triangleq P_{k,i}$, i.e., the transmitted power for the information from $CM_{k,i}$ to the CH_k. Hence, by taking the upper bound of (3.34), then the scaler $\alpha_{k,i}$ can be simplified as:

$$\alpha_{k,i} = \sqrt{\frac{P_{k,i}}{\sigma_S^2 + \sigma_N^2}}. \tag{3.35}$$

In order to estimate the received sample at CH, MMSE estimation is used. Hence, received signal at CH_k, $Z_{k,i}$, can be defined according to the linear transformation of $Y_{k,i}$ as follows:

$$Z_{k,i} = a\ Y_{k,i}. \tag{3.36}$$

where a is constants. According to the MMSE estimator the optimal case the received signal, $Z_{k,i}$, is equal to the sensor observation, $S_{k,i}$, as:

$$E[(Z_{k,i} - S_{k,i})^2] = 0. \tag{3.37}$$

Similarly, in order to find an optimal a, the derivative is calculated with respect to a and the expectation is applied after (36) is substituted in (37), as defined bellow:

$$E\left[\frac{d}{da}\left(a^2\ {Y_{k,i}}^2 - 2\ a\ Y_{k,i}\ S_{k,i} + {S_{k,i}}^2\right)\right] = 0. \tag{3.38}$$

Since the means of source $S_{k,i}$ and noise $N_{k,i}$ are zero, then (3.38) can be solved and a can be given by:

$$a = \frac{E[Y_{k,i}\ S_{k,i}]}{E[{Y_{k,i}}^2]}. \tag{3.39}$$

After finding a, $Z_{k,i}$ can expressed by using (3.28), (3.33), (3.39) in (3.36) as follows:

$$Z_{k,i} = \frac{E[Y_{k,i}\ S_{k,i}]}{E[{Y_{k,i}}^2]} Y_{k,i} = \frac{E[(\alpha_{k,i}\ X_{k,i} + W_{k,i}\ )\ S_{k,i}]}{E[(\alpha_{k,i}\ X_{k,i} + W_{k,i})^2]} Y_{k,i}. \tag{3.40}$$

Since there is no correlation between $S_{k,i}$ and $W_{k,i}$, then it can be reduced to:

$$Z_{k,i} = \frac{E[(\alpha_{k,i}\ X_{k,i}\ S_{k,i})]}{E[(\alpha_{k,i}\ X_{k,i})^2] + E[(W_{k,i})^2]} Y_{k,i} = \frac{\alpha_{k,i}\ E[X_{k,i}\ S_{k,i}]}{{\alpha_{k,i}}^2 var(\ X_{k,i}) + var(W_{k,i})} Y_{k,i}. \tag{3.41}$$

where the means of $X_{k,i}$ and $W_{k,i}$ are zero. Hence, $Z_{k,i}$ can be finally simplified as:

$$Z_{k,i} = \frac{\alpha_{k,i}\ \sigma_S^2}{P_{k,i} + \sigma_W^2}\ Y_{k,i}. \tag{3.42}$$

Furthermore, since the CH_k decodes each $Y_{k,i}$ using the MMSE estimator, the estimation, S_k, of the event information, $S_{k,i}$ can simply be computed by taking the average of m_k readings of CMs, $Z_{k,i}$. Then, the estimation, S_k, is given as follows:

$$S_k = \frac{1}{m_k} \sum_{i=1}^{m_k} \frac{\alpha_{k,i} \sigma_S^2}{P_{k,i} + \sigma_N^2} Y_{k,i}. \tag{3.43}$$

The uncoded transmission is not only optimal for single-hop communication, but also for two-hop communication model [14]. Then, the received sample, Y_k, at sink node, and it can be defined as:

$$Y_k = \alpha S_k + g_k. \tag{3.44}$$

where $\{g\}_k$ is a channel noise that between CH_k and sink node, which it is defined as a set of i.i.d Gaussian random variables with zero mean and variance σ_g^2. And the scalar α is defined with respect to its power constraint, $\hat{P}$, as shown:

$$\sum_{k=1}^{c} E[(\alpha\, S_k)^2] \leq c\, \hat{P}. \tag{3.45}$$

where the mean of S_k is zero. By substituting (3.43) in (3.45), and by taking the upper bound then, (3.45) can be written as:

$$\sum_{k=1}^{c} \alpha^2\, var(\frac{1}{m_k} \sum_{i=1}^{m_k} Z_{k,i})^2 = c\, \hat{P}. \tag{3.46}$$

which it can be extended to:

$$\sum_{k=1}^{c} \frac{\alpha^2}{{m_k}^2} \left(\sum_{i=1}^{m_k} var(Z_{k,i}) + \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} cov(Z_{k,i}\,, Z_{k,j}) \right) = c\, \hat{P}. \tag{3.47}$$

By substituting (3.28), (3.33), (3.35), and (3.42) in (3.47). Then, (3.47) can be extended as:

$$\begin{aligned} &\sum_{k=1}^{c} \frac{\alpha^2}{{m_k}^2} \left(\sum_{i=1}^{m_k} var \left(\frac{\alpha_{k,i}\, \sigma_S^2}{P_{k,i} + \sigma_W^2} \left(\alpha_{k,i} \left(S_{k,i} + N_{k,i} \right) + W_{k,i} \right) \right) \right. \\ &\qquad \left. + \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} E \left[\left(\frac{\alpha_{k,i}\, \sigma_S^2}{P_{k,i} + \sigma_W^2} \left(\alpha_{k,i} \left(S_{k,i} + N_{k,i} \right) + W_{k,i} \right) \right) \left(\frac{\alpha_{k,j}\, \sigma_S^2}{P_{k,j} + \sigma_W^2} \left(\alpha_{k,j} \left(S_{k,j} + N_{k,j} \right) + W_{k,j} \right) \right) \right] \right) \\ &\qquad = c\hat{P}. \end{aligned} \tag{3.48}$$

Since each CH is capable to adjust its transmitted power level, then $\hat{P} \triangleq P_k$, i.e., the transmitted power of $CH_k$for data forwarding to the sink node . Hence, the scaler, $\alpha_{k,i}$, can be simplified as follows:

$$\alpha_k = \sqrt{\frac{P_k\, {m_k}^2}{\sum_{i=1}^{m_k} \frac{{\alpha_{k,i}}^2 \sigma_S^4}{P_{k,i} + \sigma_W^2} \left(1 + \sum_{j=1}^{m_k} \frac{{\alpha_{k,j}}^2 \sigma_S^2}{\left(P_{k,j} + \sigma_W^2\right)} \rho(S_{k,i} S_{k,j})\right)}}. \tag{3.49}$$

where $\rho(S_{k,i}, S_{k,j})$ is the correlation coefficient between the readings of $CM_{k,i}$ and $CM_{k,j}$. For simplicity, the scaler, α_k, will be written as:

$$\alpha_k = \sqrt{\frac{P_k}{var(S_k)}} = \sqrt{\frac{P_k}{\sigma_{Sk}^2}}. \tag{3.50}$$

The sink node then tries to estimate the readings of each cluster, S_k, with MMSE estimator in a form of linear transformation. Hence, the estimation of each cluster's readings, Z_k is given as:

$$Z_k = a\, Y_k. \tag{3.51}$$

According to the MMSE estimator, Z_k has to be equal to S_k, which it can be written in the form:

$$E[(Z_k - S_k)^2] = 0. \tag{3.52}$$

Then, the derivative with respect to a is applied before the expectation as follows:

$$E\left[\frac{d}{da}\left(a^2\, {Y_k}^2 - 2\, a\, Y_k\, S_k + {S_k}^2\right)\right] = 0. \tag{3.53}$$

Since the means of source S_k and noise N_k are zeros, it can easily see that a can be written as shown:

$$a = \frac{E[Y_k\, S_k]}{E[{Y_k}^2]}. \tag{3.54}$$

By substituting (3.44) and (3.54) in (3.51), then (3.51) can be defined as:

$$Z_k = \frac{E[Y_k\, S_k]}{E[{Y_k}^2]}\, Y_k = \frac{E[(\alpha_k\, S_k + g_k) S_k]}{E[(\alpha_k\, S_k + g_k)^2]}\, Y_k = \frac{\alpha_k\, \sigma_{Sk}^2}{P_k + \sigma_g^2}\, Y_k. \tag{3.55}$$

Then, the estimated version $\hat{S}_R$ of original event S is defined by averaging all information of CHs, Z_k, with number c ,and it is demonstrated as:

$$\hat{S}_R = \frac{1}{c}\sum_{k=1}^{c} Z_k. \tag{3.56}$$

Finally, the distortion of two-hop communication model is referred as relayed distortion (D_{relay}). It is calculated at the sink node with respect to the estimated value $\hat{S}_R$ and source event, S, according to:

$$D_{relay} = E\left[\left(S - \hat{S}_R\right)^2\right] = E[S^2] - 2\,E\left[S\,\hat{S}_R\right] + E\left[\hat{S}_R^2\right]. \tag{3.57}$$

The first term, $E[S^2]$, equal to σ_S^2, where the source mean is zero. By substituting (3.28), (3.33), (3.43), (3.44), (3.55), and (3.56) in the second term, $E\left[S\,\hat{S}_R\right]$, then it can be defined as:

$$E\left[S\,\hat{S}_R\right] = E\left[S\frac{1}{c}\sum_{k=1}^{c}\frac{\alpha_k\sigma_{Sk}^2}{P_k+\sigma_g^2}\left(\alpha_k\left(\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{\alpha_{k,i}(\sigma_S^2+\sigma_N^2)}{P_{k,i}+\sigma_W^2}\,\left(\alpha_{k,i}\left(S_{k,i}+N_{k,i}\right)+W_{k,i}\right)\right)+g_k\right)\right]. \tag{3.58}$$

which it can be simplified to:

$$E\left[S\,\hat{S}_R\right] = \frac{\sigma_S^2}{c}\sum_{k=1}^{c}\frac{P_k}{P_k+\sigma_g^2}\left(\frac{1}{m_k}\sum_{i=1}^{m_k}\frac{P_{k,i}}{P_{k,i}+\sigma_W^2}\,\rho(S\,S_{k,i})\right). \tag{3.59}$$

The third term can be defined as:

$$E\left[\hat{S}_R^2\right] = var\left(\hat{S}\right). \tag{3.60}$$

where the mean of $\hat{S}$ is zero. By substituting (3.56) in (3.60), then (3.60) can be stated as shown:

$$var\left(\frac{1}{c}\sum_{k=1}^{c} Z_k\right) = \frac{1}{c^2}\left(\sum_{k=1}^{c} var(Z_k) + \sum_{k=1}^{c}\sum_{l=1}^{c} E[Z_k\,Z_l]\right). \tag{3.61}$$

By substituting (3.28), (3.33), (3.35), (3.42), and (3.44) in (3.61), then (3.61) can be extended as:

$$\frac{1}{c^2}\left(\sum_{k=1}^{c}\frac{P_k\sigma_{Sk}^2}{P_k+\sigma_g^2}+\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{P_k}{P_k+\sigma_g^2}\frac{P_l}{P_l+\sigma_g^2}\frac{1}{m_k}\frac{1}{m_l}\left(\sum_{i=1}^{m_k}\sum_{j=1}^{m_l}\frac{\alpha_{k,i}\sigma_S^2}{P_{k,i}+\sigma_W^2}\frac{\alpha_{l,j}\,\sigma_S^2}{P_{l,j}+\sigma_W^2}E\big[(\alpha_{k,i}\,(S_{k,i}+N_{k,i})+W_{k,i})\,(\alpha_{l,j}\,(S_{l,j}+N_{l,j})+W_{l,j})\big]\right)\right). \tag{3.62}$$

Then, the last term can be simplified as:

$$E[\hat{S}_R^2]=\frac{1}{c^2}\left(\sum_{k=1}^{c}\frac{P_k\sigma_{Sk}^2}{P_k+\sigma_g^2}+\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{P_k}{P_k+\sigma_g^2}\frac{P_l}{P_l+\sigma_g^2}\frac{1}{m_k}\frac{1}{m_l}\left(\sum_{i=1}^{m_k}\sum_{j=1}^{m_l}\frac{P_{k,i}}{P_{k,i}+\sigma_W^2}\,\frac{P_{l,j}}{P_{l,j}+\sigma_W^2}\,\sigma_S^2\,\rho(S_{k,i}S_{l,j})\right)\right). \tag{3.63}$$

Finally, (3.57) can be simplified by using (3.59) and (3.63) as demonstrated below:

$$D_{realy}=\sigma_S^2-\frac{1}{c^2}\sum_{k=1}^{c}\frac{P_k}{P_k+\sigma_g^2}\,\frac{1}{{m_k}^2}\sum_{i=1}^{m_k}\frac{{\alpha_{k,i}}^2\,\sigma_S^4}{P_{k,i}+\sigma_W^2}\left(2\,c\,m_k\,\rho\big(S,S_{k,i}\big)-1-\sum_{j=1}^{m_k}\frac{{\alpha_{k,j}}^2\,\sigma_S^2\rho\big(S_{k,i},S_{k,j}\big)}{\big(P_{k,j}+\sigma_W^2\big)}\right)+\frac{1}{c^2}\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{P_kP_l/m_km_l}{\big(P_k+\sigma_g^2\big)\big(P_l+\sigma_g^2\big)}\sum_{i=1}^{m_k}\sum_{j=1}^{m_l}\frac{{\alpha_{k,i}}^2\,{\alpha_{l,j}}^2\,\sigma_S^6}{\big(P_{k,i}+\sigma_W^2\big)\big(P_{l,j}+\sigma_W^2\big)}\,\rho\big(S_{k,i},S_{l,j}\big). \tag{3.64}$$

where $\rho(S,S_{k,i})$ and $\rho(S_{k,i},S_{l,j})$ are the correlation coefficients between the source event, S, and the readings of $CM_{k,i}$, $S_{k,i}$, and between the readings of $CM_{k,i}$, $S_{k,i}$, and $CM_{l,j}$, $S_{k,i}$, respectively.

The goal of this analytical derivation is to define the accuracy level of the clustered-network from EDC algorithm. Each sensor is required to do specific duty, but when its energy level is less than the predetermined threshold, the sensor changes its state to inactive state (only-harvesting) and based on that the distortion changes. The upper limit of both D_{point} (3.27) and D_{relay} (3.64) reached when all sensor nodes triggers into inactive state. In such a case, the distortions are equal to event variance σ_S^2 as stated below:

$$\sigma_S^2=max_{nodes=\emptyset}\big(D_{relay},D_{point}\big). \tag{3.65}$$

3.5. Energy Model

In this section, we attempt to investigate the mechanism of energy consumption in the sensor networks for single-hop and two-hop communication models. In both cases, a sensor node is either in active or inactive state. An active node assists in operation of the network for sensing the area or sending the data to sink. In an inactive state, sensor nodes stop sensing and sending data until they recovered their energy by recharging their batteries or by energy harvesting.

The energy consumption in the active nodes is composed of four components. These components can be defined as energy consumed for packet transmitting, packet receiving, data gathering (aggregating) and event sensing. The simplified model of energy consumption for each part is defined in this section.

3.5.1. Energy Consumption for Clustered Network

In clustered-network model Fig 3.3, sensor nodes are assembled into clusters, where in each cluster consists of a single CH, and several CMs with number m_k. Based on the EDC algorithm, CH requires to receive the observations from its CMs, and aggregates their readings to transmit towards the sink node by single-hop communication.

The required energy for transmitting and receiving L bits over the distance d between the transmitter and the receiver can be modeled according to the first order radio (Fig 3.6) model [8] defined as:

$$E_{Tx}(L,d) = \begin{cases} (E_{elec}\, L) + (\varepsilon_{fs}\, L\, d^2), & d \le d_0 \\ (E_{elec}\, L) + (\varepsilon_{mp}\, L\, d^4), & d > d_0 \end{cases} \quad (3.66)$$

$$E_{Rx}(L) = E_{elec}\, L.$$

where, $d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$.

The E_{elec} is electronic energy. The ε_{fs} and ε_{mp} are coefficients in free space model and in multi-path model, respectively. Furthermore, sensing energy for L bits can be defined as follows:

$$E_{sens}(L) = L\, T_{sens}\, I_{sens}\, V_{sub}. \quad (3.67)$$

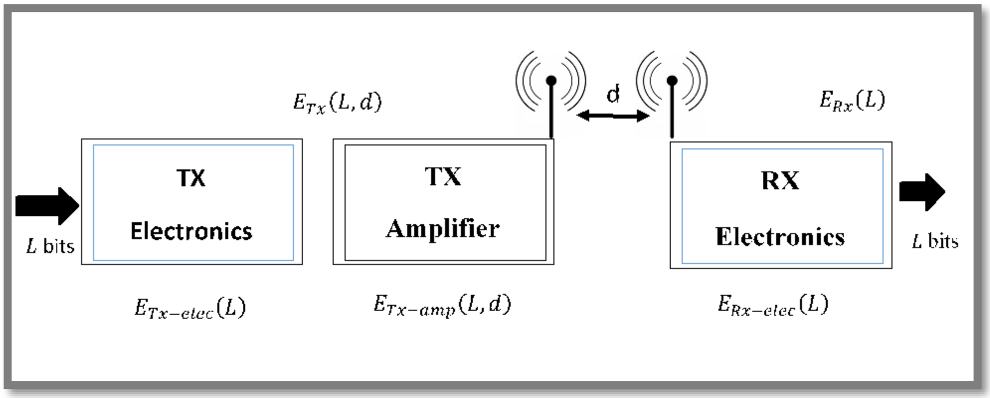


Figure 3.6. First Order Radio Model

where I_{sens} is sensing current, V_{sub} is supplied voltage and T_{sens} is sensing time. The aggregating energy form m receiving destinations can be modeled as [8]

$$E_{agg}(L, m) = L\, m\, E_{DA}. \tag{3.68}$$

where E_{DA} is aggregation energy. Based on the above equations, it is possible now to define the energy consumption of $CM_{k,i}$ and CH_k which can be demonstrated as:

$$E_{CM}(k, i) = E_{sens}(L) + E_{Tx}(L, d_{ch}(k, i)), \tag{3.69}$$

$$E_{CH}(k) = E_{agg}(L, m_k) + m_k\, E_{Rx}(L) + E_{Tx}(L, d_{BS}(k)). \tag{3.70}$$

where $d_{ch}(k, i)$ and $d_{BS}(k)$ are Euclidean distance between $CM_{k,i}$ and CH_k, and between CH_k and sink node, respectively. Furthermore, the energy consumption for single-hop communication case is modeled as:

$$E_{n_i}(i) = E_{sens}(L) + E_{Tx}(L, d_{BS}(i)) \tag{3.71}$$

where $d_{BS}(i)$ is the euclidean distance between n_i and sink node. The above equations in (66) represent the energy consumption for transmitting in joules.

However, the unit of power constraints (i.e., transmitting power, that is defined in point distortion (17) and relayed distortion (64)) is Watts. Also, the unit of above energy consumption functions is Joules. Hence, a simple conversion method is used for finding the power constraints by converting to Watts, which is ($P(watts) = E(joules)/T$), with respect to the time durations (T_{sens}, T_{agg}, and T_{trans}) as shown below:

$$T_{agg} = L\ \frac{N_{cyc}}{f_{sen}}, \qquad T_{tran} = T_{rec} = \frac{L}{R}. \tag{3.72}$$

where N_{cyc} is the number of processor cycles per bit, f_{sen} is sensor frequency and R is the data rate.

3.5.2. Harvesting Model

Energy harvesting is an interesting solution for battery recharging when nodes are positioned in an event area covered by a cellular base station. By assuming that there is radio frequency source such as GSM900 tower [13] with transmitted power P_t and antenna gain G_t. Then, it's possible from each sensor node to harvest energy from it by equipping an additional harvesting circuit with additional antenna of gain G_r. Therefore, by using Friis equation it's possible to model the receiving power, P_R, considering the free path loss model as shown:

$$P_r = P_t \frac{G_T G_R \lambda^2}{(4\pi d)^2}. \tag{3.73}$$

where λ is wave length over d distance. The harvested time, T_{har}, can be defined with respect to circuit harvesting efficiency and battery charging time.

3.6.Case Study

In this section, the point (27) and relayed (64) distortion functions are studied based on some important parameters (data rate, number of sensor nodes, event distance, and the size of event area for different values of θ_1) by means of Monte Carlo simulation (i.e., 500 iterations).

3.6.1. Data Rate (R)

The presented two distortion functions are observed by varying the data. According to (72), increasing data rate decreases transmission time, T_{tran}. It is possible to convert the transmitting energy, $E_{Tx}(L, d)$, to transmitting powers (i.e., power constraints) by dividing them with T_{tran}. Since the power constraints of the two distortion functions are transmitting powers (i.e., $(P_{k,i}, P_k)$ for D_{relay} and P_i for D_{point}), the distortion functions are affected also by the transmission time, T_{tran}. Hence, increasing the data rate, R, improves the distortion functions, (D_{relay} & D_{point}), i.e., the reliability or accuracy level of the entire network as shown in Fig 3.7.

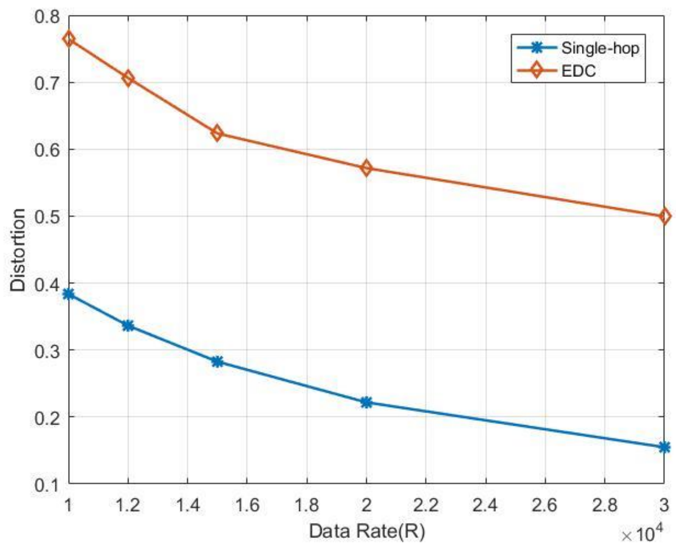


Figure 3.7 Distortion Vs Data Rate

3.6.2. Number of Sensor Nodes

Increasing the number of sensor nodes affects positively the distortion functions but until 100 sensors. In other words, by fixing the event area $(100 \times 100\ m^2)$, when more than 100 sensor nodes are deployed in the event area, relatively similar distortion levels are observed in the sink node as shown in Fig 3.8. The reason for that is related to the density property for WSN. Since the distortion isn't changed after increasing the nodes more than 100, the 100 sensor nodes are already present a dense deployment, and the nearby sensors create redundant readings. Hence, there is no new information about source event, even if more sensors are deployed.

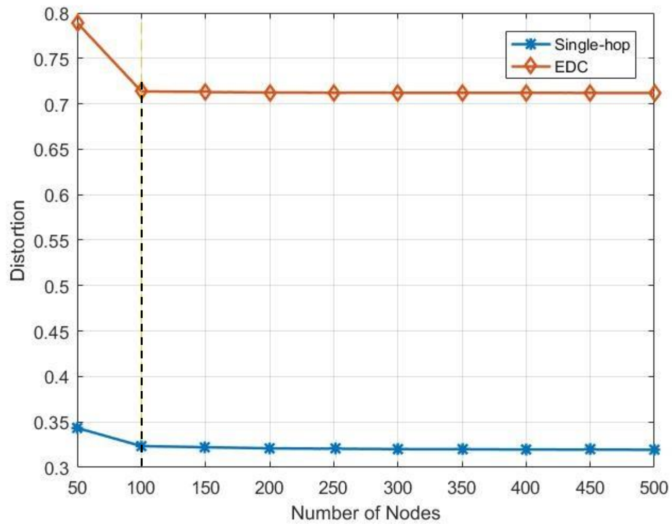


Figure 3.8 Distortion Vs Number of Nodes

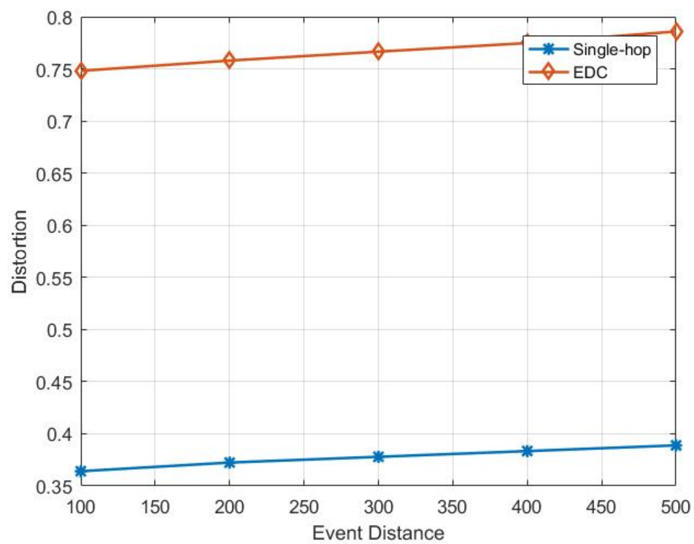


Figure 3.9 Distortion Vs Event Distance

3.6.3. Event Distance

The correlation coefficients $\left(\rho\left(S, S_{k,j}\right)\right.$ for D_{relay} and $\rho(S, S_i)$ for $\left.D_{point}\right)$ are mainly depend on the distance between event source, S, (i.e., point source that is located at n_0) and sensor nodes, n_i, Consequently, by setting 100 sensors, increasing the event distance, i.e., the distance between source event and sensor nodes (in meter), effects on the distortion functions negatively. In other words, if the sensor nodes are selected apart from the event source, they observe relatively inaccurate readings, resulting higher distortion level at the sink node for both cases, as shown in Fig 3.9.

3.6.4. Event Area

Changing the size of the event area affects in the distortion functions. In other words, when the area is increased, the separation between sensor nodes are also increased. Hence, the correlation coefficients between sensor readings (i.e., $\rho(S_{k,i}\ S_{k,j}), \rho(S_{k,i}S_{k,l})$ for D_{relay} and $\rho(S_i\ S_j)$ for D_{point}) are changed also. Because they depend on the separation distances between sensor nodes (for single-hop case) and between CMs (for two-hop case, as a result from EDC algorithm).

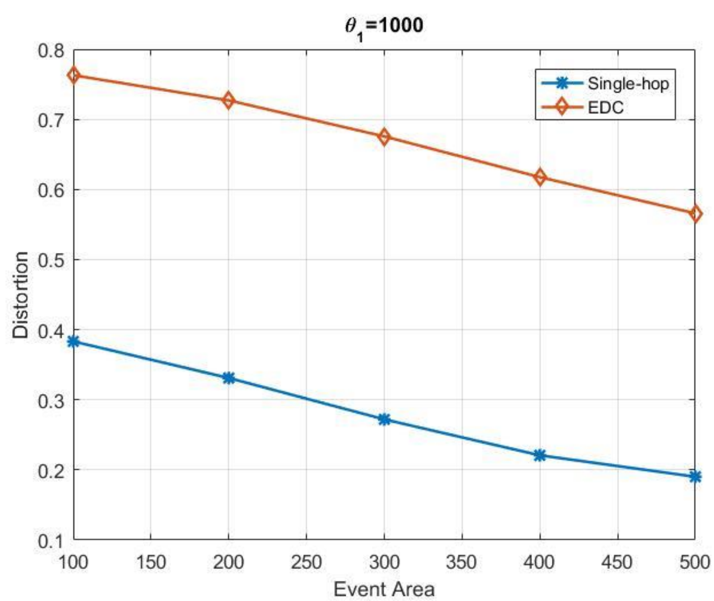


Figure 3.10 Distortion Vs Event area

Furthermore, in this analysis, power exponential model is used, which it defined in equations (4), (30) and (31). By setting $\theta_1 = 1000$, as in [3], when the correlation range is increased, the distortions is improved. In other words, when the distance between the sensor nodes increases, the distortion decreases. Because of the further apart nodes observe less correlated data. Hence, less correlated observations are more favorable in WSNs.

CHAPTER 4

SIMULATIONS

In this section, the proposed EDC clustering algorithm is simulated in MATLAB to examine its detailed performance with respect to the single-hop communication model. For this purpose, simulation results in different network scopes are reported with different parameters to observe the performance of the proposed algorithm. The simulation is applied in terms of distortion, energy consumption rate and network lifetime. Furthermore, the influence of the harvesting energy on the network performance is presented with respect to the same above terms.

4.1.Setup Parameters

In this section, some important parameters are defined. A total of 100 sensor nodes are deployed randomly in 100×100, 200×200 and 300×300 square meters event area, based on the uniform distribution. Changing event area, changes the density degree of nodes' deployment. Furthermore, each sensor node requires to forward its observations to the sink node continuously via single-hop or two-hop communication links. The sink node is assumed to be located at the center of the event area. Furthermore, a cell tower, i.e., GSM900, can be used as a renewable energy source. This tower is assumed to be located at the center of event area as a harvesting source for all sensor nodes. Then each node can harvest energy, besides its main duties.

Each sensor node has an additional circuit for energy harvesting similar to the presented circuit in [13]. Hence, each sensor node has two antennas, one for harvesting and the other one for communicating with others. The harvesting energy is available only if sensor nodes are located within specific threshold distance, d_0, (due to free path loss model). Without physical damages, sensor nodes run out from their energies with each iteration (time) with respect to the proposed energy model in Chapter 3. The simulation parameters are introduced in Table 4.1.

4.2.EDC algorithm

At the first stage of EDC algorithm, the sink node allocates all locations of sensor nodes and estimates their receiving samples based on single-hop communication and determines the resulted point distortion for all N nodes, $D_{point}(N)$. After $D_{point}(N)$ is calculated, the sink node is applied Vector Quantization (VQ). All sensor nodes are assumed to be uniformly distributed in event area.

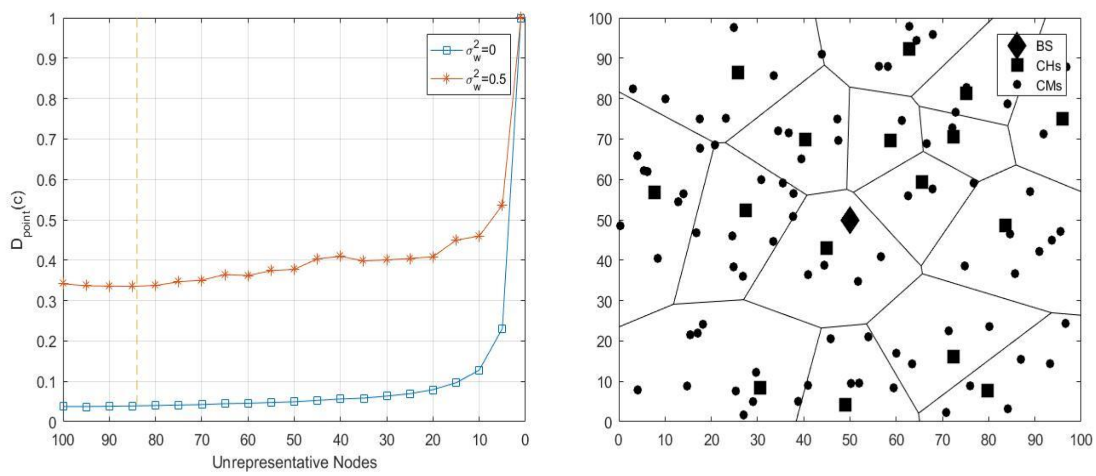
Table 4.1. Simulation Parameters

Parameters	Values
Area	(100×100, 200×200 and 300×300) m^2
Number of sensors, N	100
Sink node and GSM900 locations	(50,50)
Initial energy	0.5 J
E_{elec}	50 nJ/bit
E_{DA}	5 nJ/bit
d_0 threshold distance	87 m
ε_{fs}	10 PJ/bit/m^2
ε_{mp}	0.0013 PJ/bit/m^4
L data packet size	1000 bytes
L-broadcast packet	25 bytes
Center frequency for harvesting GSM 900	950 MHZ
G_t , G_r	17 dB , 9 dB

Hence, VQ is initiated with all sensor nodes locations as a vector source, code vector length is one. According to EDC algorithm, the sink node starts to increase the code word length until the remaining nodes (unrepresentative nodes) achieve distortion constraints as shown in Fig 4.1. The maximum achievable distortion is defined according to:

$$D_{Point}(N) + D_{thresold} \geq D_{constraint}. \tag{4.1}$$

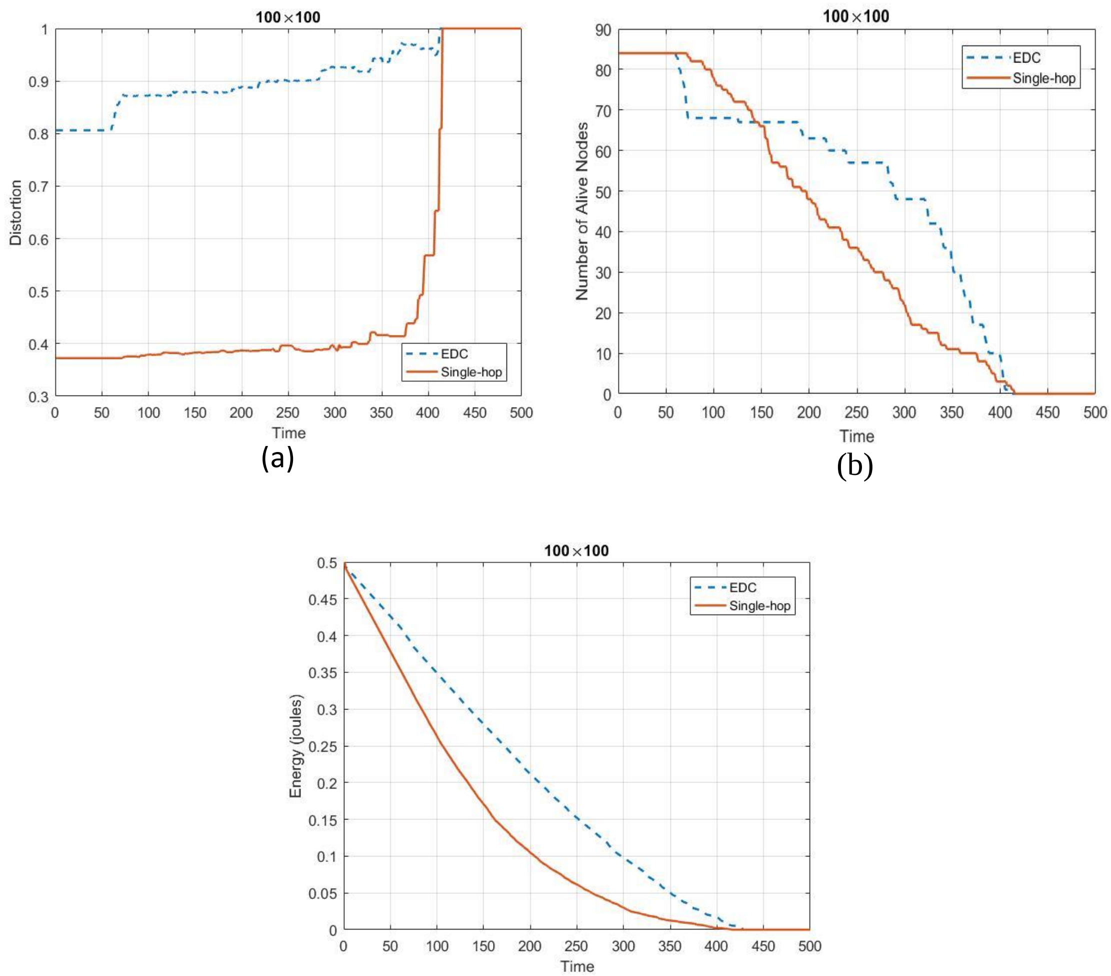
where $D_{Point}(N)$ is the minimum distortion achieved when all the sensor nodes in the event area send information to the sink [3]. And $D_{thresold}$ is threshold distortion for single-hop case, which it is assumed to be equal to 0.001 for noiseless ($\sigma_W^2 = 0$) and noisy ($\sigma_W^2 = 0.5$) channel cases. Approximately, in both cases it produces similar results. But in noisy case, the distortion is lifted up as shown in Fig 4.1 (a).



(a) (b)

Figure 4.1 Network Clustering

After repeating the experiment over 100 trials (Monte Carlo), EDC approximately selects 84 sensors nodes as unrepresentative nodes (CMs) and 16 representative nodes (CHs). Then, the representative nodes (CHs) forms Voronoi diagrams (clusters) as shown in Fig 4.1 (b). Then, Fig 4.1 (b) represents a snapshot for the resulted network with the sink node which is located at the center of event area. This network uses two-hop communication model, and hence, we can use a relayed distortion function (64) to define the accuracy level. The clustered-network is then compared with single-hop communication model. In single-hop case, the VQ is also applied for sensor nodes and the result is unrepresentative nodes that require to forward their observations. In $100 \times 100\ m^2$ event area, the distortion for single-hop, $D_{Point}(c)$, is less than the distortion for two-hop (EDC) case, D_{relay} (64). Furthermore, when the number of iterations increased, some sensors may run off from energy and hence distortion is increased. Besides for both cases, the network terminates at the same time. Hence, single-hop communication provides a better performance than two-hop (EDC) case as illustrated in Fig 4.2 (a). Also with respect to the number of alive nodes and energy consumption, single-hop provides similar performance to EDC case as shown in Fig 4.2 (b), (c). The simulation is performed for 500 seconds. The energy consumption is calculated at each iteration with respect to equations (3.69), (3.70) and (3.71).



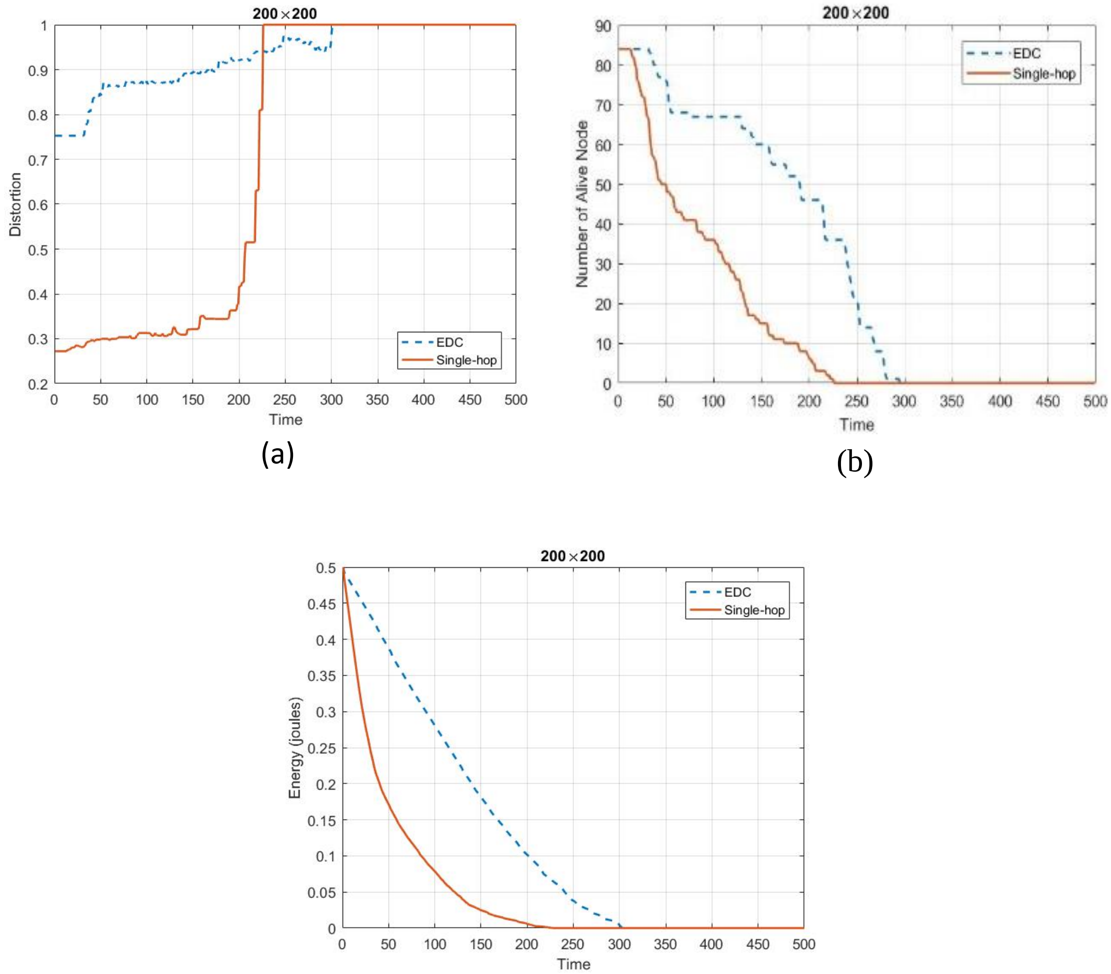
(c)

Figure 4.2. Distortion, Number of Alive nodes and Energy Consumption for $100 \times 100\ m^2$ Event Area

This similarity is observed because of the total distances of the forwarded data through single-hop case is approximately equal to the two-hop (EDC) case, as shown:

$$\sum_{i=1}^{N} d_{BS}(i) \cong \sum_{k=1}^{c} \left(\sum_{i=1}^{m_k} d_{CH}(k,i) + d_{BS}(k) \right). \tag{4.2}$$

where $d_{BS}(i)$ is a distance between node n_i and sink node. Besides, $d_{CH}(k,i)$ and $d_{BS}(k)$ are the distance between $CM_{k,i}$ and CH_k and between CH_k and sink node, respectively.

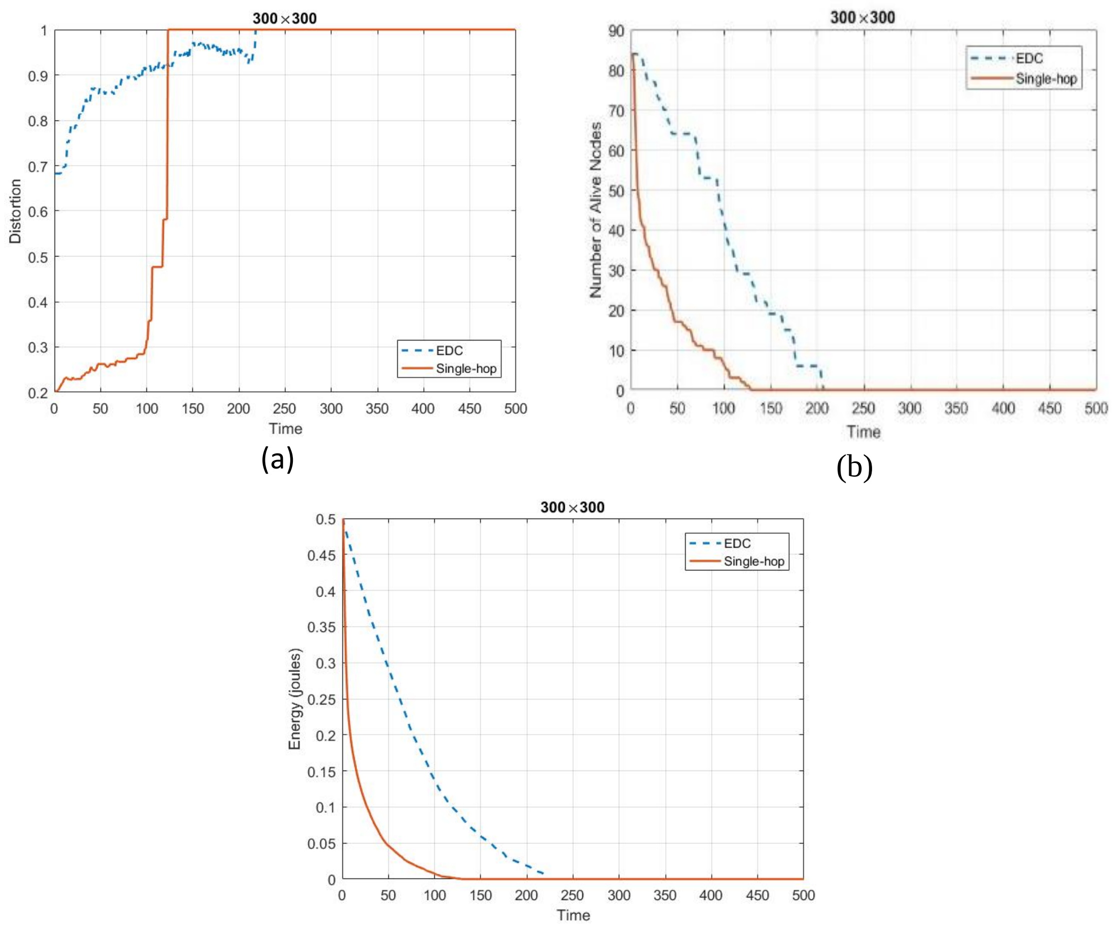


(a) (b) (c)

Figure 4.3 Distortion, Number of Alive nodes and Energy Consumption for $200 \times 200\ m^2$ Event Area

On the other hand, when the size of event area is $200 \times 200\ m^2$, EDC algorithm is provided a better performance than single-hop case. The reason is that the $D_{point}(c)$ is decreasing faster that D_{relay}, until all sensors expired as shown in Fig 4.3, because the total distances to transmit data for single-hop case is greater than EDC case, as shown:

$$\sum_{i=1}^{N} d_{BS}(i) > \sum_{k=1}^{c} \left(\sum_{i=1}^{m_k} d_{CH}(k,i) + d_{BS}(k) \right). \tag{4.3}$$



(c)

Figure 4.4 Distortion, Number of Alive nodes and Energy Consumption for $300 \times 300\ m^2$ Event Area

Similarly, in case of $300 \times 300\ m^2$ event area, the performance of EDC is much better than single-hop case as in Fig 4.4. As a result, EDC is considered to be used for event areas with sizes are greater than $100 \times 100\ m^2$ and single-hop is considered to be used for only $100 \times 100\ m^2$ because of its advantage in distortion level.

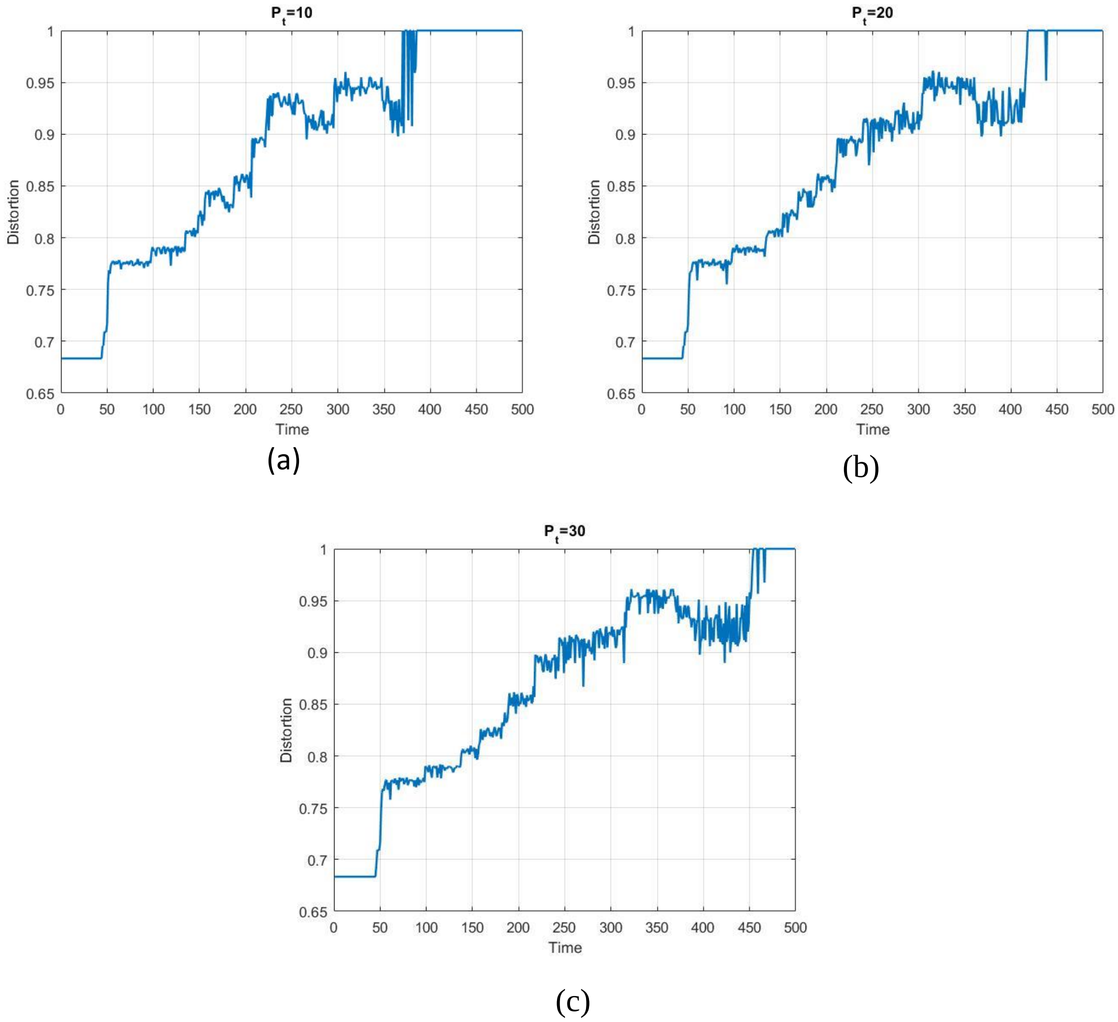
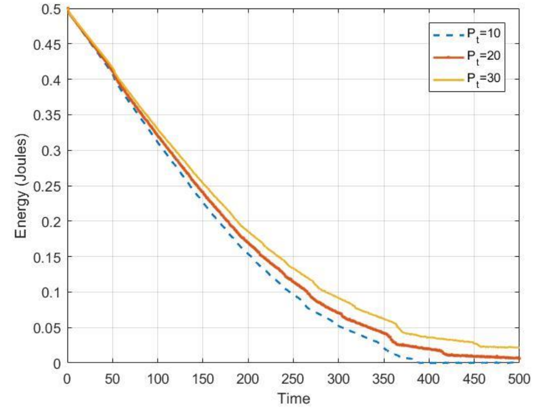
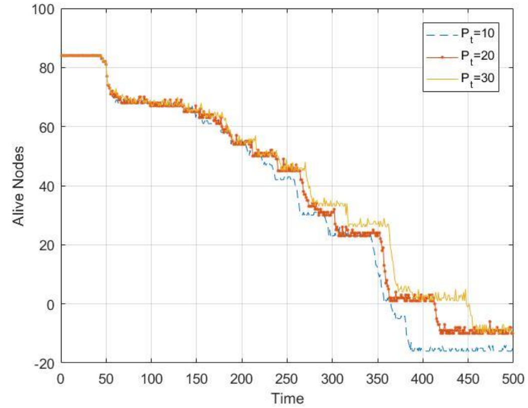


Figure 4.5 Distortion for Different Harvested Energies

Furthermore, by fixing the size of event area, another investigation is performed, where a different harvesting energy levels (i.e., transmitted power, P_t, form cell tower [13]) is applied for each sensor node. Because of harvesting, an inactive sensor node can be recharged and be active again, hence the distortion changes iteratively. Then, the distortion level improved and made it less than the upper bound as possible as shown in Fig 4.5. The upper bound in this simulation is set to be equal to one, i.e., it is the variance of source event σ_S^2. Furthermore, this improvement is happened because of the energy consumption rate, as shown in Fig 4.6 (a), is reduced at each iteration and it aided the number of nodes to be alive for longer time which is shown in Fig.4.6 (b).



(a)



(b)

Figure 4.6 Energy Consumption and Number of Alive nodes for Different Harvested Energies

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1.Conclusion

In both military and civil applications, the WSN is highly required. In most applications, densely deployment has been used because of it is inexpensive to be deployed. However, in densely deployment, the nearby sensors are spatially correlated, and hence their readings are redundant, which causes wasting in energy and increasing overheading packets. The EDC algorithm is presented to address these challenges by cluster the network based on exploiting spatial correlations between sensor nodes through Vector Quantization. The main contribution from exploiting spatial correlation is to ban the readings of cluster heads and thus their duties are reduced and the network's life time is improved. As a part of this algorithm, a theoretical framework defining distortion for single-hop and two-hop commutations model has been derived.

Additionally, the EDC algorithm has been presented for energy harvesting sensors. Energy harvesting is enhancing the accuracy level. In other word, when sensor nodes harvest energy, their life time is improved. Unlike non-harvesting case, when the network loss the accuracy when all nodes die.

5.2.Future Works

The power constraints are not random. However, by changing this assumption and model the power constraints randomly, another model it can be derived for distortion function and another results are expected. Energy harvesting technology has been used in WSN and hence it is good idea if the distortion function is connected with available harvesting power. By developing another form for distortion for multi-hop communication model, it is possible to compare EDC algorithm with other clustering algorithms.

BIBLIOGRAPHY

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey” *Computer Networks, vol. 38, pp. 393–422, 2002*.

[2] S. Akbari, “Energy harvesting for wireless sensor networks review,” *in Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on, Sept 2014, pp. 987–992*.

[3] M. C. Vuran and I. F. Akyildiz, “Spatial correlation-based collaborative medium access control in wireless sensor networks” *IEEE/ACM Transactions on Networking, vol. 14, no. 2, pp. 316–329, April 2006*.

[4] M. Gastpar and M. Vetterli, “Source-Channel Communication in Sensor Networks,” *Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp.162–177. [Online]*. Available: http://dx.doi.org/10.1007/3-540-36978-3 11

[5] Z. Liu, W. Xing, B. Zeng, Y. Wang, and D. Lu, “Distributed spatial correlation-based clustering for approximate data collection in wsns,” *in Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on, March 2013, pp. 56–63*.

[6] G. Zheng and S. Tang, “Spatial correlation-based mac protocol for event-driven wireless sensor networks*,” Journal of Networks, vol. 6, no. 2, 2011*.

[7] S. K. Gupta, N. Jain, and P. Sinha, “Clustering protocols in wireless sensor networks: A survey,” *International Journal of Applied Information Systems, vol. 5, no. 2, pp. 41–50, January 2013, published by Foundation of Computer Science, New York, USA.*

[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy efficient communication protocol for wireless microsensor networks,” *in System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, Jan 2000, pp. 10 pp. vol.2–.*

[9] M. Ye, C. Li, G. Chen, and J. Wu, “EECS: an energy efficient clustering scheme in wireless sensor networks,” *in PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference, 2005., April 2005, pp. 535–540.*

[10] S. Yoon and C. Shahabi, “Exploiting spatial correlation towards an energy efficient clustered aggregation technique (cag) [wireless sensor network applications],” *in IEEE International Conference on Communications, 2005. ICC 2005. 2005, vol. 5, May 2005, pp. 3307–3313 Vol. 5*.

[11] L. A. Villas, A. Boukerche, D. L. Guidoni, H. A. de Oliveira, R. B. de Araujo, and A. A. Loureiro, “An energy-aware spatio-temporal correlation mechanism to perform efficient

data collection in wireless sensor networks," *Comput. Commun., vol. 36, no. 9, pp. 1054–1066, May 2013. [Online]. Available:* [http://dx.doi.org/10.1016/j.comcom. 2012.04.007](http://dx.doi.org/10.1016/j.comcom.2012.04.007)

[12] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications, vol. 28, no. 1, pp. 84– 95, Jan 1980.*

[13] M. Arrawatia, M. S. Baghini, and G. Kumar, "Rf energy harvesting system from cell towers in 900mhz band," *in Communications (NCC), 2011 National Conference on, Jan 2011, pp. 1–5.*

[14] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *IEEE Transactions on Information Theory, vol. 51, no. 3, pp. 765–779, March 2005*.

[15] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," ACM Trans. Sen. Netw., vol. 8, no. 1, pp. 7:1–7:31, Aug. 2011. [Online]. Available: http://doi.acm.org/10.1145/1993042.1993049

[16] I. F. Akyildiz and E. P. Stuntebeck, "Wireless underground sensor networks: research challenges," *Ad Hoc Networks, vol. 4, no. 6, pp. 669–686, 2006*.

[17] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M. Welsh, "Deploying a wireless sensor network on an active volcano," *IEEE Internet Computing, vol. 10, no. 2, pp. 18–25, March 2006.*

[18] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer. Network., vol. 51, no. 4, pp. 921–960, Mar. 2007. [Online]. Available:* [http://dx.doi.org/10.1016/ j.comnet.2006.10.002](http://dx.doi.org/10.1016/j.comnet.2006.10.002).

[19] M. Khelifi, I. Benyahia, S. Moussaoui, and F. Nat-Abdesselam, "An overview of localization algorithms in mobile wireless sensor networks," *in 2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), July 2015, pp. 1–6.*

[20] H. Huang, S. Xiao, X. Meng, and Y. Xiong, "A remote home security system based on wireless sensor network and gsm technology," *in Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference on, vol. 1, April 2010, pp. 535–538.*

[21] M. T. Lazarescu, "Design of a wsn platform for long-term environmental monitoring for iot applications," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 3, no. 1, pp. 45–54, March 2013*.

[22] E.M. Petriu, N.D. Georganas, D.C. Petriu, D. Makrakis, V.Z. Groza, "Sensor-based information appliances," *IEEE Instrumentation and Measurement Magazine (December 2000) 31–35.*

[23] M. P. urii, Z. Tafa, G. Dimi, and V. Milutinovi, “A survey of military applications of wireless sensor networks,” *in 2012 Mediterranean Conference on Embedded Computing (MECO), June 2012, pp. 196–199*.

[24] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, January 1967*.

[25] S. Toumpis and T. Tassiulas, “Optimal deployment of large wireless sensor networks,” *IEEE Trans. Information Theory, vol. 52, pp. 2935–2953, 2006.*

[26] I. F. Akyildiz and E. P. Stuntebeck, “Wireless underground sensor networks: research challenges,” *Ad-Hoc Networks, vol. 4, pp. 669–686, 2006*.

[27] M. Li, Y. Liu, “Underground structure monitoring with wireless sensor networks,” *Proc. the IPSN, Cambridge, MA, 2007.*

[28] I. F. Akyildiz, D. Pompili and T. Melodia, “Challenges for efficient communication in underwater acoustic sensor networks,” *ACM Sigbed Review, vol. 1(2), pp. 3–8, 2004*.

[29] J. Heidemann, Y. Li, A. Syed, J. Wills and W. Ye, “Underwater sensor networking: research challenges and potential applications,” *Proc. the Technical Report ISI-TR-2005-603,* USC/Information Sciences Institute, 2005.

[30] I. F. Akyildiz, T. Melodia, K.R. Chowdhury, “A survey on wireless multimedia sensor networks,” *Computer Networks,* vol. 51, pp. 921–960, 2007.

[31] N. Bulusu, D. Estrin, L. Girod, J. Heidemann, “Scalable coordination for wireless sensor networks: self-configuring localization systems,” *Proc. Int. Symposium on Communication Theory and Applications (ISCTA 2001),* Ambleside, UK, July 2001.

[32] J. O. Berger, V. de Oliviera, and B. Sanso, “*Objective Bayesian analysis of spatially correlated data*,” J. Amer. Statistical Assoc., vol. 96, no.456, pp. 1361–1374, 2001.

[33] G. L. Stuber, *Principles of Mobile Communication*. Norwell, MA:Kluwer, 2001.

[34] V. Poor, *An Introduction to signal detection and estimation*, 2^{nd} Edition, Springer-Verlag, 1994.