A

Dissertation On

"An Efficient Spatial and Temporal Co-relation Model for Wireless Sensor Networks"

Submitted in Partial Fulfillment of the Requirement
For the Award of Degree of

Master of Technology

in

Software Technology

by

Abhishek Gupta
University Roll No. 2K12/SWT/01

Under the Esteemed Guidance of

Mr. Rajesh Kumar Yadav
Assistant Professor, Computer Science & Engineering, DTU

2012-2015

COMPUTER SCIENCE & ENGINEERING DEPARTMENT
DELHI TECHNOLOGICAL UNIVERSITY
DELHI – 110042, INDIA

CERTIFICATE

Delhi Technological University
(Government of NCT, Delhi)
Bawana Road, New Delhi-42

This is to certify that the thesis entitled "An Efficient Spatial and Temporal Co-relation Model for Wireless Sensor Networks" done by ABHISHEK GUPTA (Roll Number: 2K12/SWT/01) for the partial fulfillment of the requirements for the award of degree of Master of Technology in Software Technology in the Department of Computer Science Engineering, Delhi Technological University, New Delhi is an authentic work carried out by him under my guidance.

Project Guide:
Mr. Rajesh Kumar Yadav
Assistant Professor
Department of Computer Science Engineering
Delhi Technological University, Delhi

ACKNOWLEDGEMENT

I take this opportunity to express my deep sense of gratitude and respect towards my guide Mr. Rajesh Kumar Yadav (Assistant Professor) Department of Computer Science & Engineering.

I am very much indebted to him for his generosity, expertise and guidance I have received from him while working on this project. Without his support and timely guidance the completion of the project would have seemed a far –fetched dream. In this respect I find myself lucky to have my guide. He have guided not only with the subject matter, but also taught the proper style and techniques of documentation and presentation.

Besides my guides, I would like to thank entire teaching and non-teaching staff in the Department of Computer Science & Engineering, DTU for all their help during my tenure at DTU. Kudos to all my friends at DTU for thought provoking discussion and making stay very pleasant.

ABHISHEK GUPTA
M.Tech Software Technology
2K12/SWT/01

Abstract

A wireless sensor network (WSN) is set of distributed autonomous devices called sensors for monitoring properties like motion, vibration, pressure, temperature, sound, pollutants at different locations.

WSN is a set of sensors, located in a remote location equipped with sensing and computational and capabilities. Along with sensors, each node in WSN is mounted with a radio transceiver, an energy source (battery) and a small microcontroller. These low power and low cost devices form a set for monitoring the area of interest. Using the collaboration of sensors, the Wireless Sensor Nodes collects and sends information about the area of interest (e.g. surveillance, rain, humidity, temperature etc.) to the Processing Node. Then received information is processed at Processing Node and used by users.

Due to limited energy capacity of nodes, network overall lifetime will not be long. Giving continuous power supply to nodes or changing the used battery is a difficult task in such remote location. This could interrupt or badly affect the flow of information from area of observation. Major portion of the energy consumption of sensor node in the sensor network is happens in the information flow (transmission and reception), and only very few portion in other activities (clustering, broadcasting etc). So to increase the network lifetime once can target to make the sensor nodes into sleep mode as frequent as possible without significant loss in information. If sensor nodes are deployed densely in the area of observation there are chances of getting similar information from the neighboring nodes. Also if the frequency of sending information to the processing node by the

sensor node is high, again there are chances of observing similar information by the same node. This means observed information can be Spatially and Temporally correlated information will be observed by the deployed nodes. Thus Spatial and Temporal correlation provides a basis for proposing and developing techniques for energy saving of sensor nodes.

We are proposing a hybrid approach to achieve target of battery saving and the approach is based on Spatial and Temporal Co-relation among various sensor nodes. On sensing a physical phenomenon in an area all nodes start sending observed information to the Processing Node. In a highly dense deployed sensor networks there are lot of chances that these nodes will be sending Spatially and Temporal correlated information. Thus it is not necessary that all deployed nodes send information to the Processing Node, and some nodes which are sending spatially correlated information can be made to sleep mode.

Network is divided into number of clusters, which we have taken as 5% of the total number of nodes of a network. Nodes are assigned to the cluster having minimum distance to the cluster head having maximum energy. The distance between nodes is Euclidean Distance.

Simulations results show that our proposed technique which is based on clustering and exploiting Spatial and Temporal Correlation can lower the energy consumption of data transmission and thus increasing node and overall network lifetime.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 Introduction

1.1 General Concepts

What Is Sensor

Sensor is a device which detects some type of physical properties from its surroundings. Physical properties it detects can be light, heat, motion, moisture, pressure. It is basically a transducer which detects change in physical properties and provides the result as output. Physical properties are converted to signal which can be measured.

Sensor Types:

1. Physical property to be measured
2. Transduction principles
3. Area of usage
4. Physical characteristic
5. Technique involved
6. Used content to manufacture it

Applications such as Surveillance have a great limitation of sensor energy, as sensors are battery operated. It is always very difficult to change battery or to charge battery in such applications. So we need some methods to save sensor energy.

What is a Sensor Network?

A wireless sensor network (WSN) is a set or group of smart devices with limited processing and power capacity and have the capabilities of monitoring and low power consumption to monitor the area of interest and send observed environmental and physical measurements, related to temperature, sound, pressure, etc to the Processing Node.

A wireless sensor network (WSN) is a collection of randomly deployed sensor nodes of hundreds to thousands in numbers. Long sensor lifetime & overall network life time, shared responsibility of entire area of observation, are some considerable requirements in Wireless Sensor Networks. Forming groups of sensor nodes based on their locations: Clustering provides a very good solution to fulfill these requirements. Various clustering algorithms also differ in their objectives.

Every sensor node in sensor network node has following parts:

- Transceiver,
- Energy source
- Communication circuit for sensors and an energy source

Sensors come in various sizes from few millimeters to few centimeters. Depending on the application, area of usage accuracy many other factors cost of Sensor nodes may vary from low to very high. There are some limitations such as storage; computational capabilities, continuous supply and transmission capabilities, which are result of constraints in cost and size of sensor nodes. Wireless sensor network architecture can be either random distribution to an organized multi-hop network. The communication between the nodes can be guided or unguided.

Characteristics

Wireless sensor networks exhibits following characteristics:

- Small in size
- Intelligent devices
- Limited Memory

- Capable of data processing
- Heterogeneity
- Moving Nodes
- Self organized within network
- Deployed in rough environmental condition
- Battery constraint
- Cross-layer design

Cross-layer design in Wireless Sensor Networks is an important area for researchers these days. Main reason behind this is the limitation in conventional layered design:

1. The conventional designs are unable to cop-up with changing conditions.
2. Each layer cannot have whole data. They cannot make sure the optimization of the network.
3. Not applicable for WSNs as the interference between the various users, access confliction, fading, and the change of environment.

So to enhance the transmission performance, such as QoS (Quality of Service) high data rates, energy efficiency, etc the cross-layer designs are used to achieve best optimization. Sensor nodes can be viewed as tiny intelligent devices which can perform basic data processing tasks and have limited HW.

Processing Nodes

The Processing Nodes can be considered as bridge between the person who needs the data (actual user) and deployed sensors. Processing Nodes consists of single or multiple units, which together form Processing Node and are integral part of Wireless Sensor Networks. Processing Nodes are rich in power and other resources. Routers are also one of important component, are used to accomplish the routing between nodes. Using routers data can be forwarded from the Wireless Sensor Networks to a server.

Types of Sensor Networks

Based on application there are two types of networks as below:

Proactive Networks

The nodes in Proactive network observe the physical properties and send the observed data of and periodically switch on their sensors and transmitters. By this, they provide a blueprint of the required components with their standard values periodically.

Application which requires data monitoring at regular frequency are best example for Proactive Networks.

Reactive Networks

Whenever there are considerable changes observed in measuring physical property reactive networks transmits such information to the processing node.

Time critical applications are the best example for this type of applications.

1.2 Clustering in wireless sensor network

As we have already discussed that in a densely deployed sensor network, Spatially and Temporally correlated information can be observed. So a logical grouping based on the node's location can be done, which simply 'Clustering'. Each cluster is having a representative node termed as 'Cluster Head' (CH). All the communication between other nodes and to the processing node is done through the CH. Individual nodes will send the observed information to the CH and CH will do some processing (data aggregation and compression) and forward the aggregated data to the processing node. Thus by clustering we can avoid longer distance communication by each individual node, and only CH will do direct communication with the processing node. By this total number of data communication rounds, energy consumption of each individual node can be reduced and overall network lifetime can be increased.

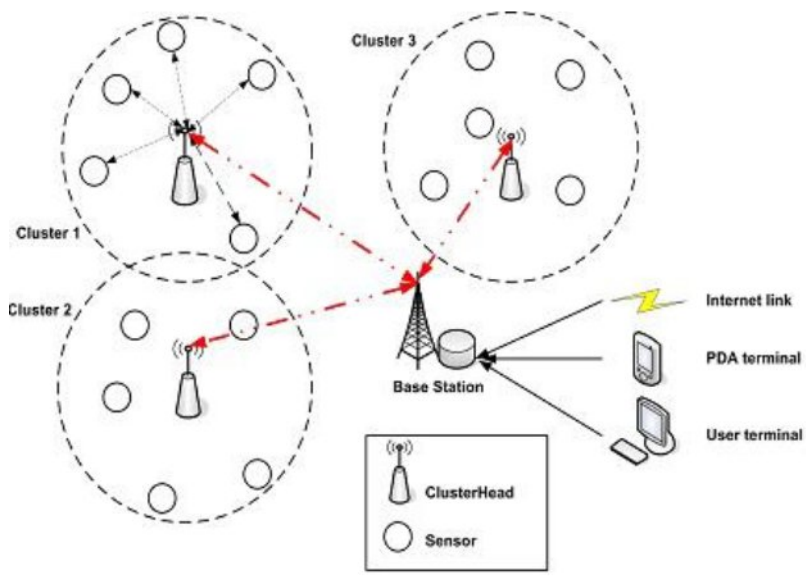

Figure 1: Sensor Networks

Advantages of Clustering:

- Transmit aggregated data to the processing node
- Very few nodes transmit data to the processing node
- Flexibility in network size
- Less message transmission as compared to direct method
- Optimized usage of battery and other resources

1.3 Motivation

Weather reporting, Warfield monitoring became main motivation for developing of Wireless Sensor Network. These days usage of WSNs have expanded to many commercial, manufacturing and daily life related applications, for example manufacturing line observation, security related, monitoring health of equipments and other industrial process monitoring and control.

The unique properties mentioned above become challenges to set up a sensor network. The key challenge in setting up and proper operation of WSN is to reduce battery dissipation of individual node and enhance the overall network life span. Since from last few years' variety of changes have been made to limit the energy requirement in WSN, as mainly energy dissipation is more for wireless transmission and reception. Main approaches till proposed were focusing at making the changes at MAC layer and network layer to minimize the energy dissipation. Two more major challenges are how to place the cluster heads over the grid and how many clusters would be there in a network. If the cluster heads are properly placed over the grid and sufficient clusters are formed, it will help to minimize the dissipation of energy and would help to increase the lifetime of the network. Clustering is always been referred as an effective method to enhance the lifetime of WSN.

1.4 Problem Statement

The purpose of this proposed technique is to present a method in which energy consumption of deployed sensor nodes in an area of observation can be minimized, so that lifetime of entire area can be minimized. To achieve this target we have exploited the benefits of Spatial and Temporal correlation. So that redundant information can be minimized. Using Spatial correlation clusters have been formed and each cluster will be assigned a cluster head. Cluster head will monitor entire area to which it is associated and will forward its data further. Also we have optimized the number of message transmission by using Supernode. Supernode will collect data from various cluster heads and send the aggregated data to the processing node.

Chapter 2 Related Work

2.1 Spatial and Temporal Correlation

Spatial Correlation

Typical WSN applications require spatially dense sensor deployment in order to achieve satisfactory coverage. As a result, multiple sensors record information about a single event in the sensor field. Due to high density in the network topology, spatially proximal sensors observations are highly correlated with the degree of correlation increasing with decreasing inter node separation.

Temporal Correlation

Wireless Sensor Networks applications may need periodic readings from the sensor nodes. Such periodic observations may have temporal correlation continuous observed information. Depending on the physical phenomenon is changing the amount of Temporal correlation between consecutive observation may change.

In addition to the collaborative nature of the WSN, the existence of above mentioned spatial and temporal correlations bring significant potential advantages for the development of efficient communication protocols well-suited for the WSN paradigm. For example, intuitively, due to the spatial correlation, data from spatially separated sensors is more useful to the processing node than highly correlated data from nodes in proximity. Therefore, it may not be necessary for every sensor node to transmit its data to the processing node; instead, a smaller number of sensor measurements might be adequate to communicate the event features to the processing node within a certain reliability/fidelity level. Similarly, for a certain event tracking application, the measurement reporting frequency, at which the sensor nodes transmit their observations, can be adjusted such that temporal-correlated phenomenon signal is captured at the processing node within a certain distortion level and with minimum energy-expenditure.

There has been some research effort to study the correlation in WSN [13, 14, 15, 16]. However, most of these existing studies investigate the information theoretical aspects of the correlation, and they do not provide efficient networking protocols which exploit the correlation in the WSN. On the other hand, there exist some proposals which attempt to exploit spatial correlation in WSN [14, 15]. However, these schemes aim to find the optimum rate to compress redundant information in the sensor observations and they also do not propose to exploit correlation for developing efficient communication protocols for the WSN. In a recent effort, the joint routing and source coding is introduced in [16] to reduce the amount of traffic generated in dense sensor networks with spatially correlated records. While joint routing and source coding reduces the number of transmitted bits; from the network point of view, the number of transmitted packets remains unchanged, which can be further minimized by regulating the network access based on the spatial correlation between the sensor nodes.

On the other hand, there already exists significant amount of research on the communication protocols for sensor networks in the literature [17]. For example, there exist some proposals to address the medium access control (MAC) problems in wireless sensor networks. However, these solutions mostly focus on energy-latency tradeoffs. S-MAC [18] aims to decrease the energy consumption by using sleep schedules with virtual clustering. A variant of S-MAC, T-MAC, incorporates variable sleep schedules to further increase the energy consumption [19]. However, in both protocols, since spatial correlation is not exploited, sensor nodes continue to send redundant data with increased latency due to sleep durations. In addition to contention-based protocols, TDMA-based protocols have also been proposed [20]. Although these protocols aim energy efficiency; both of the protocols assume a cluster-based topology, which requires significant additional processing complexity in the overall sensor network. In [21], energy efficient collision-free MAC protocol is presented. The protocol is based on a time-slotted structure and uses a distributed election scheme based on traffic requirements of each node to determine the time slot that a node should use. Although the protocol achieves high delivery ratio with tolerable delay, the performance of the protocol depends on the two-hop neighborhood information in each node, which, in the case of high density sensor networks, may result in either incomplete neighbor information due to collisions or high energy consumption due to signaling cost. In summary, none of these MAC protocols take advantage of the correlation in the WSN in order to improve energy-efficiency without compromising on the access latency.

2.2 Cluster Head Selection Methods

In recent years, the interest on clustered WSNs has generated a significant body of research works. A CH may be elected by the sensors in a cluster or pre-assigned by the network designer. A CH may also be just one of the sensors or a node that is richer in resources. Also the cluster membership of a node may be fixed or Variable. In this section, we will review CH selection algorithms.

2.2.1 LEACH

Most basic and admired protocol for clustering, known as Low Energy Adaptive Clustering Hierarchy (LEACH) which had been a basis for many further clustering protocols [3, 4]. The most important goal of LEACH is to have Cluster Heads to reduce the energy cost of transmitting data from normal nodes to a distant Processing Node. In LEACH, nodes organize themselves into local clusters with one node acting as cluster head. All non-cluster head nodes (normal nodes) transmit their data to the cluster heads. Cluster head nodes do some data aggregation and/or data fusion function on which should be transmitted to Processing Node. The cluster heads change randomly over a period of time to balance the nodes energy dissipation.

The operation of LEACH is divided into two phases: Set up Phase and steady state phase. Each round begins with a set-up (clustering) phase when clusters are organized, followed by a steady- state (transmission) phase in which data packets are transferred from normal nodes to cluster heads. After data aggregation, cluster heads will transmit the messages to the Processing Node.

Set up Phase: During this phase each node decides whether or not to become a cluster head for the current round. The election of cluster head is done with a probability function: each node selects a random number between 0 and 1 and if the number is less than T (n), the node is elected as a cluster head for current round:

$$T(n) = \frac{P}{1 - P \times (r \bmod P^{-1})} \qquad \forall n \in G$$

$$T(n) = 0 \qquad \forall n \in G$$

Where n is a random number between 0 and 1
P is the cluster-head probability and
G is the set of nodes that weren't cluster-heads the previous rounds

After this CH election, each cluster head prepares a TDMA schedule and transmits to all the cluster nodes in that respective cluster. This completes the set up phase of LEACH.

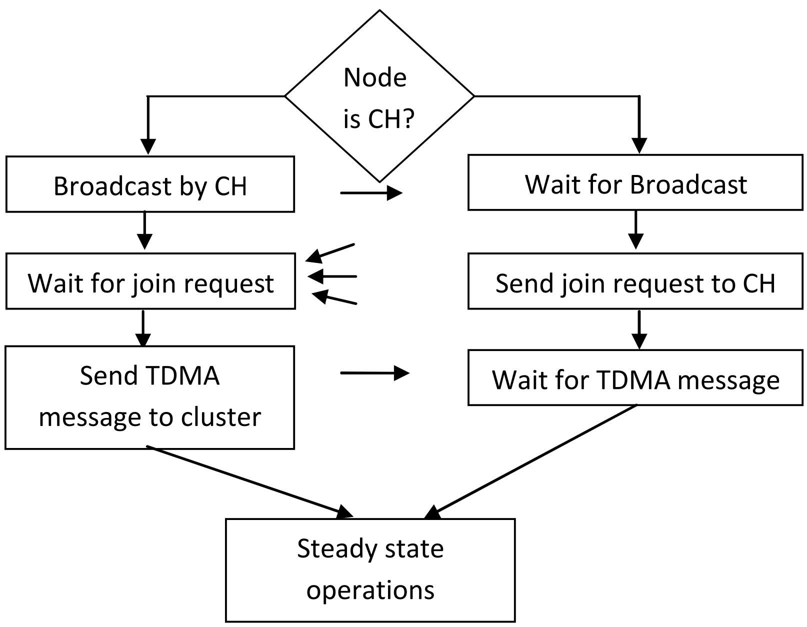

Figure 2: Flow Chart for Set UP Phase

Steady State Phase: In this phase nodes send their collected data to CH at once per frame allocated to them. This assumes that the node always has a data to transmit. The node goes to sleep mode after this transmission until next allocated transmission slot, to save the energy. The CH must keep its receiver on all the time to receive the data from cluster nodes. After reception of all the data, CH aggregates that data and transmits it to the Processing Node.

The strength of LEACH is in its CH rotation mechanism and data aggregation. But one important problem with LEACH is that it offers no guarantee about placement and/or number of cluster head nodes in every round. Therefore using a centralized clustering algorithm would produce better results. LEACH-Centralized (LEACH-C) is a Processing Node cluster formation algorithm. It uses the same steady state protocol as LEACH.

2.2.2 HEED

Cluster head selection is primarily based on the residual energy of each node [6, 7]. Since the energy consumed per bit for sensing, processing, and communication is typically known, and hence residual energy can be estimated. Intra cluster communication cost is considered as the secondary parameter to break the ties. A tie means that a node might fall within the range of more than one cluster head.

When there are multiple candidate cluster heads, the cluster head yielding lower intra-cluster communication cost are favored. The secondary clustering parameter, intra-cluster communication cost, is a function of:

(i) Cluster properties, such as cluster size, and
(ii) Whether or not variable power levels are permissible for intra-cluster communication.

If the power level used for intra cluster communication is fixed for all nodes, then the cost can be proportional to (i) NODE DEGREE , if the requirement is to distribute load among cluster heads, or (ii) 1/ NODE DEGREE, if the requirement is to create dense clusters.

This means that a node joins the cluster head with minimum degree to distribute cluster head load or joins the one with maximum degree to create dense clusters. Each node performs neighbor discovery, and broadcasts its cost to the detected neighbors. Each node sets its probability of becoming a cluster head, CH_P, as follows:

$$CH_P = \max (C_P * (E_{Res}/E_{max}), Pmin)$$

Where, C_P is the initial percentage of cluster heads among n nodes (it was set to 0.05), while E_{Res} and E_{max} are the residual and the maximum energy of a node (corresponding to the fully charged battery), respectively. The value of CH_P is not allowed to fall below the threshold P_{min}.

2.2.3 TEEN

TEEN [8] (Threshold sensitive Energy Efficient sensor Network protocol) is targeted at reactive networks and is the first protocol developed for reactive networks.

Functioning

At every cluster change time, in addition to the attributes, the cluster-head broadcasts to its members.

Hard Threshold (HT): This is a threshold value for the sensed attribute. It is the absolute value of the attribute beyond which, the node sensing this value must switch on its transmitter and report to its cluster head.

Soft Threshold (ST): This is a small change in the value of the sensed attribute which triggers the node to switch on its transmitter and transmit. The nodes sense their environment continuously. The first time a parameter from the attribute set reaches its hard threshold value, the node switches on its transmitter and sends the sensed data. The sensed value is stored in an internal variable in the node, called the sensed value (SV). The nodes will next transmit data in the current cluster period, only when both the following conditions are true:

1. The current value of the sensed attribute is greater than the hard threshold.
2. The current value of the sensed attribute differs from SV by an amount equal to or greater than the soft threshold.

Whenever a node transmits data, SV is set equal to the current value of the sensed attribute. Thus, the hard threshold tries to reduce the number of transmissions by allowing the nodes to transmit only when the sensed attribute is in the range of interest. The soft threshold further reduces the number of transmissions by eliminating all the transmissions which might have otherwise occurred when there is little or no change in the sensed attribute once the hard threshold.

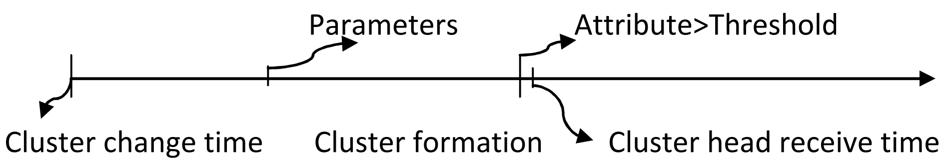

Figure 3: Time line for proactive protocol

Important Features

The main features of this scheme are as follows:

1. Time critical data reaches the user almost instantaneously. So, this scheme is eminently suited for time critical data sensing applications.
2. Message transmission consumes much more energy than data sensing. So, even though the nodes sense continuously, the energy consumption in this scheme can potentially be much less than in the proactive network, because data transmission is done less frequently.
3. The soft threshold can be varied, depending on the criticality of the sensed attribute and the target application.

4. A smaller value of the soft threshold gives a more accurate picture of the network, at the expense of increased energy consumption. Thus, the user can control the trade-off between energy efficiency and accuracy.

5. At every cluster change time, the attributes are broadcast afresh and so, the user can change them as required.

The main drawback of this scheme is that, if the thresholds are not reached, the nodes will never communicate; the user will not get any data from the network at all and will not come to know even if all the nodes die. Thus, this scheme is not well suited for applications where the user needs to get data on a regular basis. Another possible problem with this scheme is that a practical implementation would have to ensure that there are no collisions in the cluster. TDMA scheduling of the nodes can be used to avoid this problem. This will however introduce a delay in the reporting of the time-critical data. CDMA is another possible solution to this problem.

Example Applications

This protocol is best suited for time critical applications such as intrusion detection, explosion detection etc

2.3 PEGASIS

The main idea in PEGASIS [5] is for each node to receive from and transmit to close neighbors and take turns being the leader for transmission to the Processing Node. This approach distributes the energy load evenly among the sensor nodes in the network. Nodes are initially placed randomly in the observation area. The nodes will be organized to form a chain, which can either be accomplished by the sensor nodes themselves using a greedy algorithm starting from some node. Alternatively, the Processing Node can compute this chain and broadcast it to all the sensor nodes.

For constructing the chain, all nodes are considered to have global knowledge of the network and employ the greedy algorithm. The greedy approach to constructing the chain works well and this is done before the first round of communication. To construct the chain, start with the furthest node from the Processing Node. Begin with this node in order to make sure that nodes farther from the Processing Node have close neighbors, as in the greedy algorithm the neighbor distances will increase gradually since nodes already on the chain cannot be revisited. Figure shows node A connecting to node B, node B connecting to node C, and node C connecting to node D in that order. When a node dies, the chain is reconstructed in the same manner to bypass the dead node.

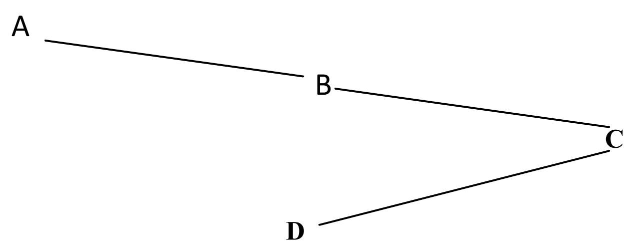

Figure 4: Chain construction

For gathering data in each round, each node receives data from one neighbor, fuses with its own data, and transmits to the other neighbor on the chain. Nodes take turns transmitting to the Processing Node. Thus, the leader in each round of communication will be at a random position on the chain, which is important for nodes to die at random locations. The idea in nodes dying at random places is to make the sensor network robust to failures. In a given round, use a simple control token passing approach initiated by the leader to start the data transmission from the ends of the chain. The cost is very small since the token size is very small. In Figure 5, node C2 is the leader, and it will pass the token along the chain to node C0. Node C0 will pass its data towards node C2. After C2 receives data from node C1, it will pass the token to node C4, and node C4 will pass its data towards node C2.

PEGASIS performs data fusion at every node except the end nodes in the chain.

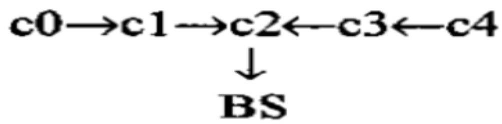

Figure 5: Token passing approach

Each node will fuse its neighbor’s data with its own to generate a single packet of the same length and then transmit that to its other neighbor (if it has two neighbors). Thus, in PEGASIS each node will receive and transmit one packet in each round and be the leader once every 100 rounds. The greedy chain construction performs well with different size networks and random node placements. In constructing the chain, it is possible that some nodes may have relatively distant neighbors along the chain. Such nodes will dissipate more energy in each round compared to other sensors. Performance of PEGASIS is improved by not allowing such nodes to become leaders. This is accomplished by setting a threshold on neighbor distance to be leaders. Performance can be further improved slightly by applying a threshold adaptive to the remaining energy levels in nodes. Whenever a node dies, the chain will be reconstructed and the threshold can be changed to determine which nodes can be leaders.

Chapter 3 Proposed Technique

Clustering based on Spatial and temporal correlation is proved to be an efficient way of saving sensor node energy, and by this increasing overall network lifetime. To use this effectiveness we have proposed a new technique, which is a hybrid technique of existing techniques which is based on Spatial and Temporal Correlation, HEED, PEGASIS. Main motivation of proposing this technique is to use Spatial and Temporal correlation in which those nodes which have highest residual energy will be selected as Cluster representative i.e. Cluster head and data communication will use Supernode.

3.1 Correlation Models

3.1.1 Spatial Correlation Model

Spatial correlation model used in our proposed hybrid technique is based on [1]

Sensor Deploying Model

The Spatial correlation model in our technique is taken is as follows: Area of observation is a circular area in which random and dense distribution of sensor nodes are done.

The followings are a few assumptions:

(i) Sensor nodes will have a sensing radius, all the activities inside this radius will be captured and no information will be captured outside this radius.

(ii) Sensor nodes can communicate to a distance which is much greater than their sensing range.

(iii) Only aggregated data will be sent to the processing node.

(iv) Once deployed node's position will be fixed and then every node's location and other related data will be shared to every other node.

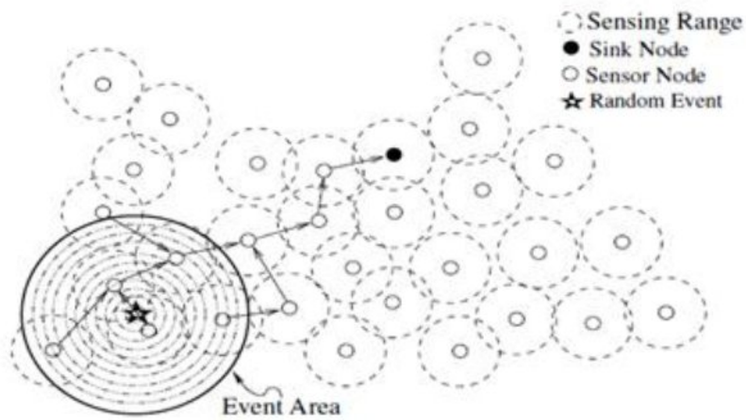

Figure 6: Model architecture

The Correlation Model

The correlation characteristics on which the degree of spatial correlation between sensor nodes in the area of monitoring are:

1. Interdistance between nodes
2. Node's sensing range

$$K_\theta(d) = \frac{\cos^{-1}\left(\frac{d}{\theta}\right)}{\pi} - \frac{d}{\pi\theta^2}\cdot\sqrt{(\theta^2 - d^2)} \quad \text{For } 0 \le d \le \theta;$$
$$= 0 \quad \text{For } d > \theta.$$

θ = 2r, and d = distance between nodes

Above equation clearly shows that the correlation between sensor nodes is a direct function of Euclidian distance between nodes. To simplify our calculation in our proposed model we have taken the Euclidian distance between nodes as main clustering criteria.

Cluster formation:

---

1. Initialize every node and insert into an empty set E
2. **For** every node Define a minimum distance d_{min} between nodes for being cluster member
3. Each node n in the area **do while E is not empty :**
4. Calculate distance of every node from every other node $d_{inter-node}$
5. If $d_{inter-node} <= d_{min}$:

   Insert node n into set S = S {n}
6. End if
7. E= E- {n}
8. End while
9. Construct new cell C = E
10. N=n - {C}
11. End **For**
12. Return C = {C1, C2, C3…}
13. Cluster formation completed

3.1.2 Temporal Correlation Model

Wireless Sensor Networks applications may need periodic readings from the sensor nodes. Such periodic observations may have temporal correlation continuous observed information. Depending on the physical phenomenon is changing the amount of Temporal correlation [2] between consecutive observation may change.

We consider that every node has multiple data to send (say n). Each data is divided into L different chunks. Each sensor is has allocated separate memory for Base signal, which is collection of data which is extracted from the observed values. This Base signal is used to transmit along with the partial message signal. Base signal is then used at the Processing Node to reconstruct the original signal.

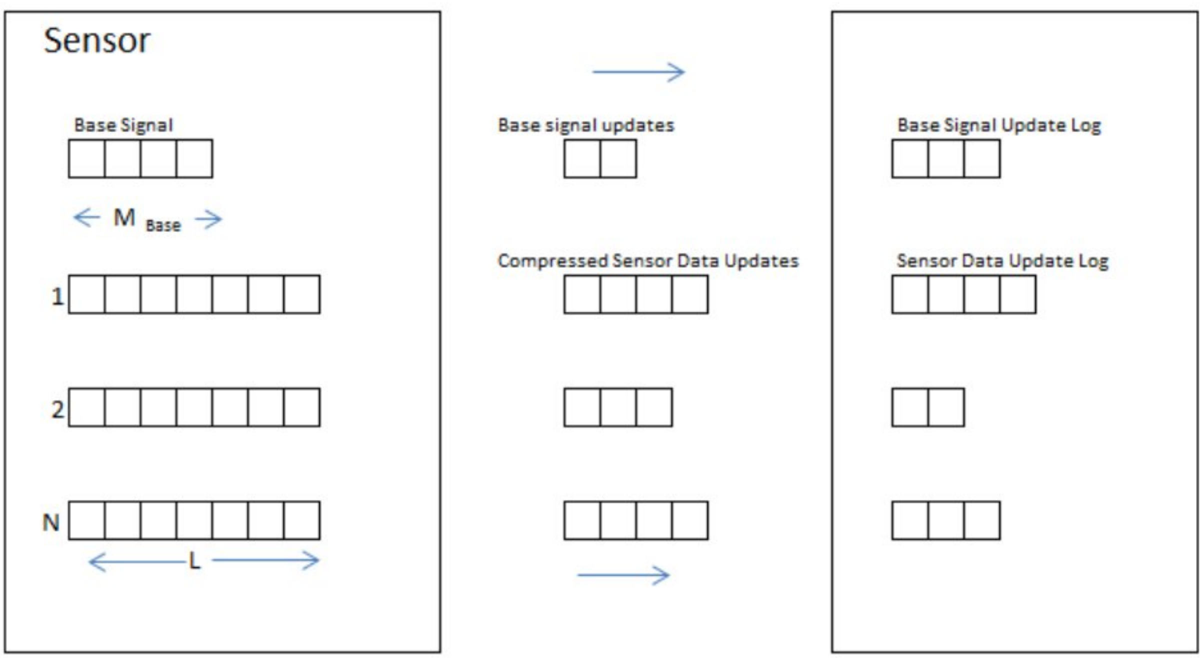

Figure 7: Data compression-Temporal Correlation

When enough data is collected (till there is no remaining space in sensor memory), recent data values are approximated by signal value B (smaller than L). Resulting approximated value shall be transmitted to the Processing Node, which is nothing but the compressed data. A history/log is maintained at the Processing Node which is generated by adding latest information received from the various nodes. A dedicated space is allocated for each sensor which transmits information to the receiving end i.e. Processing Node. It is depicted in above figure.

Base signal is collection of differential information which is gathered from the observed values by the sensor nodes. Base signal is then used at the Processing Node to recover the approximated/compressed signals sent by specific sensor node.

Sensor nodes transmits signal to the Processing Node which are approximated using the memorized data by the Base signal. Information in the Base signal can be refreshed every time whenever there is new information is seen by the sensor nodes. This ensures the best quality of approximation. On receiving such updated in base signal transmission of base signal also occurs with the actual data and it is updated in Base signal field of dedicated space for each sensor in the Processing Node. The base signal created using differential samples of the original observations by the sensor nodes at specific intervals.

The process is initiated by extracting smallest data unit of each observation. Sorting is done by the Processing Node using their starting position and there is no requirement of end point.

3.2 Energy Dissipation in transmission

The communication model used in this technique is similar to that used in [3,5] which is most basic communication model. A node dissipates $P_{trans/rec}$ = 50 nJ/bit to send and receive the information from and to other nodes and P_{amp} = 100 pJ/bit/m^2 energy is dissipated to amplify the signal so that information to the other nodes can be received within a set signal to noise ratio (S/N).

Data transmission dissipates energy in two ways, Energy required in transmission of data to distance l : $P_{Tx\text{-}elec}$ and amplifying signal to maintain required S/N ratio $P_{Tx\text{-}amp}$.
Data reception dissipates energy in receiving P_{Rx}

Energy dissipation required in data transmission with n bits and distance l:

$$P_T(n,l) = P_{trans}(n) + P_{amp}(n,l)$$

$$P_T(n,l) = P_{trans}*n + P_{amp}*n*l^2$$

Our model shows that the energy consumption is the transmission is the combination of transmission energy plus the amplifying energy. Receiving requires the energy to receive the transmitted signal. So the cost of transmission is almost double of the receiving cost. Also receiving is also consuming a considerable amount of energy. So our approach should be to minimize both transmission and reception of the signal.

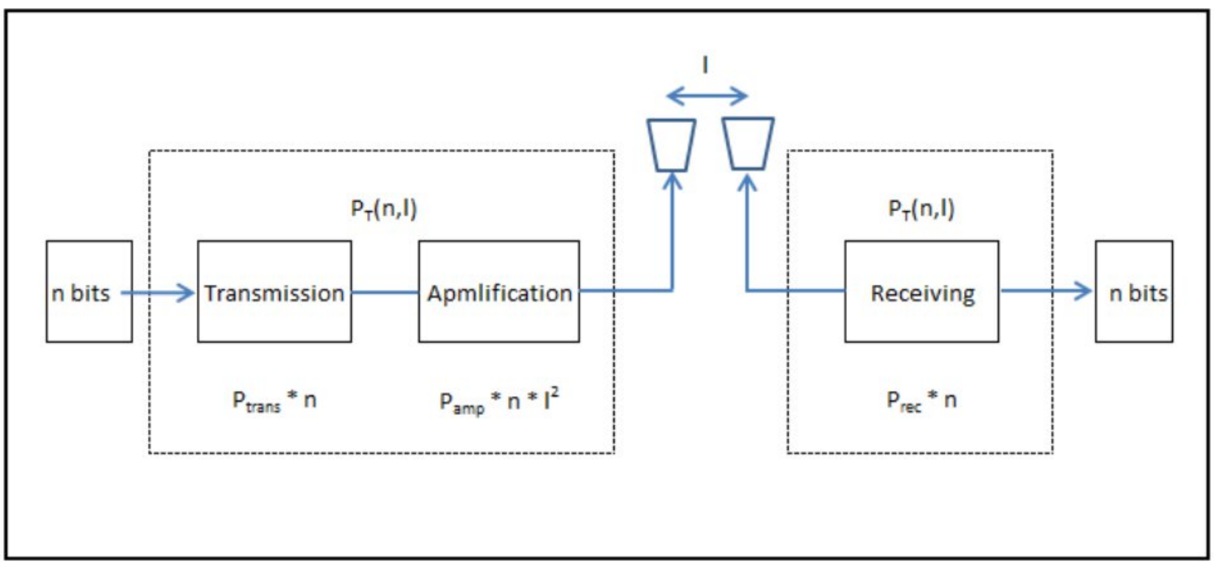

Figure 8: Radio Transmission Model

Reception dissipates energy as below:

$$P_R(n; l) = P_{rec}(n)$$

$$P_R(n; l) = P_{rec} * n$$

3.3 Assumptions: Proposed Technique

1. Nodes are distributed randomly in remote location
2. Nodes are deployed to observe weather data which includes temperature, air pressure, humidity, wind speed.
3. There will be a Spatial correlation between the observations given by sensor nodes
4. There will be a Temporal correlation between the observations given by sensor nodes.
5. For a given signal to noise ratio (SNR), energy consumption in transmission and reception between two nodes will be same.
6. All sensor nodes will be considered at same level of energy before starting our simulation.

Idle time dissipation is ignored for all type of networks

3.4 Flow Chart

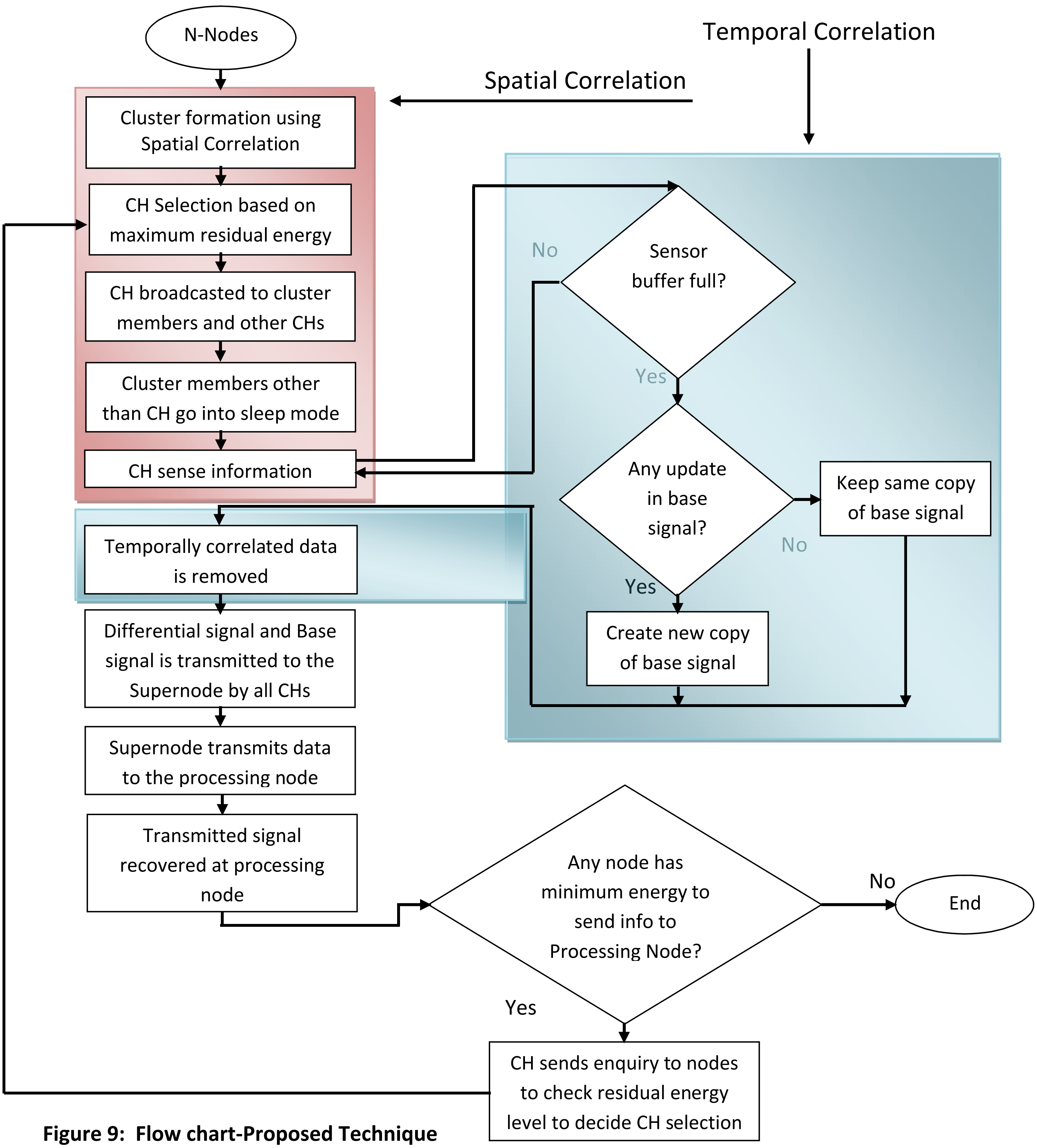

Figure 9: Flow chart-Proposed Technique

3.5 Performance Evaluation

For performance evaluation we will implement our code in 'C' language and output will compared between various techniques and our proposed technique.

Comparison of various techniques with our proposed technique will be based on:

1. Number of live nodes after each round of communication
2. Total number of communication rounds-Network total lifetime

In simulation a network of 100 randomly deployed nodes and a Processing Node is at fixed position at far distance from sensor networks. All the Nodes have been initialized with an initial energy 0.5 J.

3.6 Different techniques for Comparison

5 runs for each technique including our proposed method have been done, readings have been taken and graph plotted. The technique in which more number of nodes will be live and maximum number of rounds of data communication will occur shall be considered as most efficient technique.

1. Direct technique: Every node (100 in our case) will send its data directly to Processing Node
2. Clustering_Each Node: Cluster Head will receive data from each node which are attached to it. CH will aggregate information at cluster head will send to the Processing Node.
3. Clustering_Only CH: Only Cluster head will be active and other nodes will be in sleep mode. Thus entire observation of the area will be sent by the CH only.
4. In our **proposed technique** Clusters will be formed on basis of Spatial Co-relation and only Cluster head will send its data to Super node of Cluster heads. Finally only Super node will send aggregated data to the Processing Node.

Chapter 4 Simulation and Results

4.1 Simulation Environment

For our experiments, we simulated an environment with randomly distributed nodes in 100x100 unit area. 100 random nodes are deployed in this area. Processing Node is placed at far distance from the monitoring area:

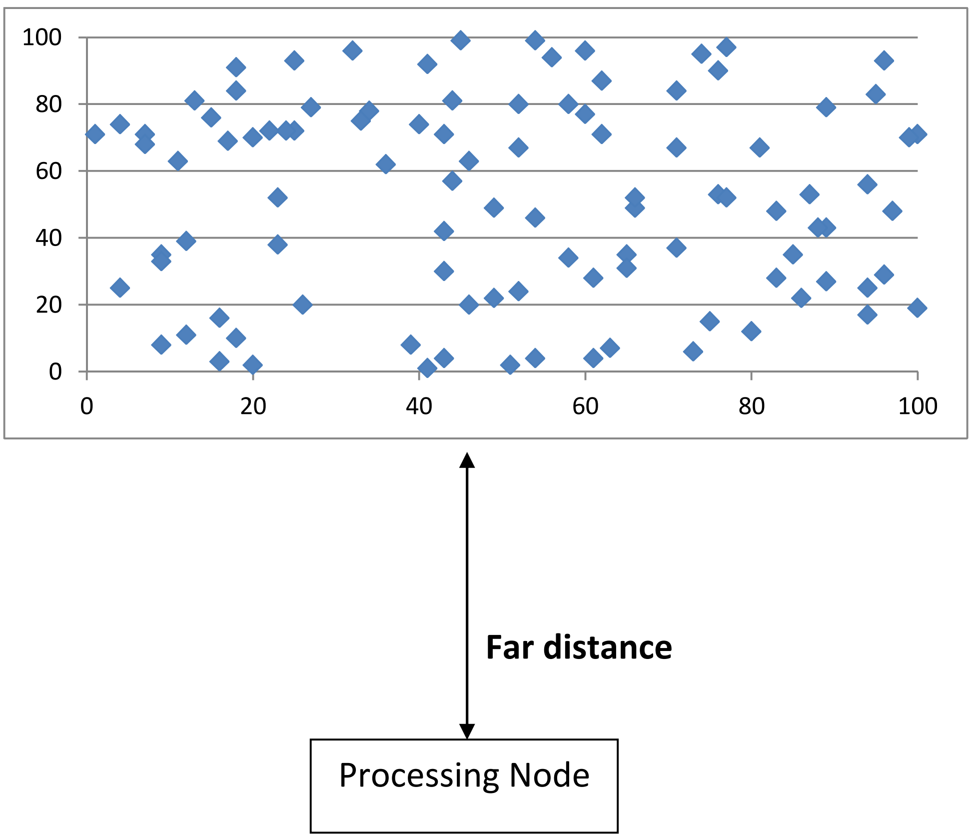

Figure 10: Sensor node deployment

4.2 Readings

Node Initial energy = 10; Number of bits transferred = 2000

Iteration 1:

1. Node distribution :

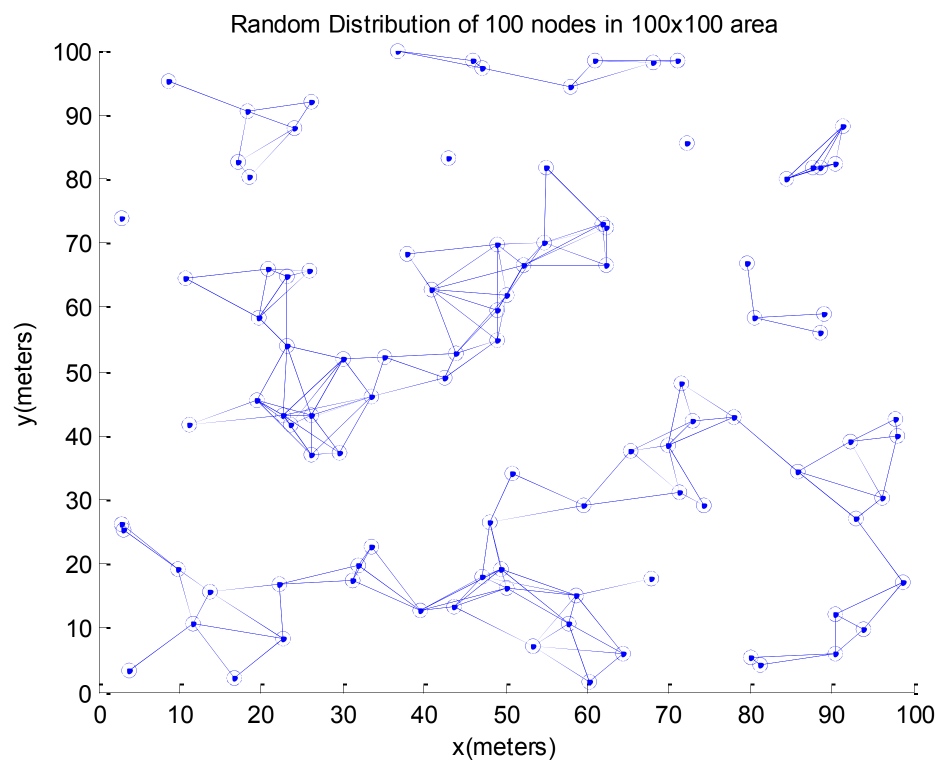

Figure 11: Cluster formation using correlation (1)

2. Comparison of communication rounds at different instances of died sensor nodes :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	59	78	127	720
	Clustering_All Nodes	19	66	185	815
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	97	583	1967	8674

Iteration 2:

1. Node distribution :

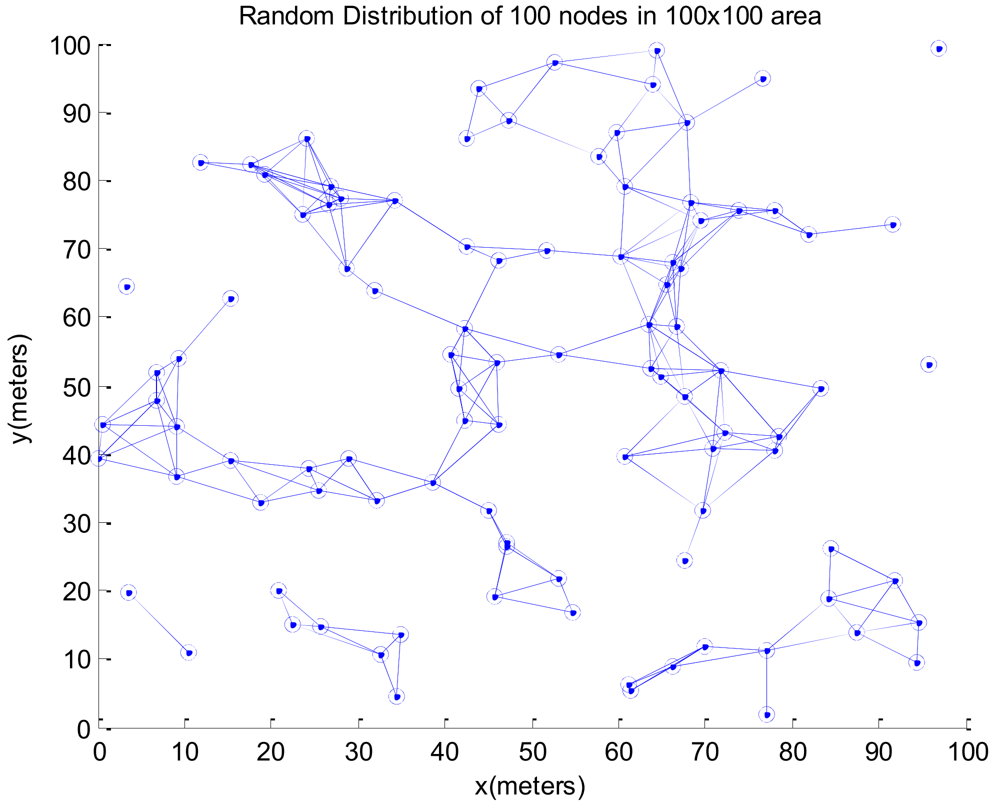

Figure 12: Cluster formation using correlation (2)

2. Comparison of communication rounds at different instances of died sensor nodes :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	51	73	114	674
	Clustering_All Nodes	16	53	156	1000
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	89	413	1604	7106

Iteration 3:

1. Node distribution :

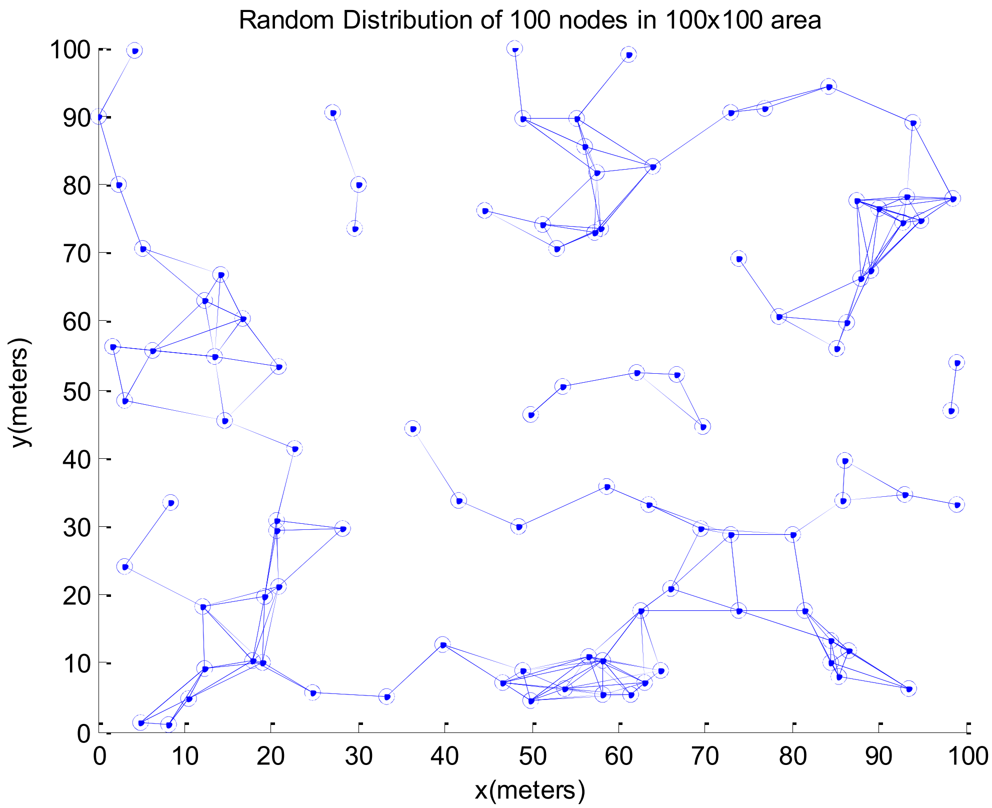

Figure 13: Cluster formation using correlation (3)

2. Comparison of communication rounds at different instances of died sensor nodes :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	52	70	107	755
	Clustering_All Nodes	16	57	162	967
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	112	522	1581	6692

Iteration 4:

1. Node distribution :

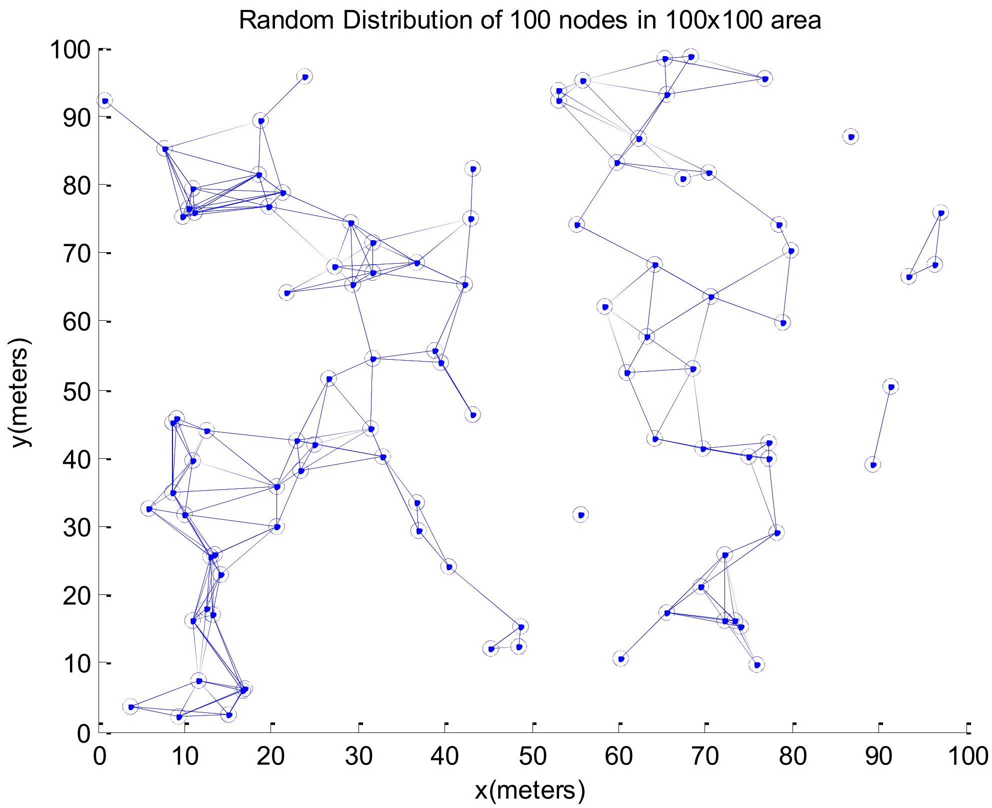

Figure 14: Cluster formation using correlation (4)

2. Comparison of communication rounds at different instances of died sensor nodes :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	56	77	114	733
	Clustering_All Nodes	21	67	210	1042
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	100	564	1860	8864

Iteration 5:

1. Node distribution :

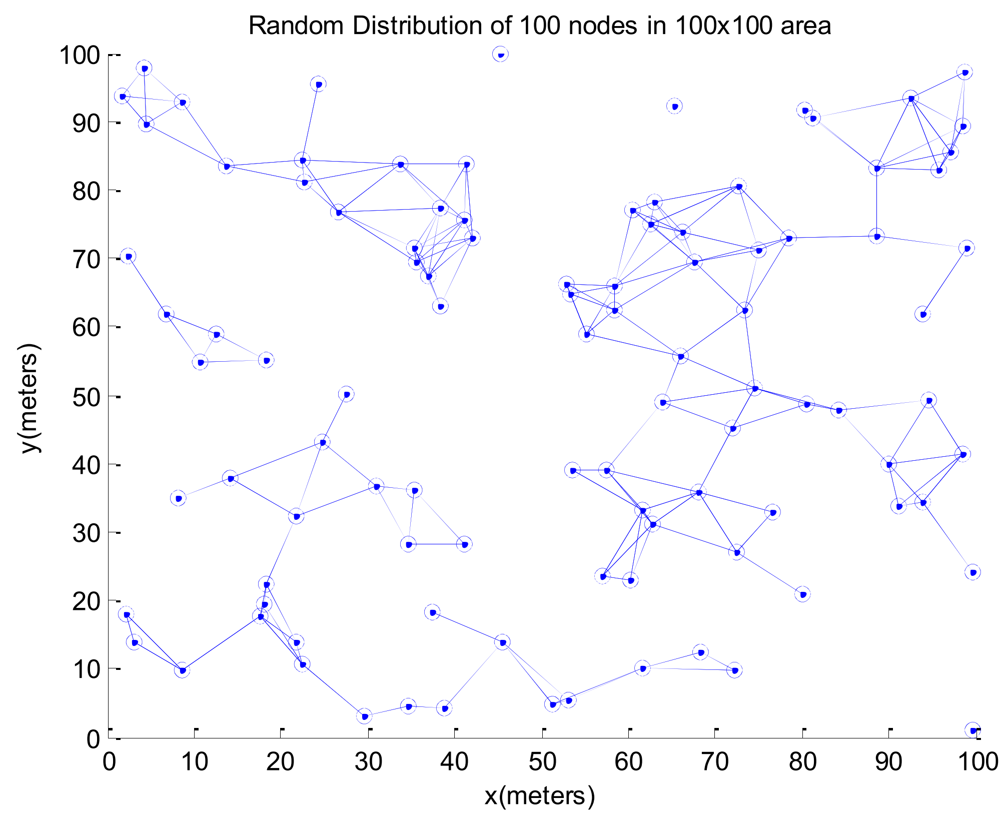

Figure 15: Cluster formation using correlation (5)

2. Comparison of communication rounds at different instances of died sensor nodes :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	53	74	100	820
	Clustering_All Nodes	13	52	154	1033
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	99	635	1823	6024

1. Average Comparison of communication rounds at different instances of died sensor nodes of 5 Iterations :

Initial Energy	Protocol	5%	20%	50%	100%
10	Direct	54	74	112	740
	Clustering_All Nodes	17	59	173	971
	Clustering_Only CH	200	800	2000	4000
	Clustering_Supernode	99	543	1767	7472

2. Node death vs. Number of rounds :

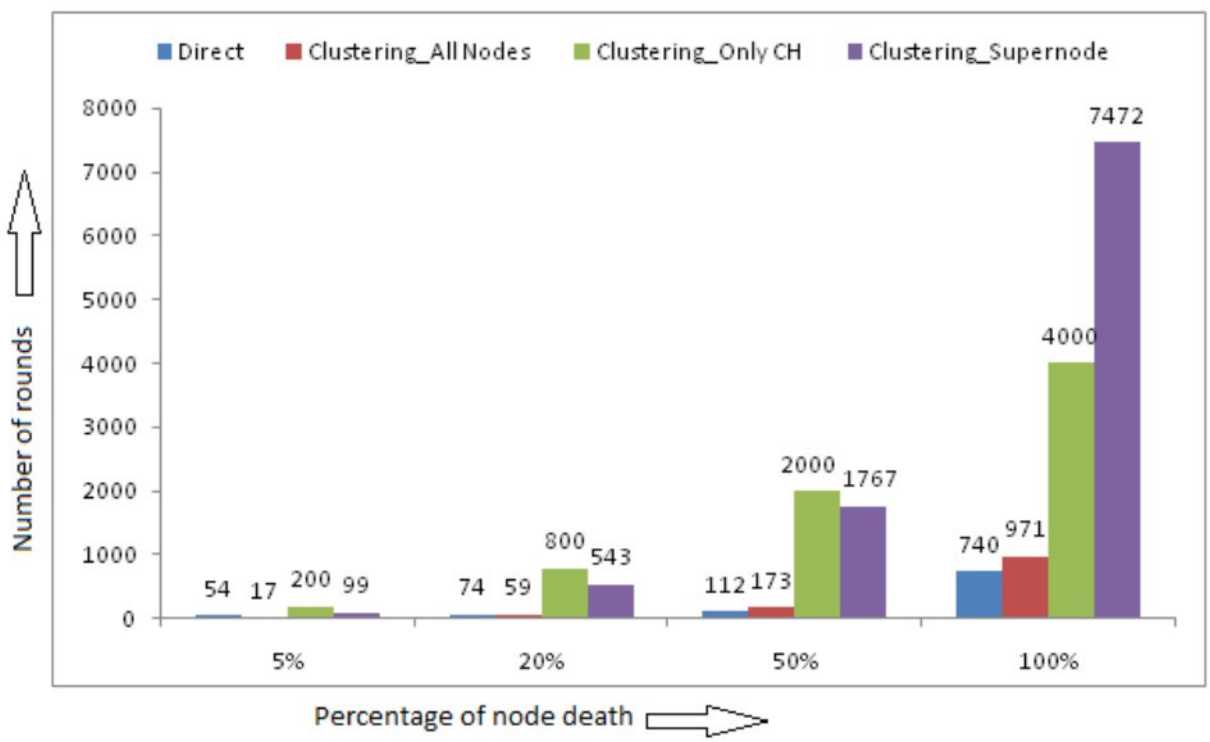

Figure 16: Node death vs. Number of rounds

3. Network Life span :

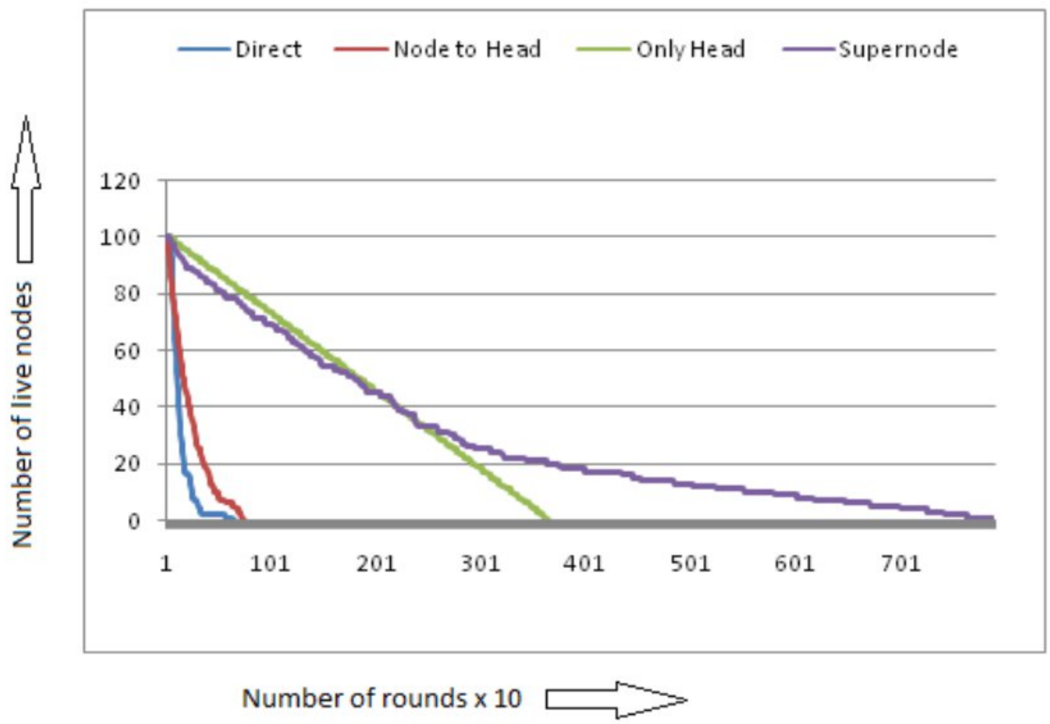

Figure 17: System lifetime comparison for various techniques

Chapter 5 Conclusion and Future Work

We have seen the importance of Correlation (both Spatial and Temporal) in clustering in Wireless Sensor networks. It reduces communication overhead significantly by reducing correlated data: Spatially as well as temporal. Thus reduces energy consumption in data transmission in WSN. Reduction in energy consumption results in enhanced overall lifetime of network. Clustering play a vital role in reducing energy consumption in transmission and increasing the network lifetime.

By using Correlation model we can see the correlated information in certain area of observation. Nearby Sensor nodes based on their location observes similar information and thus the observed information is correlated. As observed information is correlated, only few nodes can participate in data transmission without loss of information. Spatial correlation helps to avoid the redundant information transmission to the Processing Node. Spatial correlation allows compressing similar information observed at various time spaces. Base signal which stores the differential information observed by the sensor nodes. Base signal helps to reconstruct the partial information at the Processing Node which is transmitted to the Processing Node.

If deployed nodes are having data to be sent at different time intervals then various approximation, synchronization and communication techniques need modifications. Main reason behind this is that in most of techniques data compression and aggregation considering same transmission frequency from all the nodes.

Proposed technique in the this research has not considered the energy loss in creating cluster, and internal communication within cluster such as CH signals other nodes to go into sleep mode, which node is currently CH, who is having highest probability of becoming next CH etc. Our proposed technique can be further explored and compared with similar kind of techniques.

Chapter 6 References

1. Rajeev K. Shakya, Yatindra Nath Singh, and Nishchal K. Verma "A Novel Spatial Correlation Model for Wireless Sensor Network Applications" IEEE 2012.
2. Antonios Deligiannakis and Yannis Kotidis "Exploiting Spatio-temporal Correlations for Data Processing in Sensor Networks" Springer-Verlag Berlin Heidelberg 2008.
3. W. Heinzelman, A. Chandrakasan, and H.Balakrishnan. "Energy-Efficient Communication Protocol for Wireless Microsensor Networks ". In Proceedings of the Hawaii Conference on System Sciences, Jan. 2000
4. Lalita Yadav and , Ch. Sunitha "Low Energy Adaptive Clustering Hierarchy in Wireless Sensor Network (LEACH)" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4661-4664 www.ijcsit.com 4661
5. Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan "PEGASIS: Power-Efficient GAthering in Sensor Information Systems" Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000
6. Ossama Younis and Sonia Fahmy "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks"
7. Harneet Kour and Ajay K. Sharma "Hybrid Energy Efficient Distributed Protocol for Heterogeneous Wireless Sensor Network" International Journal of Computer Applications 2010
8. Arati Manjeshwar and Dharma P. Agrawal "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks" IEEE 2001
9. Vuran M. C., Akan O. B., Akyildiz I. F., "Spatio-Temporal Correlation: Theory and Applications Wireless Sensor Networks," Computer Networks Journal (Elsevier Science), vol. 45, no. 3, pp. 245 -259, June 2004.
10. Akyildiz I. F., Vuran M. C., Akan O. B., "On Exploiting Spatial and Temporal Correlation in Wireless Sensor Networks," in Proc. WiOpt'04: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, pp. 71 -80, March 2004.

11. S.B. Pourpeighambar, M. Sabaei “Spatial correlation aware protocols for efficient data aggregation of moving object in wireless sensor networks” 2013 Sharif University of Technology
12. Sundeep Pattem, Bhaskar Krishnamachari and Ramesh Govindan “The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks” IPSN 2004
13. Michael Gastpar and Martin Vetterli "Source-Channel Communication in Sensor Networks" Springer-Verlag Berlin Heidelberg 2003
14. Pradhan, S.S "Distributed source coding: symmetric rates and applications to sensor networks" Data Compression Conference, 2000. Proceedings. DCC 2000
15. Pradhan, S.S, Kusuma, J. ; Ramchandran, K. "Distributed compression in a dense microsensor network" Signal Processing Magazine, IEEE (Volume:19 , Issue: 2) 2002
16. A. Scaglione, S.D. Servetto, “On the interdependence of routing and data compression in multi-hop sensor networks”, in: Proceedings of the ACM MOBICOM 2002, Atlanta, GA, USA, September 2002.
17. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Computer Networks 38 (4) (2002) 393–422.
18. W. Ye, J. Heidemann, D. Estrin, An energy-efficient MAC protocol for wireless sensor networks, in: Proceedings of the INFOCOM 2002 vol. 3, 2002, pp. 1567–1576.
19. T. van Dam, K. Langendoen, An adaptive energy-efficient MAC protocol for wireless sensor networks, in: Proceedings of the ACM SenSys 2003, Los Angeles, CA, November 2003.
20. K.A. Arisha, M.A. Youssef, M.Y. Younis, Energy-aware management in cluster-based sensor networks, Computer. Networks 43 (5) (2003) 649–668.
21. V. Rajendran, K. Obraczka, J.J. Garcia-Luna-Aceves, Energy-efficient, collision-free medium access control for wireless sensor networks, in: Proceedings of the ACM SenSys 2003, Los Angeles, CA, November 2003.