An Energy Efficient Distributed Clustering Algorithm for Ad Hoc Deployed Wireless Sensor Networks in Building Monitoring Applications

Sankalpa Gamwarige[*]
Chulantha Kulasekere[**]
Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka.
Email: sankalpa@ent.mrt.ac.lk[*], ekulasek@ent.mrt.ac.lk[**]

ABSTRACT: In recent years Wireless Sensor Networks (WSNs) have been deployed for Building Monitoring (BM) as they provide a low cost and reconfigurable alternative to centralized cable based sensor systems. Using WSNs gives rise to unique issues in its practical usage. Lifetime of a WSN is one such crucial issue to be addressed during deployment. Clustering is an effective way of extending the lifetime of a WSN. In this article we propose a distributed and energy driven clustering algorithm where the selection of the cluster heads (CHs) are based on relative residual energy level of sensors. Furthermore, the CHs are rotated only when their energy drops below a dynamic threshold computed by the algorithm. As a result, the overheads in the inter sensor communications will be reduced and thereby the proposed algorithm will favor more powerful nodes over the weaker ones to prolong the lifetime of the entire WSN. This will effectively prolong the usability of the monitoring system and thus the underlying safety of the building.

The results will show that the proposed algorithm performs better when compared to existing clustering algorithms. Further we present theoretical analysis of the performance of the proposed algorithm in terms of correctness and complexity and explain how to identify the optimal values for key parameters such as transmission range R and re-clustering trigger threshold function value c in order to maximize the network lifetime.

KEYWORDS: Building Monitoring, Wireless Sensor Network, Self-organizing, Clustering

1 INTRODUCTION

Potential use of ad hoc deployed large WSNs for BM has been a popular research topic in recent times (Wang et al. 2007; Wua et al. 2009; Xu et al. 2004). In general, wireless sensors are used to monitor building environment conditions such as temperature, light intensity, CO/CO_2 level and building structure health. Traditionally, building monitoring systems used expensive high reliable macro sensors wired to a central data sink. The cost of such sensors is exorbitant to be deployed in many of the commercial buildings. The conditions have worsened as the size of the structures has exponentially increased. Over the past decade many low cost sensors have been developed (Heidemann & Govindan 2004; Culler et al. 2004). The reliability of an individual micro sensor node is less when compared to its expensive macro sensor counterpart. However, when a large number of such low cost micro sensor nodes are deployed and used within a collaborative data processing environment, they can easily match the performance of a macro sensor. However installation of these low cost sensor nodes using wired technology is very expensive and impractical when it comes to large BM systems. However connecting of low power, low bit rate wireless transceivers coupled with simple microcontrollers enhances their applicability and reduces cost. That is, for large buildings, ad hoc WSNs are fast becoming the platform of choice for monitoring applications. Such application has two characteristics that we will concentrate on. Namely the random distributed placement of sensors and the low power availability per node. For the purpose of communication which reduces the power consumption of a sensor, one has to look at the self organization of the sensors which are randomly distributed deployed. Additionally it is known that the Radio communication consumes most of the energy of a sensor node. Therefore, energy efficient communication and data gathering mechanisms are key issues that have to be considered for the successful deployment of such networks in practice (Estrin et al. 1999).

Most of the WSN based BM applications requires periodic data collection from the distributed sensors to one central location. Such a many-to-one data communication pattern is referred as convergecast (Cheng et al. 2008). The energy expenditure of such periodic convergecast network can be reduced by (1) Compressing the traffic volume using in-network collaborative data processing (2) Multi-hop communication to reduce required communication power (3) Decreasing wasteful energy consumption as a result of idle listening on wireless channel, overhearing, retransmissions due to packet collisions and protocol overhead for exchanging control packets.

Further clustering has been identified as an effective energy saving WSN organization framework under which above techniques can be adopted compared to other methods such as direct transmission to Base Station (BS), or minimum distance transmission i.e. relay the data through nearest neighbor (Ibriq & Mahgoub 2004; Heinzelman et al. 2002; Younis et al. 2006). In direct transmission, the nodes that are far away from the BS die rapidly as a result of long distance transmission. On the contrary in minimum distance transmission, the nodes that are closer to the BS die rapidly as they have to relay all the data packets coming from nodes located beyond its position.

Cluster based network organization framework partitions the network into disjoint clusters where each cluster consists of one Cluster Head (CH) and multiple member nodes. Any WSN clustering algorithm faces two challenges (1) How should clusters be formed? (2) How many clusters are required or cluster geometric dimensions? (Wang et al. 2004a)

The first question includes two aspects: how to select the CHs and how to associate a non CH node to a particular CH. Based on how this question is answered, existing clustering schemes can be classified into different categories. For example, clustering scheme can operate as centralized (e.g BCDCP (Muruganathan et al. 2005), EGSOM (Guru et al. 2005)) or distributed (e.g. LEACH (Heinzelman et al. 2002)); static or dynamic (e.g. ANTCLUST based (Kamimura et al. 2004)); a scheme can be applicable only for homogeneous energy networks i.e. all nodes in the network have same level of energy when the first time clustering algorithm applied (e.g. LEACH) or even for a heterogeneous energy network (Nodes have different amounts of energy at the beginning of the first time the algorithm is used. For example, application of a new algorithm to an existing network or addition of new nodes to an existing network) (e.g. HEED (Younis & Fahmy 2004)); the CH selection is weight independent i.e. randomized (e.g LEACH) or weight associated (e.g. HEED); procedure for CH selection can be finalized in one step (e.g LEACH) or iteratively (e.g. HEED, MEDIC (Zhao & Liang 2007)). Each of the above categories have their own advantages and disadvantages. In general, any scheme with more complex control can lead to near optimal energy efficient solutions (i.e. All nodes die at same time / even energy usage among all nodes). However, this also increases the overhead of control and coordination mechanism in terms of energy consumption.

The class of dynamic, distributive and randomized (DDR) clustering algorithms is promising in providing energy-efficient, load balancing, scalable and robust communications in WSNs due to their low complexity, good feasibility, and high effectiveness. This is the main reason that LEACH (and its derivatives such as SEP (Smaragdakis et al. 2004)) has attracted immense attention and has become a well studied and popularly referred baseline since its appearance (Wang et al. 2004a). However, LEACH has issues such as (1) Performance in heterogeneous energy networks (2) Non uniform cluster formation (3) LEACH produced the required number of CHs only 20% of the time. (4) There may be situations that entire network would be served with just one CH (5) A node with insufficient residual energy can occasionally become a CH even though there are neighboring nodes with more battery power and (6) Time based CH rotation (Zhao et al. 2007; Kim & Youn et al. 2005). Many other clustering schemes were presented to overcome some of the identified problems of DDR clustering algorithms like LEACH. However, they had issues with the lifetime measurements due to energy overhead as a result of the complex cluster setup algorithms (e.g. Younis & Fahmy 2004 have admitted that 'LEACH protocol expends less energy in clustering and produces longer lifetime than HEED') and some unacceptable assumptions such as location awareness of nodes using GPS or some form of localization technique (ANTCLUST based). Even most of the alternatives to LEACH algorithms such as HEED, SEP and ANTCLUST based have time driven CH role rotation mechanisms. That is, the role of CH will be changed after a predetermined number of data gathering rounds. However, none of these algorithms provide guidelines on how to identify the optimum number of data transmission rounds before re-clustering. EDAC (Wang et al. 2004b) is an algorithm which extends the LEACH with energy driven CH rotation instead of original time driven method. However, EDAC also suffer with the problems found in LEACH such as non uniform cluster formation, inability to produce required number of clusters and the entire network having the possibility of being served by even one CH. If such an adverse scenario is created at the first time the clusters are formed, then this can be propagated to the future as there is no re-clustering of the entire network but just a handing over of the CH role to a suitable node. This would result in bad life time performance of the EDAC similar to LEACH in most occasions.

Based on these studies we propose a new weight based dynamic distributed clustering algorithm with energy driven CH rotation which specially address the drawbacks identified in existing algorithms. In this article, we present this new Energy Driven Cluster-Head Rotation (EDCR) algorithm (Gamwarige & Kulasekere 2005, 2007) which has made sure to produces Local Energy Balancing i.e. nodes in a given local neighborhood guarantee to deplete their battery energy at the same rate. Even though EDCR is weight based we specifically gave attention to reduce overhead of control and coordination messages, whereby managing to get better results as opposed to HEED. For this purpose the algorithm uses minimum communication within a limited neighborhood to select the node which has the most residual energy as the CH i.e. the algorithm guarantees there are no any other nodes with higher energy than the CH in its neighborhood. This would result in very low communication overheads during the selection phase of the algorithm. Furthermore, as the results will indicate, this has allowed the formation of well distributed CHs in the system similar to the ones found in HEED and ANTCLUST. Further, the reduction in energy consumption of nodes is achieved by initiating this localized communication protocol only at the point of CH rotation. This is the key factor that has lead to reducing the overheads compared to other algorithms.

The rest of the article is organized in the following manner. In Section 2 we discuss the preliminary details related to the WSN model, energy consumption model of a sensor and the assumptions related to the lifetime of the network. In Section 3 we discuss the salient features of the proposed algorithm. Section 4 will carry out a discussion of how to find the optimal values for the EDCR algorithm parameters. The complexity, correctness and behavior of the algorithm is analyzed in the Section 5. In Section 6 the simulation results in a comparative form for several algorithms including the proposed EDCR algorithm is presented. We give our conclusions and proposed future work in Section 7.

2 SENSOR NETWORK MODEL

The preliminary assumptions used to model the WSN are identical to the previous literature (Heinzelman et al. 2002; Kamimura et al. 2004; Younis & Fahmy 2004; Smaragdakis et al. 2004). In summary they are given below.

2.1 *Assumptions*

1. All nodes have the equal processing and communication capabilities. Further each node is equipped with same size of batteries.
2. Base Station (BS) does not have any energy limitations. We assume 100% reliability and availability of the BS due to the fact that many to one communication. BS can reach any CH asynchronously and has the ability to command them. CH to BS communication is contention based MAC.
3. Time Division Multiple Access (TDMA) scheduled direct data transmission from non CH member nodes to its CH. TDMA is appropriate due to its simplicity, low overhead, short communication duty cycle, and no packet collisions. TDMA is only effective for scenarios in which the number of transmitting nodes is relatively stable over time (Hoang & Motani 2007). This is true in periodic data gathering from regular member nodes to their CH. Direct transmission between regular members to its CH is more energy efficient for small to medium size clusters due to short distance. Further energy requirement for receiving is comparable to typical intra cluster transmissions of a small to medium size cluster. Therefore we can save energy of regular nodes by using TDMA as it allows switching off receivers most of the time eliminating idle listening.
4. All nodes use contention-based MAC protocols during cluster setup phase.
5. Availability of symmetric radio communication model, i.e if a node can reach another node then the second can reach the first using same amount of energy.
6. Nodes have the capability of adjusting the transmission power (Chipcon 2006).
7. The required transmitting power is calculated based on the received signal strength, i.e. the availability of Receive Signal Strength Indicators in the motes (Chipcon 2006).
8. Sensor nodes are uniform randomly distributed in a given rectangular region. This implies that the proposed algorithm can be applied for a WSN deployed in ad hoc manner. The assumption that all nodes are uniform randomly distributed implies that the node distribution is 2D Poisson point process with intensity $\lambda = N/A$ where N is the total number of nodes and A is the distributed area (Mhatre et al. 2005).
9. Sensor nodes can aggregate or fuse multiple data packets to one packet. This implies that in network data aggregation is possible at CH nodes. Most of other WSN clustering algorithms of the similar class (e.g. LEACH, HEED, ANTCLUST etc) have considered infinite data compressibility at the CH node with the assumption of perfect data correlation. This is acceptable for a scenario in identifying the maximum, minimum or average value of a

given type of sensor measurement. On the other hand (Zhao & Liang 2007) has proposed a more realistic data correlation model named Exponential Data Correlation in which the aggregated data packet of two nodes apart from r distance is given by $l\,(1+\eta)$ where l is the length of data packet from each node and $\eta = 1 - e^{-\alpha r}$ in which $0 < \alpha < 1$. $\alpha = 0$ is identical to the perfect data correlation.

It should be noted that we have relaxed the following two assumptions used in many of the existing literature.

1. Homogeneous energy of nodes: Algorithms like LEACH performs well only when all the nodes are equipped with equal energy batteries at the beginning. Hence LEACH cannot be used in already deployed networks or in a repaired WSN, i.e. some new nodes have deployed to an existing network.
2. Location awareness of nodes: Some algorithms like ANTCLUST based require nodes to know its location. In order to achieve this either use of GPS systems or extra algorithms to do node localization is required. Either method is energy consuming and costly.

2.2 *Energy Consumption Model*

The wireless transceiver circuit energy consumption model is given by equations (1) and (2). In this model a sensor node consumes E_{elec} (nJ/bit) energy at the transmitter or receiver circuitry and ε_{amp} (pJ/bit/m^2) energy at the transmitter amplifier. The ε_{amp} and n are in line with the radio propagation path loss constant and exponent respectively of a given environment. A sensor node expends energy $E_{Tx}(l,d)$ or $E_{Rx}(l)$ in transmitting or receiving a l bit message to or from distance d respectively.

$$E_{Tx}(l,d) = E_{elec} \times l + \varepsilon_{amp} \times l \times d^n \quad (1)$$

$$E_{Rx}(l) = E_{elec} \times l \quad (2)$$

Furthermore a CH node consumes E_{DA} (nJ/bit/message) energy in aggregating multiple sensor data. This energy consumption model follows Heinzelman et al. 2000, 2002.

Note: In reality n may vary between 1.8 to 6 depending on the environment conditions. Typical theoretical modeling of environments $n = 2$ is named as Free Space (FS) model and $n = 4$ as Multi-path Fading (MF) model.

2.3 *Life Time of the Sensor Network*

The definition of the life time of a WSN depends on the application where the sensors are deployed. There are three commonly used definitions in the literature (Younis & Fahmy 2004; Handy et al. 2002).

- First Node Dies (FND): This definition is appropriate in situations where death of a single node deteriorates the quality of the network. E.g. Intrusion Detection systems.
- Percentage of Nodes Alive (PNA): Time until a certain percentage of nodes are still alive. This definition is more appropriate for most of the applications with a requirement for a certain percentage of nodes alive for the network to output credible information. Here we assume that some of the sensors are producing correlated data so that some amount of redundancy is built into the network. The Half of the Nodes Alive (HNA) metric is a special case of this.
- Last Node Dies (LND): Though this parameter can be considered as a way to measure the lifetime of a WSN its practical applicability is very limited.

The goal of any good self organizing WSN protocol is to increase the life time of all sensors in the network. As shown in Figure 1, the ideal situation is represented when all sensors die at the same time. Thereafter a new set of sensors may be deployed without replacing some of them. In general, ad hoc WSN are deployed in areas where the sensors are hard to reach after deployment. Hence selective sensor replacement is not practical. Hence in this paper we use the PNA and FND metrics to measure performance of the WSN. In the case of PNA we have assumed 95% of nodes alive which balances the quality of the information gathered and the correlation between the information gathered by the sensor nodes in the network. However the exact % is very much application dependant. In summary the objective can be defined as shifting of the knee point of the graph given in Figure 1 to right while maintaining a right angle at the knee point.

2.4 *Objectives*

Now we outline the base objectives of our algorithm.

1. CH will be the node with highest residual energy in a given neighborhood.
2. A node joins a closest CH with most residual energy using local information.
3. The CH rotation is initiated if one of the existing CHs find that it does not have enough energy to continue its role.

4. The CHs should be well distributed.
5. All decisions are distributed using local information.

Collectively we aim to achieve the WSN lifetime to follow the ideal situation shown in Figure 1.

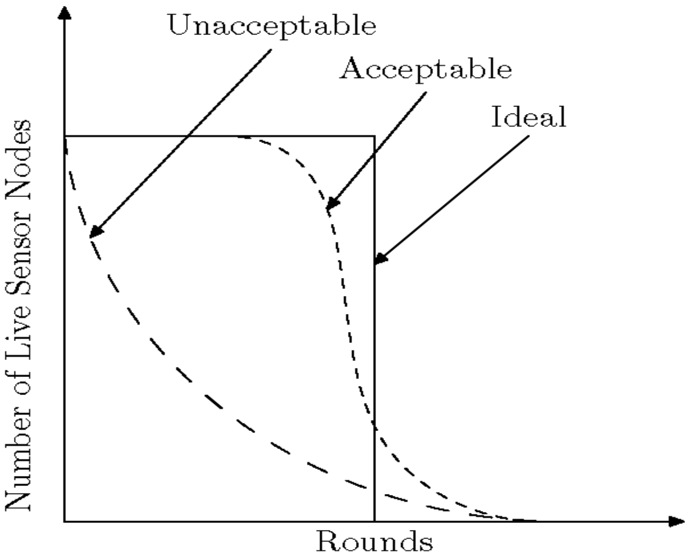

Figure 1: # of Live Sensor Nodes at the End of Each Round

3 OVERVIEW OF THE ALGORITHM

3.1 *Nomenclature*

Table 1 gives a brief definition of some notations used in the proposed EDCR algorithm for the ease of understanding.

Table 1. Brief description of some notations used in the algorithm

P_i	$\in [0,1]$ represents the relative position of the node i with respect to the other nodes in its neighborhood in terms of its residual energy level.
$\mathcal{N}_i^{\mu}$	The set of sensor nodes within a neighborhood of radius μ from node j excluding the node j.
$\mathcal{H}$	The set of all CHs at a given moment.
$\mathcal{M}_i$	The set of member nodes in a cluster headed by CH i including itself.
$\mathcal{SN}_i$	CH i's second degree neighborhood.
$E^t_{res_i}$	Residual energy of node i at given any time instance t.
$E^t_{res_i}\|_{t=\tau}$	Residual energy of node i at $t = \tau$.
λ_i^{τ}	Dynamic energy threshold value of a given CH node i which becomes a CH at time $t = \tau$.When its residual energy drops below this value it calls for a new CH selection phase with the help of the BS.
$d_{x,y}$ $\|x - y\|$	Euclidean distance between two points/locations x and y.
$P_{RX_{i,j}}$	The received signal strength of the signal transmitted by node i at the node j.
P_{TX_i}	The transmitted signal strength of a data packet by node i.
c	The CH role rotation triggering dynamic energy threshold level calculation parameter.
R	CH candidacy broadcasting range.

The proposed EDCR algorithm has CH Candidacy, CH Selection, routine Data Gathering and CH Rotation phases. In the following subsections we discuss each phase in detail.

3.2 *Cluster Head Candidacy*

CH selection is done using the descending price auction, also known as Dutch auction (Zhao & Liang 2007) principle. Dutch auction principle can make sure that the most suitable CH i.e. the node with highest residual energy in a given neighborhood is selected with minimum energy overhead without having multiple iterations. This is realized as follows.

All sensor nodes initially consider themselves as potential candidates of being a CH. However a sensor node with more residual energy has a chance to advertise its candidacy earlier than others within a neighborhood of R. Those sensor nodes that receive an advertisement from any other sensor node will abandon their quest to become a CH. This ensures that a node with a higher residual energy always ends up being a CH within its neighborhood R.

Assume that the CH advertisement phase is limited to a time interval of T time units and that the sensor node i announces its candidacy within a radius of R at a time instance T_i given by equation (3)

$$T_i = T(1 - P_i) + k_i \quad (3)$$

Here k_i is a random time unit introduced to reduce the possibility of collisions among sensor node advertisements with identical P_i in the same neighborhood and $P_i \in [0,1]$ represents the relative position of the node i with respect to the other nodes in its neighborhood in terms of its residual energy level. In other words, the node i with the highest residual energy would be assigned the largest value of P_i in a given neighborhood R. Hence from equation (3), this node will have a T_i which is the smallest in the neighborhood resulting in it being chosen as the CH. Furthermore, the initial conditions that apply to equation (3) are different for homogeneous and heterogeneous sensor networks.

1) *Homogeneous sensor network:* For the initial round $P_i = 1$ is assumed $\forall\, i$ since, all sensors are considered to be equipped with similar batteries and hence equal in residual energy. Then $T_i = k_i$ from equation (3). For all subsequent rounds $P_i \leq 1 \,\forall\, i$ and the sensor

node with the smallest T_i found using equation (3) will broadcast its CH candidacy.

2) *Heterogeneous sensor network:* For heterogeneous WSNs we assume that the initial state to have $P_i \leq 1\ \forall\ i$ and same as for each subsequent rounds. This is a direct result of the WSN having dissimilar residual energies at deployment or the algorithm is applied to an already existing WSN.

Calculation of P_i for different rounds is given by equation (6) of Section 3.5. The neighborhood R is computed assuming that the WSN will consist of an optimum number of clusters k_{opt} as discussed in Section 4.1. For each node j, the set of sensor nodes within a neighborhood of radius μ from j is denoted by $\mathcal{N}_j^{\ \mu}$. Furthermore we define a set $\mathcal{H}$ where

$$\mathcal{H} = \{i | \text{set of all nodes } i \text{ where node } i \text{ is a CH}\}$$

Observation: - For any node i with $j \in \mathcal{N}_i^R$ we have $T_i < T_j \Rightarrow E^t_{res_i} > E^t_{res_j}$. Where $E^t_{res_i}$ and $E^t_{res_j}$ are the residual energies of nodes i and j at this moment.

3.3 *Cluster Head Selection*

Any node j which is not a CH will select its CH CH_j using (4).

$$CH_j = \left\{ i \middle| \max_{i \in \mathcal{H} \cap \mathcal{N}_j^R} D_{i,j} \right\} \tag{4}$$

where

$$D_{i,j} = E^t_{res_i} \frac{P_{Rx_{i,j}}}{P_{Tx_i}} \tag{5}$$

Here $P_{Rx_{i,j}}$ and P_{Tx_i} represents the received signal power from node i to node j and the transmitted power of the advertisement message for node i respectively. The CH advertisement message will contain both $E^t_{res_i}$ and P_{Tx_i} which will be used in (4). Furthermore, $D_{i,j}$ will achieve the following:

1. $E^t_{res_i}$ will allow us to select a CH node with higher residual energy over other CH nodes with lower residual energy. For example if we have CH nodes at an equal distance from j, the factor $\frac{P_{Rx_{i,j}}}{P_{Tx_i}}$ will be constant. Hence the dominating factor would be $E^t_{res_i}$ and as a result the higher energy node will be selected. This will prolong the weaker CH nodes' network lifetime since now the high energy node has taken the burden of processing an additional node.
2. $\frac{P_{Rx_{i,j}}}{P_{Tx_i}}$ allows us to select the closet CH node which will help to reduce the energy consumption of node j. For example if we have CH nodes' that have equal $E^t_{res_i}$ but are placed at unequal distances from j, the CH node which is closer will be selected. This will prolong the lifetime of sensor node j since the node j will be using lesser power P_{Tx_i} to reach the CH in all subsequent communications. Furthermore since
$$P_{Rx_{i,j}} \propto \frac{P_{Tx_i}}{d^n_{i,j}}$$
the resulting $D_{i,j}$ can be used with any communication model.

The combination of the above facts will ensure that effectively prolong the life time of the entire WSN. Furthermore the CH node i calculates a dynamic threshold λ_i^τ based on the current residual energy condition of the node at the time $t = \tau$ i.e. the moment it broadcasts its CH Candidacy using following formula.

$$\lambda_i^\tau = c.E^t_{res_i}|_{t=\tau}$$

where $c \in [0,1]$ is a predetermined constant. The use of such a threshold to generate an event driven CH rotation will be explained in Section 3.4.

Subsequently a CH j calculates its TDMA schedule for the nodes who joined its cluster and broadcast the schedule among them. Apart from the slots allocated for each member node in its cluster, the TDMA schedule will have a time slot reserved for the CH to send any messages to its members if any. This slot will also be used to send control information if any. In a normal data gathering round this slot will not carry any communication and will not generate overheads that will expend energy. However, the CH will use this time slot to update its members at the time of a CH rotation. All the member nodes will keep awake during this time slot to identify if there are any control messages from the CH.

Note: We can define a given CH i cluster as a set of nodes including CH i as $\mathcal{M}_i \ \left(\subseteq \left(\mathcal{N}_i^R \cup i\right)\right)$ given by

$$\mathcal{M}_i = \left\{ j \middle| \begin{array}{c} \text{set of all nodes } j \text{ such that} \\ D_{i,j} > D_{k,j} \text{ where } j \in \mathcal{N}_i^R \\ \text{and } k \in \left(\mathcal{H} \cap \mathcal{N}_j^R\right) \end{array} \right\} \bigcup i \quad ; \text{ for any } i \in \mathcal{H}$$

3.4 *Data Transmission*

The next phase of the algorithm is data transmission where the nodes go into normal routine operation of periodic data gathering. Non CH nodes $j \in (\mathcal{M}_i \setminus i)$ send their data in the allotted time slot according to the TDMA schedule to their CH i. The CH uses a data fusion algorithm to merge the received data from its cluster $\mathcal{M}_i$ before sending to the BS. (*Note*: During this period algorithm refrain from exchanging control messages which results an overhead.)

3.5 *Cluster Head Rotation*

When a CH node i finds its residual energy falling below the threshold value λ_i^{τ}, it triggers a new CH candidacy event by informing the BS that it is unable to perform its duties as a CH any more. Subsequently the BS will inform this to all other CHs thus initiating a CH rotation phase. (Note that most of the previous energy aware WSN clustering algorithms such as LEACH, HEED, SEP and ANTCLUST have a predetermined time point to initiate a CH rotation phase.) Then all CHs use their immediate next chance in the TDMA slot to communicate this fact to its neighborhood, and further request nodes to send their residual energy along with the data in its allotted slot.

A CH i computes the maximum residual energy component of its cluster $\mathcal{M}_i$, using

$$E_{res_i,max} = \max_{j \in \mathcal{M}_i} \{E_{res}\}$$

It will then broadcast this information to all nodes $j \in \mathcal{H} \cap \mathcal{N}_i^{2R+\epsilon}$. Here $\epsilon \left(\leq \frac{1}{2\sqrt{\lambda}}\right)$ [1] is a small positive number which represents a degree of uncertainty when computing the distance to neighbor CHs. Furthermore $2R + \epsilon$ would be the maximum expected distance from a given CH to any of its immediate neighbor CHs as shown in Figure 2. Based on this CH i can get access to the maximum residual energy information of its second degree neighborhood $\mathcal{SN}_i$, where

$$\mathcal{SN}_i = \left\{ \mathrm{j} \,\middle|\, \mathrm{j} \in \left(\bigcup_{\mathrm{k} \in \mathcal{H} \cap \mathcal{N}_i^{2\mathrm{R}+\epsilon}} \mathcal{M}_k \right) \bigcup \mathcal{M}_i \right\}$$

This $\mathcal{SN}_i$ can extend up to $3R + \epsilon$ in any direction to the CH i. Then the CH i updates its member base with the highest available residual energy level of it's $\mathcal{SN}_i$ in the immediate next TDMA slot and triggers a cluster formation phase.

The use of $\mathcal{SN}_i$ information to derive the relative energy position P_i of a node i is more meaningful since it will dispel any ambiguity when it comes to nodes at a border of two clusters. Furthermore it guarantees that a given node will know its residual energy level with respect to its immediate neighborhood or even further. The relative residual energy level is computed using

$$P_i = \frac{E_{res_i}}{E_{res_i,sup}} \tag{6}$$

where

$$E_{res_i,sup} = \max\left\{ \max_{j \in \mathcal{H} \cap \mathcal{N}_i^{2R+\epsilon}} \left\{E_{res_j,max}\right\},\ E_{res_i,max} \right\}$$

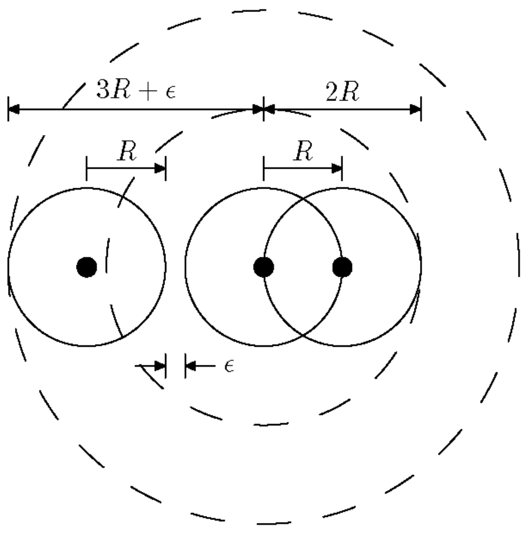

Figure 2: Second Degree Neighborhood

The next step would be to initiate a CH candidacy phase as explained previously. Above described CH rotation process is graphically shown in Figure 3.

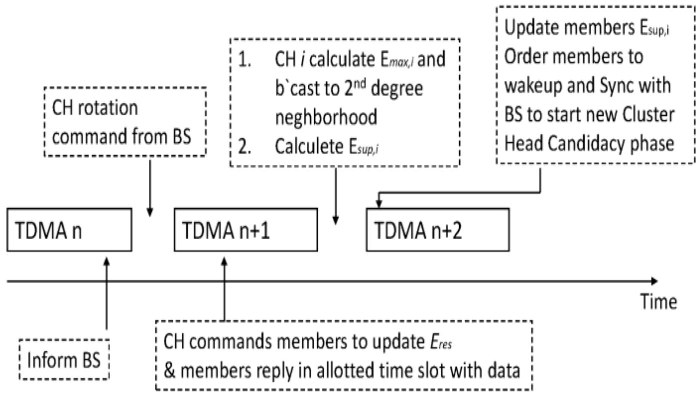

Figure 3: CH Rotation process once it triggered by any CH

Note : It should be highlighted that the number of CH role rotations are significantly below the number of periodic data gathering rounds in a given WSN lifetime.

4 OPTIMUM PARAMETERS OF THE EDCR ALGORITHM

The performance of the EDCR algorithm depends on the proper selection of the CH candidacy broadcasting radius R and proper selection of CH role rotation trigger function λ_i^{τ} parameter c. In what follows we will be deriving optimum values for these parameters to maximize the WSN lifetime.

[1] This is a reasonable upper bound assuming that the maximum distance would result when there are two nodes on the boundary of the neighboring CH announcement range along the line joining the two CHs. $\frac{1}{2\sqrt{\lambda}}$ is the average distance between any two neighboring sensor nodes assuming them to be 2D Poisson points. Refer Appendix for the proof.

The CH candidacy broadcasting radius R is the main factor which determines the expected number of clusters. As the CH candidacy broadcasting range increases much larger clusters form resulting few clusters covering the entire WSN area. This results in member nodes to communicate far distance to reach their CH and each CH to handle more nodes. On the other hand as R reduces there would be more CHs and most of them need to communicate with the far distance BS result in large energy dissipation. Hence selection of optimum R is crucial for the prolonging the WSN lifetime.

4.1 *Optimum CH candidacy broadcasting range, R_{opt}*

We will follow the data-centric analysis of energy consumption method proposed by Zhao & Liang 2007 in deriving the optimal CH announcement broadcasting range R_{opt}. For this purpose let's consider a WSN with N number of similar sensor nodes uniform randomly distributed in a rectangular region of $a \times b$ with the BS located at (p, q). Let's assume one corner of the rectangular area as the $(0, 0)$ location of the Cartesian map. Let's assume that there are k uniform clusters produced. Let's approximate the shape of a cluster to a disk of radius R_c centered at the CH node. Hence

$$R_c = \sqrt{\frac{ab}{\pi k}} \tag{7}$$

Now let's calculate the total energy cost of transmitting one bit of information from each node to the BS. Let's define $J_{CH(i)}$ as the amount of energy dissipated in the whole system to transfer one bit of information originated from the CH i node and $J_{CM(j)}$ as the amount of energy dissipated in the whole system to transfer one bit of information originated from a cluster member node j belonging to the CH node i cluster.

Since the average distance of any CH i to BS is large we can assume that CH to BS communication follows MF model. Hence we can assume that propagation path loss constant $\varepsilon_{amp} = \varepsilon_{mf}$ and exponent $n = 4$.

$$J_{CH(i)=} E_{DA} + E_{elec} + \varepsilon_{mf}\, d_{i,BS}^4 \tag{8}$$

where first term E_{DA} represents the data aggregation cost and the other two terms represents the cost of transmitting one bit to the BS from CH i.

As individual clusters are of small to medium in size we can assume that the FS propagation model would follows in intra cluster communication where propagation constant $\varepsilon = \varepsilon_{fs}$ and propagation exponent $n = 2$.

$$\begin{aligned} J_{CM(j)=} \; & E_{elec} + \varepsilon_{fs} d_{i,j}^2 + E_{elec} + E_{DA} \\ & + \eta\left(d_{i,j}\right)\left(E_{elec} + \varepsilon_{mf}\, d_{i,BS}^4 \quad \right) \end{aligned} \tag{9}$$

The first two terms of equation (9) represents the energy cost of the cluster member j in sending a bit to its CH i, next two terms represents the data receiving and data aggregation cost at the CH i of this bit and the rest of the terms refer to the extra energy required to transfer the aggregated or compressed data bit from the corresponding CH i to the BS. The term $\eta\left(d_{i,j}\right)$ shows the compressibility of the information due to data correlation given by

$$\eta\left(d_{i,j}\right) = \left(1 - e^{-\alpha d_{i,j}}\right) \tag{10}$$

where $0 < \alpha < 1$. Hence we can denote the expected total energy cost of the entire sensor network to collect one bit of data from each node as

$$J_{total} = k\left(J_{CH} + \left(\frac{N}{k} - 1\right) J_{CM}\right) \tag{11}$$

where J_{CH} and J_{CM} are the average energy cost of sending one bit of information generated from a CH and a cluster member respectively. J_{CH} and J_{CM} are given by equations (12) and (13) respectively.

$$J_{CH} \;=\; E_{DA} + E_{elec} + \varepsilon_{mf} E[d_{i,BS}^4] \tag{12}$$

$$\begin{aligned} J_{CM} = \; & E_{elec} + \varepsilon_{fs} E\left[d_{i,j}^2\right] + E_{DA} + E_{elec} \\ & + E\left[\eta\left(d_{i,j}\right)\right]\left(E_{elec} + \varepsilon_{mf} E\left[d_{i,BS}^4\right]\right) \end{aligned} \tag{13}$$

where

$$E\left[d_{i,j}^2\right] = \int_0^{2\pi}\int_0^{R_c} \frac{r^2}{\pi R_c^2} r dr d\theta = \frac{R_c^2}{2} = \frac{ab}{2\pi k} \tag{14}$$

$$E\left[d_{i,BS}^4\right] = \int_0^b \int_0^a \frac{((x-p)^2 + (y-q)^2)^2}{ab} dx\, dy \tag{15}$$

$$\begin{aligned} E\left[\eta\left(d_{i,j}\right)\right] &= \int_0^{2\pi}\int_0^{R_c} \frac{\left(1 - e^{-\alpha r}\right)}{\pi R_c^2} r\, dr\, d\theta \\ &= 1 + \frac{2\pi k}{\alpha^2\, ab}\left(e^{-\alpha\sqrt{\frac{ab}{\pi k}}}\left(1 + \alpha\sqrt{\frac{ab}{\pi k}}\right) - 1\right) \end{aligned} \tag{16}$$

The optimal value for the k with respect to the energy consumption can be found by setting $\frac{\partial J_{total}}{\partial k} = 0$. According to equation (15), $E\left[d_{i,BS}^4\right]$ is invariant of k. Using this fact with equations (11), (12) and (13) we can write

$$\begin{aligned} \frac{\partial J_{total}}{\partial k} &= J_{CH} - J_{CM} + (N-k)\,\frac{\partial J_{CM}}{\partial k} \\ &= \varepsilon_{mf} E\left[d_{i,BS}^4\right] - E_{elec} - \frac{\varepsilon_{fs} ab}{2\pi k} - E\left[\eta\left(d_{i,j}\right)\right]\left(E_{elec} + \varepsilon_{mf} E\left[d_{i,BS}^4\right]\right) \\ &\quad + (N-k)\left(-\frac{\varepsilon_{fs} ab}{2\pi k^2}\right. \\ &\quad \left. + \frac{\partial E\left[\eta\left(d_{i,j}\right)\right]}{\partial k}\left(E_{elec} + \varepsilon_{mf} E\left[d_{i,BS}^4\right]\right)\right) \end{aligned} \tag{17}$$

where

$$\frac{\partial E[\eta(d_{i,j})]}{\partial k} = \frac{2\pi}{\alpha^2\, ab}\left(e^{-\alpha\sqrt{\frac{ab}{\pi k}}}\left(1+\alpha\sqrt{\frac{ab}{\pi k}}\right)-1\right) + \frac{e^{-\alpha\sqrt{\frac{ab}{\pi k}}}}{k} \quad (18)$$

Once we substitute equation (16) and (18) to (17) and equate it to zero we can identify the optimum k. Since it is difficult to solve it algebraically we should use a numerical technique in finding the answer.

In our simulation work we will be comparing the EDCR algorithm with other algorithms such as LEACH, HEED and ANTCLUST which has simulated using the assumption of perfect data compressibility. Hence we will use $\alpha = 0$ when we want to compare EDCR with those algorithms. When $\alpha \to 0$ we can show $E[\eta(d_{i,j})] \to 0$ and $\frac{\partial E[\eta(d_{i,j})]}{\partial k} \to 0$. Therefore using equation (18) we can derive

$$k_{opt\,(\alpha\to 0)} = \sqrt{\frac{\varepsilon_{fs}abN}{2\pi(\varepsilon_{mf}E[d_{i,BS}^4] - E_{elec})}} \quad (19)$$

Now we can use the equation (8) to determine the R_c. The relationship with R_c and R can be derived using the equation (22) given in Section 5. According to this

$$|\mathcal{M}_i| = \frac{|\mathcal{N}_i^R|}{2} \Longrightarrow \lambda\pi R_c^2 = \frac{\lambda\pi R^2}{2}$$
$$R = \sqrt{2}R_c \quad (20)$$

Hence we can write the optimum CH broadcasting range for a perfect data correlation system as

$$R_{opt\,(\alpha\to 0)} = \left(\frac{8ab(\varepsilon_{mf}E[d_{i,BS}^4] - E_{elec})}{\pi N\varepsilon_{fs}}\right)^{\frac{1}{4}} \quad (21)$$

Note : Above equation is derived assuming MF model follows between CH to BS communication. Most literature consider $\varepsilon_{mf} = .0013\text{pJ/bit/m}^4$ and $E_{elec} = 50\text{nJ/bit}$. In such scenario this equation can be applied if the CH to BS distance is more than 87m. However when the WSN dimensions are small and average CH to BS distance is less than this we can expect FS model to dominate the radio propagation. In such condition we can deduce

$$R_{opt\,(\alpha\to 0)} = \left(\frac{8ab(\varepsilon_{fs}E[d_{i,BS}^2] - E_{elec})}{\pi N\varepsilon_{fs}}\right)^{\frac{1}{4}}$$

provided that $E[d_{i,BS}^2] > E_{elec}$.

The R_{opt} computed above will minimize the total energy consumption of the entire WSN. However this would not guarantee that the network lifetime curve would have a sharp edge. In order to achieve this sharp edge it is necessary to evenly rotate the CH role as well. The rotation of a CH role is triggered when any CH i finds that its current residual energy has dropped below λ_i^τ . Therefore proper calculation of this λ_i^τ is essential to achieve the desired objectives of the EDRC algorithm. The following section analyses the effect of λ_i^τ on the lifetime of the WSN.

4.2 *The Effect of λ_i^τ on the WSN Lifetime*

Most of the existing clustering algorithms (Heinzelman et al. 2002; Smaragdakis et al. 2004; Younis & Fahmy 2004; Kamimura et al. 2004) have time triggered CH rotation schemes, i.e. the algorithms rotate the CH role after a predetermined number of data gathering rounds. If a CH selection phase is triggered with a smaller number of data transmission rounds, it will result in excessive overhead during the CH selection phase. On the other hand if the number of data transmission rounds is large before a CH selection is triggered, the CH nodes would not have enough energy to act as ordinary sensor nodes after relinquishing the CH role. Therefore proper selection of optimum number of data transmission rounds is crucial for the system performance in terms of extending the lifetime of the entire WSN. However none of the existing algorithms have addressed this issue even though it is critical for a time based CH rotation algorithm. This issue is irrelevant for energy driven cluster head rotation algorithms such as EDAC (Wang et al. 2004b) and EDCR. However energy driven algorithms do need to determine at what energy level of a CH, it requires changing the CH role. EDCR uses a dynamically calculated CH energy threshold value λ_i^τ using $\lambda_i^\tau = c.E_{res_i}^t|_{t=\tau}$ where $E_{res_i}^t|_{t=\tau}$ is the residual energy of a CH i when it broadcasts its CH candidacy at time $t = \tau$ and $c \in [0,1]$ is a predetermined constant. That is, the number of data gathering rounds can change dynamically depending on the energy of the CH via parameter c. We highlight two such cases.

Case 1. If $c \to 1$ then there will be frequent CH rotations. This allows an even distribution of the CH role among nodes in the WSN where each node expends its energy at the same pace resulting in a sharp edge in the lifetime curve. However frequent CH rotations would result in considerable energy overhead in control and coordination messages during cluster set up. Therefore overall use of energy for useful work will be less. Hence even though the lifetime curve has a sharp edge, the useful lifetime of the WSN is reduced.

Case 2. On the other hand when $c \rightarrow 0$ the CH rotations will be less frequent resulting in low overheads. However now CH nodes would not have enough energy to act as regular nodes after relinquishing the CH role. This would result in a lifetime curve that is less steep. Ideally based on PNA lifetime measurement metric we may select an optimal value for $c = c_{opt}$ as shown in Figure 4. Gamwarige & Kulasekere 2007 has proposed an analytical technique in determining the c_{opt}.

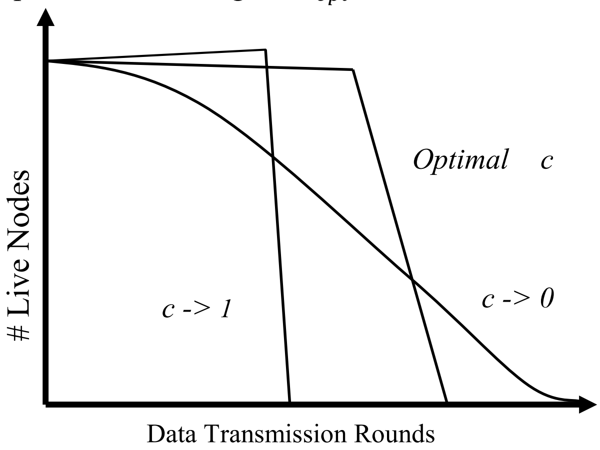

Figure 4: Lifetime of WSN with respect to the change of $\boldsymbol{c}$

5 CORRECTNESS, COMPLEXITY AND BEHAVIOR ANALYSIS OF THE EDCR ALGORITHM

Analysis related to the correctness, complexity and behavior of the proposed EDCR algorithm is presented in this section.

Observation 1. The EDCR algorithm is completely distributed. A node i can either elect to be a CH based on locally calculated candidacy announcement time T_i or join a cluster according to the overheard CH j announcement messages within its neighborhood ($j \in \mathcal{H} \cap \mathcal{N}_i^{2R+\epsilon}$). Each node i calculates T_i based on the information collected from its previous round CH-ℓ's second degree neighborhood $\mathcal{SN}_\ell$.

Observation 2. In the EDCR algorithm a new CH selection phase is initiated by the BS at an instant and subsequently terminated after a fixed amount of time (CH candidacy announcement period, T + *additional time allowed to complete joining a cluster + time taken for the CH to send the TDMA schedule*) irrespective of the number of nodes N, i.e. the time complexity of algorithm is $\mathcal{O}(1)$.

Lemma 1: At the end of the CH candidacy phase a node is either a CH or has identified a cluster in which it can act as an ordinary member node.

Proof. At the beginning of the CH candidacy phase all nodes mark themselves as potential CHs. However for node i if $T_i > T_j$ where $j \in \mathcal{N}_i^R$ it will become a member of CH j. However if $T_i < \min_{\forall j} T_j$ then node i becomes a CH. Further if $T_i = T$ and $\mathcal{N}_i^R = \{\phi\}$ then node i becomes a CH. Based on this by time T all nodes are either a CH or discovered by at least one CH.

Lemma 2: The probability that two nodes within each other's CH announcement range R are both CHs is very small. i.e. CHs are well distributed.

Proof. The only possibility that this can happen is when there are two (or more) undiscovered neighbor nodes i and j (where $j \in \mathcal{N}_i^R$ which implies $i \in \mathcal{N}_j^R$) having $P_i = P_j$. Then the deterministic component of T_i and T_j are the same. However based on k_i and k_j the node which first announce CH message becomes a CH and other will abandon its candidacy quest. If $k_x = (\gamma - \alpha_x)\tau/\gamma$ where $\alpha_x \in [1, \gamma]$ is a random integer such that it has a uniform p.d.f given by $p_{\alpha_x} = (1/\gamma)$ and τ is an appropriate fixed time duration. Then the two nodes i and j making their CH announcement at the same time has a probability of$(1/\gamma^2)$. Similarly n such identical units making announcement at once is $(1/\gamma^n)$. Typically we can assume that a mote has the ability to generate 1000 random discrete numbers. Then the probability of two or more nodes announcing at the same time is less than $.0001\%$.

Observation 3. CH distribution point process: As we have assumed in this research, each ad hoc deployed sensor node represents a Poisson point in 2D space with intensity $\lambda = N/(a \times b)$. Further EDCR does not allow two CHs to be within a distance R. Further according to *Lemma 1* it ensures all the nodes are either discovered by a CH (i.e. there is a CH within a distance R of a regular node) or itself is a CH. The algorithm uses a parameter T_i which represents the inverse of relative energy level P_i of a given node i in its neighborhood. In other words higher the relative node energy lower the T_i of a given node i. Node i with lowest T_i would be elected as the CH in that neighborhood. T_i is purely random during the initial deployment since all nodes would equipped with equal energy batteries. Based on these information we can conclude EDCR algorithm resultant CHs represents a dependent thinning point process on original 2D Poisson point process. Hence we can explain the resultant CH distribution as follows.

Let $\mathcal{S}$ represents the set of all deployed nodes. Where $\mathcal{S}$ is a finite measure subset of $\mathbb{R}^2$ with $|\mathcal{S}| = N$.The clustering process yields a random set $\mathcal{H} \subseteq \mathcal{S}$ of secondary points which we call CHs with the property that $|h_i - h_j| > R$ where $h_i, h_j \in \mathcal{H}$ and $i \neq j$. Note that $\mathcal{S} \setminus \mathcal{H}$ are the non CH member nodes. For any node $m_k \in \mathcal{S} \setminus \mathcal{H}$ we have $|m_k - hi<R$ and $Tmk>Thi$ at least for one CH node $hi \in \mathcal{H}$. We note that m_k is a member of the cluster with CH h_i when $|m_k - h_i| < |m_k - h_j| < R$.

According to Mat´ern 1986 such a dependant thinning process on 2D Poisson points is referred as Mat´ern Type III dependant thinning or hard-core point process. Further Mat´ern 1986 has shown that even though this point process is more close to practical situations and natural phenomena it is mathematically intractable to determine the resultant point distribution. However Bettstetter 2004 has provided an empirical formula for the resultant CH distribution of a different clustering algorithm which also resembles to Mat´ern Type III point process. According to Bettstetter 2004 the resultant point process has a point density of $\lambda_c = \frac{\lambda}{1+\mu/2}$ where $\mu = \pi R^2 \lambda$. In a typical application $\mu \gg 1$. Hence

$$\frac{\lambda}{\lambda_c} \approx \frac{\lambda \pi R^2}{2} \Longrightarrow |\mathcal{M}_i| \approx \frac{|\mathcal{N}_i^R|}{2} \tag{22}$$

In other words this empirical formula shows that a given cluster size is almost half of the size of the CH broadcasting range R covered neighborhood.

Lemma 3: The total overhead in exchanging control messages in the WSN has complexity $\mathcal{O}(N)$.

Proof. EDCR algorithm sends small fixed length control messages during each cluster set up period without iterations as found in HEED. A CH node sends one message each of CH announcement, TDMA schedule announcement, residual energy request from its members, maximum energy level announcement among its cluster members to neighbor CHs and update of its members with the maximum energy level within its second degree neighborhood. Further only one from all CHs will send a CH rotation request message where as all non CH nodes will send cluster join request messages. Furthermore note that answer to residual energy request message is carried out using existing data transmission packet. Hence it is ignored in the computation. i.e. Total Overhead Messages $= 5k_{exp} + 1 + (N - k_{exp})$, where $k_{exp} = E[Number\ of\ CHs]$ and typically $k_{exp} \ll N$. Thus the total message overhead is $\mathcal{O}(N)$.

In LEACH the CHs transmit 2 control messages and non CH nodes send only one message. Total Overhead Messages in LEACH $= 2NP + N(1 - P)$, where P is the percentage of CHs (Wang et al. 2004a). However in LEACH the CH announcement messages must be broadcast to cover the entire WSN where as in EDCR its messages are limited to a radius R and a radius $2R$.

On the other hand a complex algorithm like HEED has total control message overhead of $N_{iter} \times N$ as given in Younis & Fahmy 2004. Typically a high energy node will iterate up to $N_{iter} = 6$ rounds and low energy node may goes beyond $N_{iter} = 15$ rounds.

Considering that $k_{exp} \ll N$ in EDCR and $P < 0.1$ in LEACH we can approximate:

EDCR Total Message Overhead per round $\approx N$
LEACH Total Message Overhead per round $\approx N$
HEED Total Message Overhead per round $\approx N_{iter} \times N$

As previously discussed in the section 1, complex weight based algorithms can achieve a sharp edge in the lifetime curve compared to pure randomized algorithms. However in many of the weight based algorithms the control message overhead is high. This will adversely affect the total lifetime of the system. The goal of most researchers such as Wang et al. 2004a was to derive a good weight based dynamic and distributed clustering algorithm which has sharp edge lifetime curve with low control message overhead similar to a dynamic distributed and random (DDR) algorithm. The EDCR algorithm has archived this goal. The analysis done in this section has proven that the EDCR algorithm has achieved the expected objectives. Further the simulation results presented in the next section confirms us the above analysis.

6 SIMULATION RESULTS

The proposed EDCR algorithm was evaluated using the MATLAB simulation platform. Performance of the EDCR algorithm in terms of the network life time was compared with the existing WSN clustering algorithms. Then the cluster distribution / CH location of the EDCR algorithm was tested using simulation results. Finally the correctness of the proposed analytical methods in deriving R_{opt} and c_{opt} were tested.

6.1 *Comparison with similar algorithms*

We compared the performance of proposed EDCR algorithm with existing WSN clustering algorithms such as LEACH, SEP, HEED and ANTCLUST under both homogeneous and heterogeneous energy networks using theoretical radio propagation models used in existing literature for simulation named Free Space (FS) (Heinzelman et al. 2000; Kamimura et al. 2004) and simplified Multi-path Fading (MF) (Heinzelman et al. 2002; Younis & Fahmy 2004; Smaragdakis et al. 2004) models.

For simulation work we have considered two types of WSN set-ups in terms of the BS location. One type is where the BS is located at the center of the area being monitored. This type of BS set-up allows us to obtain maximum life time for all sensors. This is because the distance between sensor nodes and BS is evenly spread around the average value by locating the BS at the center of the area of interest. However there are instances where the BS cannot be located at the center of the sensor bed. In such a case the average distance to a sensor node is skewed and as a result the WSN life time would be reduced.

The typical energy of a possible battery on a sensor node is much more than the values we used for the simulation purpose. The reason behind use of much small energy level in the battery for simulation is mainly to reduce the time taken to complete the simulation and it does not effect on the final conclusion based on the simulation results. Further almost all other proposed algorithms have used such scale down level of energy content of the batteries for simulation work as well. We have neglected the complexities in associated with the underline WSN MAC protocols of both contention based and TDMA based. We believe that the neglect of this aspect would not have effect on the evaluation of the proposed EDCR algorithm compared with rest of the algorithms of the same class such as LEACH, HEED, ANTCLUST etc. Further we used the perfect data compressibility assumption in these simulations as it was assumed in the previous simulations done by the above mentioned algorithms.

In order to present the comparison of the EDCR algorithm with LEACH, HEED, ANCLUST and SEP we use following scenarios under FS propagation model.

Case FS1: Homogeneous Network of 200 nodes each with 0.5J energy randomly dispersed in a 100×100 region with BS located at (50,50).

Case FS2: Homogeneous Network of 200 nodes each with 0.5J energy randomly dispersed in a 100×100 region with BS located at (50,150).

Case FS3: Heterogeneous Network of 200 nodes with energies 0.3J to 0.8J (randomly assigned) randomly dispersed in a 100×100 region with BS located at (50,50).

Case FS4: Heterogeneous Network of 200 nodes with energies 0.3J to 0.8J (randomly assigned) randomly dispersed in a 100×100 region with BS located at (50,150).

Note:– FS1 and FS2 are situations where applying EDCR for a new WSN. On the other hand FS3 and FS4 are situations where applying EDCR algorithm for an existing WSN possibly with some new nodes to replace malfunctioning nodes. FS1 and FS3 are scenarios where BS is at the center of the area under monitor. FS2 and FS4 demonstrate a situation where it is not possible to set up the BS at the center of the area under monitoring resulting to set up the BS far away from the area of interested.

We set E_{elec} at 50 nJ/bit, ε_{amp} at 100 pJ/bit/m^2 and E_{DA} at 5 nJ/bit/message. Advertisement or setup packets were chosen 60 bits in length and normal data packets were chosen to be 2000 bits long. These values have been chosen to consistent with Heinzelman et al. 2000 and Kamimura et al. 2004.

For simulation purposes we set LEACH as having on average 5% nodes as CHs. EDCR and ANTCLUST is assumed to have a Broadcasting Radius of 25m and 37m respectively for scenario where BS was at (50,50) and (50,150). Additionally for the ANTCLUST algorithm we have assumed 10% as social nodes having a broadcast radius of 15m and 18m for base station locations (50,50) and (50,150) respectively. In computing λ_i^τ we have assumed $c = 0.7$.

Remark:– Note that SEP algorithm assumes two types of energy nodes. Hence it cannot be considered in a strict homogeneous network. On the other hand it cannot be also considered in a heterogeneous network where random energies are assigned to nodes. To overcome this issue we have considered SEP in a homogeneous network where there are 20% of odes having 4 times ($0.5\text{J} \times 4 = 2\text{J}$) extra energy. The simulation results for above mentioned scenarios under FS model are shown in Figure 5.

As we have discussed above some researchers have used simplified MF model in simulations. Therefore we tested the behavior of EDCR algorithm compared with others of similar class under MF model too. The simplified MF model assumes a d^2 model (FS Model) for distances less than $d_0(= 87m)$ and assumes a d^4 model for distance greater than d_0. Typically the intra cluster communication would follow the d^2 model where as the CH to BS communication would follow the d^4model. Hence to ensure that both these communication models are evaluated during the simulation we have taken the BS location to be far away from the Sensor

Bed. To evaluate the effectiveness of the proposed algorithm under the MF model we selected the following two cases:

Case MF1: Homogeneous Network of 400 nodes with 0.5J energy randomly dispersed in a 200×100 region with BS located at $(100,200)$.

Case MF2: Heterogeneous Network of 400 nodes with energies 0.3J to 0.8J (randomly assigned) randomly dispersed in a 200×100 region with BS located at $(100,200)$.

For MF model simulation we used E_{elec} at $50\ \text{nJ/bit}$, ε_{amp-mf} at $.0013\ \text{pJ/bit/m}^4$, ε_{amp-fs} at $10\ \text{pJ/bit/m}^2$ and E_{DA} at $5\ \text{nJ/bit/message}$ as same as used in Heinzelman et al. 2002; Younis & Fahmy 2004; Smaragdakis et al. 2004. Both data and control packet length were same as FS model simulation.

Figure 6 shows number of sensor nodes remaining alive Vs the number of data transmission rounds for the Case MF1 and MF2 listed above. Again we have compared the results of our algorithm with LEACH, HEED, ANTCLUST and SEP in homogeneous energy network scenarios in MP1. In the heterogeneous energy network case i.e. MP2, we have compared with LEACH, HEED and ANTCLUST but not with SEP as it cannot be applied in this condition for the same reason given earlier in this section. For experiment purposes we set LEACH is having on average 5%of cluster heads. EDCR had CH broadcasting range of 55m. The parameter c was set at 0.7 when calculating the λ_i^τ of EDCR. ANTCLUST and HEED too had CH advertisement range of 55m. On average 10% of nodes become Social sensor nodes in ANTCLUST and had a range of 25m.

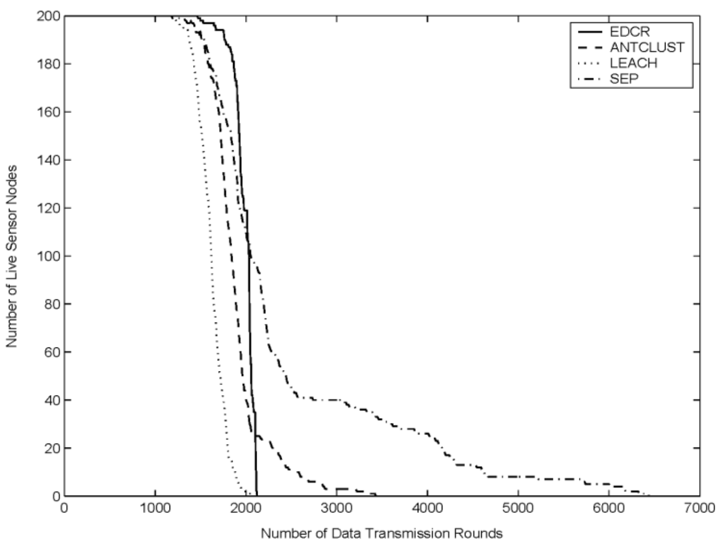

Figure 5.1 Case FS1

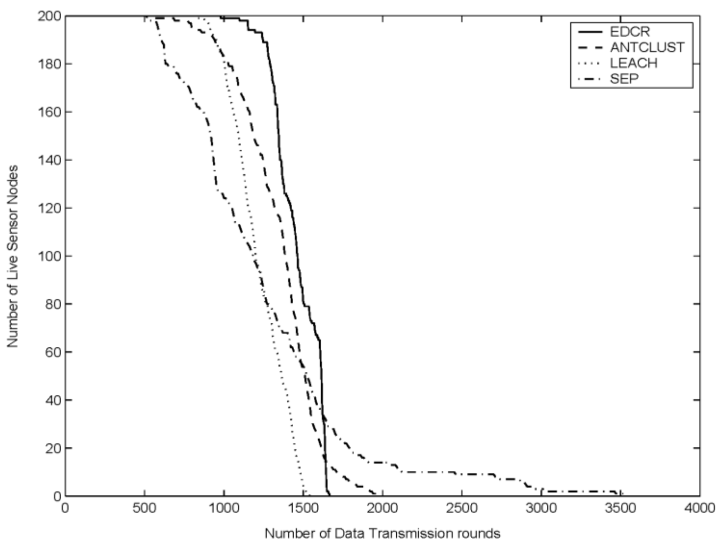

Figure 5.2 Case FS2

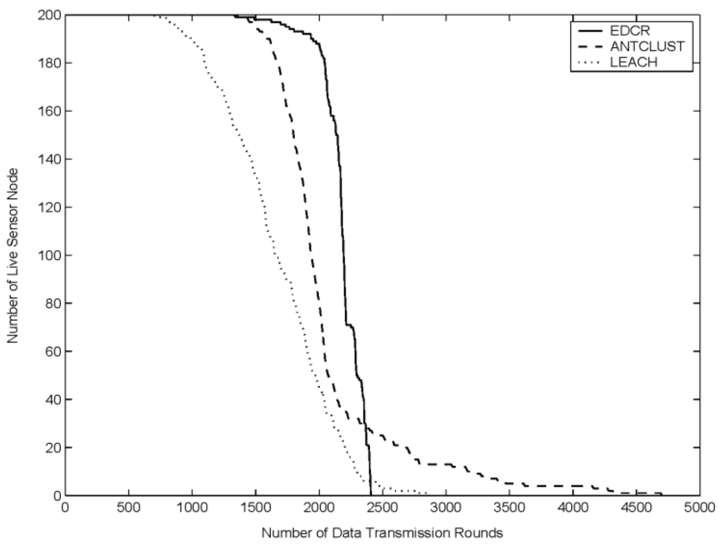

Figure 5.3 Case FS3

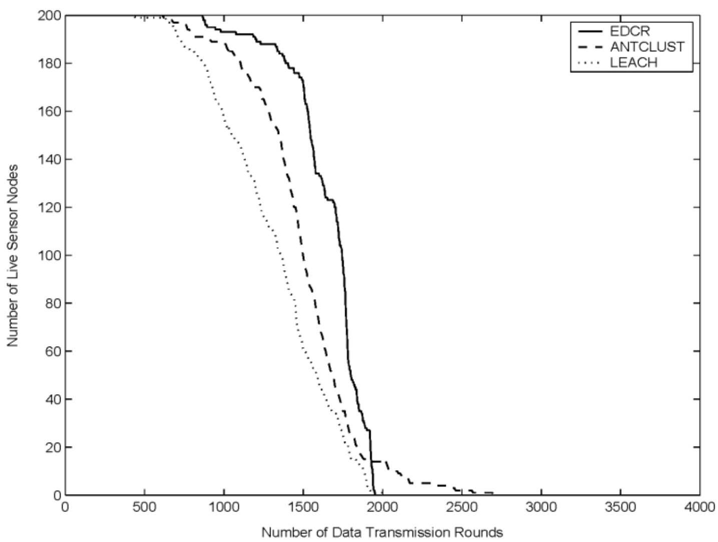

Figure 5.4 Case FS4

Figure 5. 'Number of Live Nodes' vs. 'Data Transmission Rounds' simulation under FS Model

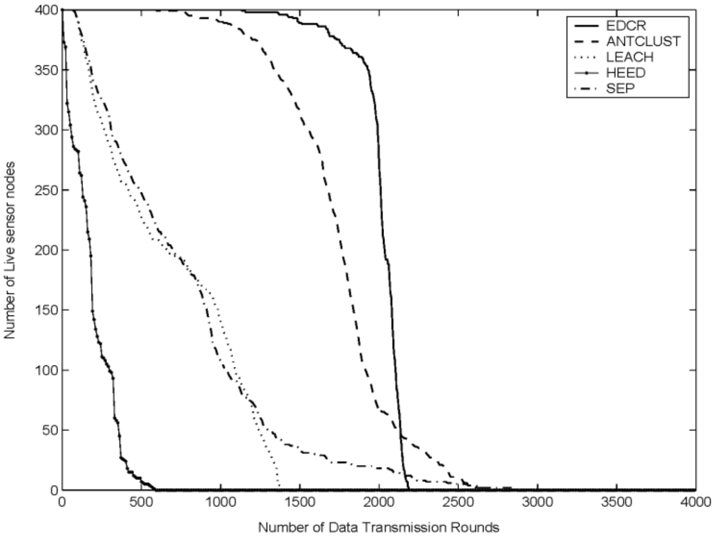

Figure 6.1 Case MF1

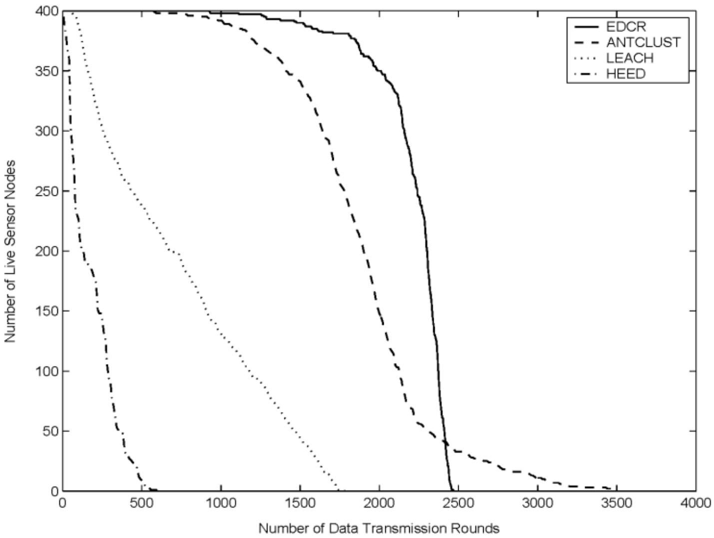

Figure 6.2 Case MF2

Figure 6. 'Number of Live Nodes' vs. 'Data Transmission Rounds' simulation under MF Model

The above simulation results show that the EDCR has outperformed LEACH, HEED, SEP and ANTCLUST based algorithm in both homogeneous and heterogeneous energy WSN scenarios under both radio propagation models with respect to the performance metrics FND and PNA (with 95% nodes alive). The reasons for EDCR to perform much better are its low overhead, energy based CH selection and rotation resulting in even local energy balancing. In MF model the required energy to reach a node or BS over a distance $d(> d_0(= 87m))$ is proportional to d^4. On the other hand in the FS model it is proportional to d^2 irrespective of the distance. Hence under the MF model algorithms with large overhead have much adverse performance especially when the communication distances are over d_0. As we have shown in our analysis the EDCR algorithm has very minimal overhead compared to other algorithms. Hence the performance of the EDCR is much better compared other algorithms under the MF model. Based on these results we can assume that EDCR would have similar performance compared with other algorithms in practical environments where radio propagation path loss exponent $1.8 < n < 6$. Simulation results have proven that the EDCR algorithm has a near ideal lifetime curve with low energy overhead compared to other algorithms of its class.

6.2 *Distribution of Clusters and CH position*

We expect EDCR algorithm to produce well distributed clusters and their corresponding CHs to be located close to the center of cluster area. In order to demonstrate the actual results lets select an arbitrary WSN setup of 200 nodes randomly distributed in an area 250×175 with BS located at the center with $R = 40$. Figure 7 graphically present the cluster setup after 200 data transmission rounds. The diagram shows that the EDCR algorithm produces fairly balanced, well distributed clusters. Further we should expect on average 11.5 nodes per cluster according to the equation (22) given in Section 5. Actual node distribution among different clusters of the WSN setup shown in Figure 7 is tabulated in Table 2. The distribution of nodes among different clusters has a mean of 11.1, standard deviation of 2.5 and a median of 11. Further this tabulation shows us 13 out of 18 clusters have member nodes 11 ± 2. This shows us that the EDCR algorithm produces well distributed even size clusters at any given moment. Further Figure 7 shows that the CHs are located pretty close to the centre of each cluster area.

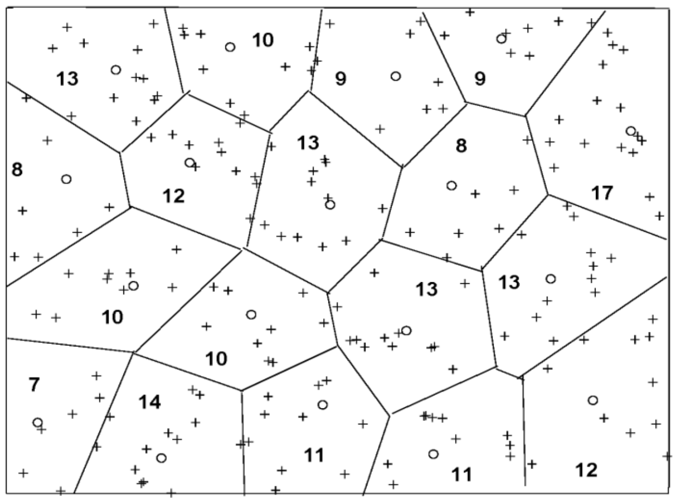

Figure 7. Node distribution among all clusters

Table 2. Distribution of member nodes among different clusters

#M	7	8	9	10	11	12	13	14	17
#C	1	2	2	3	2	2	4	1	1

#M = No Of Members, #C = No Of Clusters

6.3 *Correctness of* R_{opt}

The applicability of using the analytical method proposed in finding R_{opt} for any given WSN setup with different data correlation values $(0 < \alpha < 1)$ was tested using simulation. We have tabulated the results of following scenarios of WSN setup given by
Case 1 : 200×200, BS $(100{,}200)$, 400 nodes
Case 2 : 200×200, BS $(100{,}200)$, 600 nodes
Case 3 : 270×150, BS $(135{,}200)$, 400 nodes
in Table 3. R_{opt}^{t} and R_{opt}^{p} represents the theoretically calculated value and the actual value realized based on simulation experiments respectively. '$\%\Delta_L$' represents the % difference of the life time of the WSN when used the $R = R_{opt}^{t}$ and $R = R_{opt}^{p}$. The results tabulated in Table 3 proves that the R_{opt}^{t} is a good value to be chosen for R to optimize the WSN lifetime.

Table 3. Comparison of the average life with R_{opt}^{t} & R_{opt}^{p}

Case	α	R_{opt}^{t}	R_{opt}^{p}	$\%\Delta_L$
1	0.000	58	55	1.53
	0.001	49	55	3.07
	0.010	28	35	3.78
	0.100	14	15	0.92
2	0.000	55	51	0.11
	0.001	43	41	1.95
	0.010	25	26	1.85
	0.100	12	11	0.13
3	0.000	68	61	3.40
	0.001	55	66	0.84
	0.010	30	36	0.28
	0.100	15	16	0.71

6.4 *Correctness of* c_{opt}

We will demonstrate the applicability of analytical technique proposed in Gamwarige & Kulasekere 2007 in deriving c_{opt} in this subsection. Figure 8 demonstrate how the change of c effect the WSN lifetime. The sensor bed has 200 homogeneous energy sensor nodes each containing 1J energy, deployed in a 100×100 region with BS located at $(50, 150)$ and $R = 43$. We can derive a theoretical c_{opt} of 0.752 when we follow the technique proposed in Gamwarige & Kulasekere 2007. The results shown in Figure 8 confirm that the relevance of this method in finding c_{opt} .

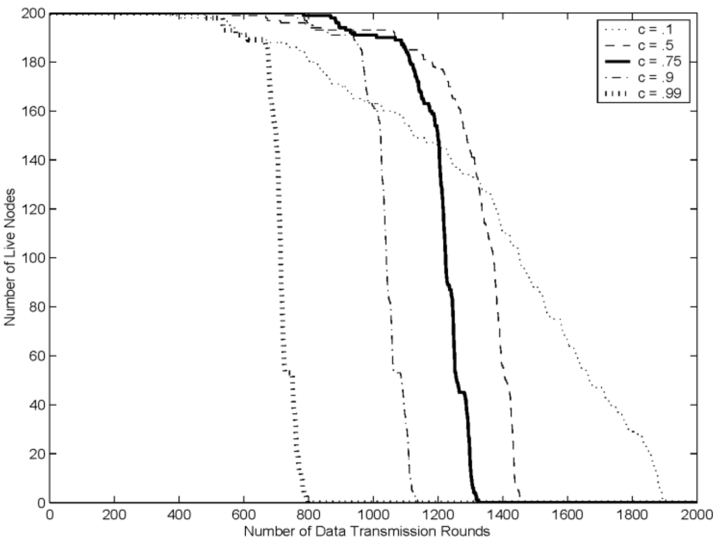

Figure 8. Lifetime Curves of a WSN for different c

7 CONCLUSION AND FUTURE WORK

In this paper we have presented a new energy efficient distributed clustering algorithm named EDCR for ad hoc deployed WSNs that uses the residual energy of sensor nodes for selection and rotation of CHs. This algorithm uses local information up to the second degree neighborhood in order to make these decisions. Further the algorithm is event driven during the CH rotation phase which ensures that the resulting operation is energy efficient. The CH selection mechanism which is based on network topology has ensured that high energy nodes are favored over weaker ones, making even balancing of energy among nodes in a given neighborhood, resulting in extending the useful lifetime of the entire WSN.

When WSNs are deployed in BM applications, lifetime of the sensor network is of crucial importance. In large buildings, the redeployment of the sensor bed can be prolonged as it is costly to selectively replace sensors which die during normal operations. The proposed algorithm will ensure the longest time interval between redeployment when compared with other existing clustering algorithms. The results also propose design strategies which can be used prior to deployment to identify where the sensors can be spread as well.

The investigation of the performance of the EDCR algorithm in hierarchical multi-hop network with global energy balancing would be a useful in addition to the work proposed in this paper.

REFERENCES

[1] Bettstetter, C. 2006. The Cluster Density of a Distributed Clustering Algorithm in Ad-hoc Networks. In Proceedings of IEEE International Conference on Communications, 2004, June 2004, pp. 20–24.

[2]]Chipcon 2006. CC1100 Data Sheet. http://www.chipcon.com/files/CC1100 Data Sheet 1 1.pdf, August 2006.

[3] Cheng, Z., Perillo, M., and Heinzelman, W. B. 2008. General Network Lifetime and Cost Models for Evaluating Sensor Network Deployment Strategies. IEEE Transactions on Mobile Computing, vol. 7, April 2008, pp. 484–497.

[4] Culler, D., Estrin, D., and Srivastava, M. 2004. Overview of Sensor Networks. IEEE Computer, vol. 37, August 2004, pp. 41–49.

[5] Estrin, D., Govindan, R., Heidemann, J. , and Kumar, S. 1999. Next Century Challenges: Scalable Coordination in Sensor Networks. In Proceedings of the Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 1999, pp. 263–270.

[6] Gamwarige, S., and Kulasekere, E.C. 2005. An Algorithm for Energy Driven Cluster Head Rotation in a Distributed Wireless Sensor Network. In Proceedings of the International Conference on Information and Automation (ICIA2005), (Colombo, Sri Lanka), December 2005, pp. 354–359.

[7] Gamwarige, S., and Kulasekere, E.C. 2007. Optimization of Cluster Head Rotation in Energy Constrained Wireless Sensor Networks. In Proceedings of the International Conference on Wireless and Optical Communication Networks WOCN2007, (Singapore), July 2007.

[8] Guru, S. M., Hsu, A., Halgamuge, S., and Fernando, S. 2005. An Extended Growing Self-organizing Map for Selection of Clusters in Sensor Networks. International Journal of Distributed Sensor Networks, vol. 1, April 2005, pp. 227–243.

[9] Handy, M. J., Haase, M., and Timmermann, D. 2002. Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-head Selection. In Proceedings of 4th International Workshop on Mobile and Wireless Communications Network, World Scientific Publishing Co. Pte. Ltd., September 2002, pp. 368–372

[10] Heidemann, J., and Govindan, R. 2004. An Overview of Embedded Sensor Networks. Technical Report ISI-TR-2004-594, USC/Information Sciences Institute, November 2004.

[11] Heinzelman, W. B., Chandrakasan, A.P, and Balakrishnan, H. 200. Energy-efficient Communication Protocol for Wireless Microsensor Networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS), January 2000, pp. 3005–3014.

[12] Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. 2002. An Application-specific Protocol Architecture for Wireless Microsensor Networks. IEEE Transactions on Wireless Communications, vol. 1, October 2002, pp. 660–670.

[13] Hoang, A. T., and Motani, M. 2007. Collaborative Broadcasting and Compression in Cluster-based Wireless Sensor Networks. ACM Transactions on Sensor Networks (TOSN), vol. 3, August 2007.

[14] Ibriq, J., and Mahgoub, I. 2004. Cluster-based Routing in Wireless Sensor Networks: Issues and Challenges. In Proceedings of the 2004 Symposium on Performance Evaluation of Computer Telecommunication Systems (SPECTS'04), 2004, pp. 759–766.

[15] Kamimura, J., Wakamiya, N., and Murata, M. 2004. Energy-efficient Clustering Method for Data Gathering in Sensor Networks. In Proceedings of the First Workshop on Broadband Advanced Sensor Networks, (San Jose, California), October 2004.

[16] Kim , K. T., and Youn, H. Y. 2005. Energy-driven Adaptive Clustering Hierarchy (EDACH) for wireless sensor networks. Embedded and Ubiquitous Computing (Lecture Notes in Computer Science), 2005, pp. 1098–1107.

[17] Mat´ern, B. 1986. Spatial Variation (Lecture Notes in Statistics), 2nd Edition, Springer-Verlag, New York, NY., 1986.

[18] Mhatre, V. P., Rosenberg, C., Kofman, D., Mazumdar, R., and Shroff, N. 2005. A Minimum Cost Heterogeneous Sensor Network with a Lifetime Constraint. IEEE Transactions on Mobile Computing, vol. 4, January/February 2005, pp. 4–15.

[19] Muruganathan, S. D., Ma, D. C. F., Bhasin, R. I., and Fapojuwo, A. O. 2005. A Centralized Energy-efficient Routing Protocol for Wireless Sensor Networks. IEEE Communications, vol. 43, March 2005, pp. 8–13.

[20] Smaragdakis, G. , Matta, I., and Bestavros, A. 2004. SEP: A Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks. In Proceedings of the International Workshop on SANPA, (Boston), August 2004.

[21] Wang, M., Cao, J., Chen, B., Xu, Y., and Li, J. 2007. Distributed Processing in Wireless Sensor Networks for Structural Health Monitoring. Ubiquitous Intelligence and Computing (2007), pp. 103-112, doi: 10.1007/978-3-540-73549-6_11

[22] Wang, Q., Hassanein, H. , and Takahara, G. 2004a. Stochastic Modeling of Distributed, Dynamic, Randomized Clustering protocols for Wireless Sensor Networks," in Proceedings of the 2004 International Conference on Parallel Processing Workshops (ICPPW'04), 2004, pp. 456–463.

[23] Wang,Y., Zhao, Q., and Zheng, D. 2004b. Energy-driven Adaptive Clustering Data Collection Protocol in Wireless Sensor Networks. In Proceedings of the 2004 International Conference on Intelligent Mechatronics and Automation (ICIMA2004), (UESTC, Chengdu, China), August 2004, pp. 599–604.

[24] Wua, J., Yuana, S., Jib S., Zhouc, G., Wanga, Y., and Wanga, Z. 2009. Multi-agent System Design and Evaluation for Collaborative Wireless Sensor Network in Large Structure Health Monitoring. Expert Systems with Applications (2009), doi:10.1016/j.eswa.2009.06.098

[25] Xu, N., Rangwala, S., Chintalapudi, K. K., Ganesan, D., Broad, A., Govindan, R., and Estrin, D. 2004 A Wireless Sensor Network for Structural Monitoring. In Proceedings of the 2nd international Conference on Embedded Networked Sensor Systems (SenSys '04), (Baltimore, MD, USA), November 2004, pp. 13-24.

[26] Younis, O., and Fahmy, S. 2004. HEED: A Hybrid, Energy-efficient, Distributed Clustering Approach for Ad-hoc Sensor Networks. IEEE Transactions on Mobile Computing, vol. 3, October-December 2004, pp. 366–379.

[27] Younis, O., Krunz, M., and Ramasubramanian, S. 2006. Node Clustering in Wireless Sensor Networks:

Recent Developments and Deployment Challenges. IEEE Network, vol. 20, May 2006, pp. 20–25.
[28] Zhao, L., and Liang, Q. 2007. Medium-contention Based Energy-efficient Distributed Clustering (MED-IC) for Wireless Sensor Networks. International Journal of Distributed Sensor Networks, vol. 3, October 2007, pp. 347–369.

APPENDIX

Expected distance between two immediate neighboring nodes

The expected distance $\overline{D}$ of two adjoining sensor nodes can be calculated using the fact that the ad hoc deployed nodes are distributed as a 2D Poisson Point Process with intensity λ. Let's determine the (random) distance D between a node and its nearest neighbor node. For $x > 0$, the cumulative p.d.f. of D is given by

$$F_D(x) = P(D \leq x) = 1 - P(D > x)$$
$$= 1 - P(\text{No other nodes in the disk of area } \pi x^2 \text{ center at itself})$$
$$= 1 - e^{-\lambda\pi x^2}$$

Hence the p.d.f. is

$$f_D(x) = \frac{dF_D(x)}{dx} = 2\lambda\pi x e^{-\lambda\pi x^2}$$

Therefore expected distance between two adjoining nodes $\overline{D}$,

$$\overline{D} = \int_0^\infty x\, f_D(x)dx = \frac{1}{2\sqrt{\lambda}}$$