

Distributed PCA and Consensus Based Energy Efficient Routing Protocol for WSNs

M. BEHZAD¹, M. S. JAVAID², M. A. PARAHCA³ AND S. KHAN¹

¹COMSATS Institute of Information Technology
Islamabad 44000, Pakistan

²University of Hafr al Batin

Hafr al Batin 31911, Saudi Arabia

³Center for Advanced Studies in Engineering
Islamabad 44000, Pakistan

E-mail: muzammil.behzad@ieee.org; {sharjeel188; greatparacha007; salmanmalk7}@gmail.com

Routing protocols with consensus approaches are extensively used for realizing distributed algorithms and have recently found many applications along with use of principal component analysis (PCA). Their vigorous implementation to increase network performance in wireless sensor networks (WSNs) is of extreme significance. In this work, we present our proposed novel routing algorithm; M-BEHZAD, to maximize system performance, stability period and throughput of nodes deployed in a remote field. Distributed principal component analysis (DPCA) based threshold aware communication and fixed clustering method have been used. Selection of cluster heads (CHs) and division of observation field make our proposed scheme more robust. Mathematical representation of the coverage model and 3-Tier communication design for reducing transmission distance is implemented. This ultimately results in improved system performance and considerably minimizes the coverage and energy holes leading to extended lifespan. Comprehensive simulations results are presented validating the applicability of our approach.

Keywords: coverage holes, energy holes, hemisphere zoning (HZ), principal component analysis (PCA), routing protocol, wireless sensor networks (WSNs)

1. INTRODUCTION

Wireless sensor networks (WSNs) contain small sensor nodes having capability of sensing several ecological characteristics like humidity, temperature, pressure and light. These can deliver effective communication via wireless links. These nodes send the sensed information to base station (BS) using multi-hop or direct transmission. Once BS node receives the sensed information, it forwards that data to end users after preprocessing.

Initialization and distribution of nodes is the first phase in developing a WSN. Usually, the deployed sensors are battery-operated having limited power supply. The nodes are normally deployed in potential environments such as observing precarious applications over battlefield or in any remote or hostile arenas. Once dispersed in fields, they are left unattended making it impossible to recharge or renew their batteries. However, data collected by these nodes is highly critical and may be of strategic importance. Hence, efficient energy consumption in WSNs is a critical measure to form a vigorous network.

Received December 26, 2015; revised March 8, 2016; accepted May 19, 2017.
Communicated by Phone Lin.

Extending network lifespan has been one of the active research areas in WSNs. Many approaches are proposed for optimization of energy consumption. To tackle this problem, efficient utilization of nodes' energies has been carried out by proposing robust protocols. Clustering is an efficient method used to evade inefficient consumption of energy. Dynamic clustering deals with varying clusters and number of associated nodes, while static clustering deals with fixed clusters and number of associated nodes.

Based on the nature of WSNs, it can be categorized either as proactive networks [1] or reactive networks [2]. Nodes in proactive network keep their sensors and transmitters turned on all the time, sense the attribute and transmit reference data. Therefore, they are suitable for applications requiring periodic data observation. In reactive network, sensors respond promptly to sudden fluctuations in desired characteristic and otherwise turn off their transmitters; henceforth, these are appropriate for applications with time response.

Sensor nodes in WSNs have certain capacity of initial energies when deployed. Because of these initial energies of the nodes, a WSN can be classified into homogenous and heterogeneous network. In the former case, nodes deployed in the network field have same initial energies, whereas in the latter case, nodes may have different initial energies. Heterogeneous networks may be further categorized as two-level [3] or multilevel [4].

Data aggregation and transmission are the energy hunger processes in WSNs. Accordingly, they need to be optimized to achieve minimal possible energy usage for longer stability. To handle multidimensional data of larger size, principal component analysis (PCA) has been widely used. PCA mainly aims at identifying patterns by finding the covariance matrix and then computing configurations to decrease dimensions. The aim is to decompose a feature space (information data containing $n \times k$ samples) over a reduced subspace representing the observed information. However, the centralized version of PCA is extremely energy inefficient as compared to its distributed version [5].

In this research work, we propose and assess a novel DPCA based efficient divide-and-rule (DDR) communication technique for WSNs, M-BEHZAD: Maximum residual energy Based Energy efficacy using Hemisphere Zoning with Advanced DDR method for WSNs. We exploit heterogeneous and reactive characteristics to ultimately prolong the network lifetime. We have used Distributed PCA using Consensus Algorithm (DPCA-CA) to estimate the covariance matrix of network and then transmitting the compressed data with restoration at BS. The CH selection uses maximum remaining energy and is synchronized with neighboring CHs using a 3-Tier communication structure. Moreover, we introduce HZ for division of network field and present a mathematical coverage model as well.

Rest of the paper is ordered in the following fashion: Section 2 deals with background, related work done by other authors and their approaches, while Section 3 discusses the need of this work. Detailed implementation of our approach is presented in Section 4, whereas Section 5 explains the experimentations and results. Section 6 finally summarizes the paper.

2. RELATED WORK

W. Heinzelman proposed a routing protocol, LEACH [1], utilizing clustering. The communication system in this protocol is multi-hop. Distribution of CHs is not uniform

as their selection is totally probabilistic resulting in unbalanced distribution of CHs in the network. This eventually leads to a speedy death of sensor nodes. In [2], A. Manjeshwar, *et al.*, proposed a reactive protocol, TEEN, in which sensors used to respond promptly to sudden fluctuations in desired characteristic. Even though it lasted long in terms of network life, it is restricted only to temperature based applications.

Heterogeneous wireless sensor network with different amount of initial energies of sensor nodes was presented by authors of [3, 4]. In [3], two-level heterogeneous network based SEP scheme is proposed, taking care of two types of nodes based on initial energies: 1) normal nodes, and 2) advance nodes carrying additional energy than normal sensors. SEP elongates the network's stability period, defined as time interval till death of first node. But it's not appropriate for extensively used multi-level heterogeneous WSNs that contain more than two nodes having different levels of initial energies. DEEC protocol is proposed in [4] which deploy multi-level heterogeneous WSNs. Nodes in DEEC with high energies are selected to be CHs. Therefore, it attains extended lifespan as compared to SEP.

Authors of [6, 7] implemented static clustering scheme. In [6], authors proposed REECH-ME by taking advantage of static clusters. Their method attains extended network lifespan as compared to LEACH. However, imbalanced areas of clusters result in energy holes. On the other hand, static clustering technique to prolong network lifespan by dividing the circular network into different regions, DREEM-ME, is proposed in [7]. Both of these fails to uniformly allocate sensors to distributed areas. Another routing scheme for WSNs is presented by A. Ahmad, *et al.*, in [8]. They tackled the issue of unbalanced energy consumption generating energy and coverage holes. This scheme beats LEACH [1] and REECH-ME [6] in terms of energy utilization and stability period of WSNs.

A distributed coverage HORA algorithm is proposed in [11]. It discusses poor performance due to non-uniform distribution and generation of energy holes because of frequent sensing. A pixel based method is used to reduce energy consumption. To tackle coverage holes, specific nodes are moved meanwhile retaining coverage area of its neighbors same. This scheme resulted in considerable amount of coverage holes' reduction.

In [12], Y. Gu, *et al.*, proposed Efficient Scheduling for the Mobile Sink in Wireless Sensor Networks with Delay Constraint (ESWC). The authors focused on implementation of sink mobility to increase network stability period. The work provides general and practical unified formulation analyzing jointly the sink mobility, routing, and delay of the network. Their work includes polynomial-time optimal algorithm comparing advantages of mobile and non-mobile sink. Various sink paths and its effects on lifetime, delay and throughput are also discussed by the authors. Authors of [13] proposed two routing protocols for Terrestrial WSNs: HEER and MHEER. CH selection in MHEER is based on max energy of the region nodes and has fixed number of CHs in each round. The authors also implemented sink mobility on these schemes namely HEER-SM and MHEER-SM.

3. MOTIVATION

In this part, we discuss flaws in existing WSNs schemes towards practical scenarios. Utilization of energy in an efficient is the primary goal of such schemes since these ener-

gies are very limited specifically for low energy carrying devices. It's been noted that insignificant consideration has been given to time critical applications proposed in the recent literature. Many of proposed techniques try to decrease number of coverage and energy holes but they assume a sensor network to be proactive. Moreover, the data aggregated and transmitted using PCA results in early death of network.

We are of the opinion that WSNs should be reactive. We also believe that dynamic clustering, TEEN [2], ultimately results in speedy death of nodes. Also, heterogeneous networks extend lifetime of network in comparison with a homogenous network. Consequently, in our work, we have focused to develop a protocol which can satisfy the appetite for practical signal and image processing applications along with several WSNs applications. Specifically, we have used DPCA-CA for data compression and recovery.

4. PROPOSED M-BEHZAD SCHEME DESIGN

Initially, the radio model of communication field, as shown in Fig. 1, is described followed by explanation of HZ based network. Afterwards, DPCA-CA based 3-Tier communication structural design is presented. Subsequently, selection mechanism for CHs and working of proposed protocol is described. After that, we explain the mathematical modeling of energy balanced consumption in various coronas.

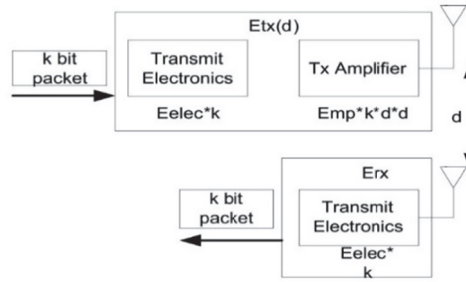


Fig. 1. Radio model.

4.1 Communication Radio Model

We have used the standard first order radio model for nodes' communication as these are widely used in the literature for a practical WSN scenario (see *e.g.* [1, 6-8]). Table 1 summarizes the initializing parameters that we are using. To make our scheme a more practical one, d^2 energy losses due to channel transmission are also considered. Accordingly, following are the expressions for transmitting a k -bit data at a distance d :

$$d_o = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (1)$$

$$\text{for distance} < d_o \quad E_{Tx}(k, d) = k \times E_{elec} + k \times d^2 \times E_{fs} \quad (2)$$

$$\text{for distance} \geq d_o \quad E_{Tx}(k, d) = k \times E_{elec} + k \times d^4 \times E_{mp} \quad (3)$$

$$\text{Energy consumption for reception: } E_{Rx}(k) = k \times E_{elec} \quad (4)$$

Table 1. Radio parameters.

Operations	Energy Dissipated
Circuitry Energy [E_{elec}]	50 nJ/bit
Aggregation/DPCA Energy	5 nJ/bit/signal
Amp. Energy [E_{fs}] (for distance $< d_0$)	10 pJ/bit/4m ²
Amp. Energy [E_{mp}] (for distance $\geq d_0$)	0.0013 pJ/bit/m ⁴

4.2 Network Model Utilizing Hemisphere Zoning (HZ)

In our work, the network area is 100m×100m with BS in the center at $C_p(x_1, y_1)$. We have divided the area by introducing HZ that helps us in an efficient division of network field because this way we can uniformly distribute M nodes for improved performance. The network area is related with real time geographical location and is considered as a sphere of diameter $2\eta\beta$. Using geometry concepts, the reference field is sliced out into horizontal and vertical hemispheres. We call them vertical and horizontal slice HZs.

4.2.1 Formation of clusters

As stated previously, whole network area is divided into $\eta = 3$ (for $M = 100$ nodes) equidistant circular areas each centered at $C_p(x_1, y_1)$. Nine regions are formed using this division followed by uniform random distribution of nodes. The number of coronas are defined by η , and hence the number of clusters are a function of network field area and number of nodes. However, this should be selected keeping in mind the possibility of overlapping regions due to extra clusters that ultimately results in energy and coverage holes. The coronas are hence named as: Internal (I_c), Middle (M_c) and Outer corona (O_c).

4.2.2 Deployment of nodes and division of coronas

As depicted in Fig. 2, each corona is further sliced out into four sub-regions. Regions $M2$ through $M5$ of M_c and $M6$, through $M9$ of O_c are formed. Because of its proximity to BS, region $M1$ is not divided further. Subsequently, uniform random distribution of nodes takes place where each region gets equal number of nodes from 80% of overall nodes while region 1 gets 20% of nodes. For non-uniform distribution, deploying more advanced nodes in such zones is effective as the coverage radius of the advanced nodes is larger. Also, the number of clusters should be increased because it will result in maximum coverage. Making sensor density higher at the center of the terrain will help remove the energy holes but it will invoke increase in coverage holes in the non-center areas. Consequently, a trade-off is needed for sensor deployment to avoid energy and coverage holes to a similar extent. In a checkerboard grid deployment, the sensors either result in improper coverage or overlapping coverage asking for energy holes. Therefore, uniform random deployment with a proper sensor nodes density in the clusters always results in a superior performance.

4.3 3-Tier Communication Architecture

We are utilizing multi-hop scheme for data transmission in our proposed model. Based on its communication methodology, we are naming it as 3-Tier architecture. During Tier-1 phase, all normal nodes, that are not CHs, forward their reference information

to respective region's CHs. During Tier-2 phase, CHs of O_c transmit their data to nearest CHs of M_c . For optimal consumption, CHs of O_c compute distances to the next level CHs and transmit information to nearest cluster heads. For example, CH of M_9 will compute its distance with CHs of region M_2 , M_4 and M_5 , and will transmit its information data to minimum distanced cluster head. During the final Tier-3 phase, nodes of I_c and CHs of M_c send their data to BS. This proposed 3-Tier communication architecture is shown in Fig. 3.

4.4 DPCA-CA

PCA is a way of choosing basis vectors to represent signals by using the statistics or correlation matrix. It mainly aims at identifying patterns by finding the covariance matrix and then computing configurations to decrease dimensions of reference information with least loss of data. The standard, undistributed and centralized version of principal component analysis algorithm needs that all M information samples from all sensor nodes should be collected at BS either by single-hop or multi-hop scheme. By such access to complete information data set Y , BS can execute PCA by finding covariance matrix of the data and then compressing the data. However, this results in significant loss of energy. We can write a signal \underline{x} in terms of N basis expansions as:

$$\underline{x} = \sum_p a_p \underline{\Psi}_p. \quad (5)$$

The objective is to find $k < N$ basis such that the signal can be recovered successfully *i.e.*,

$$\underline{x} \approx \sum_{p=1}^k a_p \underline{\Psi}_p = [\underline{\Psi}_1 \quad \underline{\Psi}_2 \quad \cdots \quad \underline{\Psi}_k] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \underline{\Psi}_k \underline{a}_k. \quad (6)$$

So, we have to minimize the squared error, by choosing the best k , given by

$$e^2(\underline{\Psi}_k | \underline{x}) = \min_{\underline{a}_k} | \underline{x} - \underline{\Psi}_k \underline{a}_k |^2.$$

The problem is thus $\min_{\underline{\Psi}_k} E_{\underline{x}} \{e^2(\underline{\Psi}_k | \underline{x})\}$ subject to the constraint:

$$\underline{\Psi}_k^H \underline{\Psi}_k = \underline{I}$$

where $(.)^H$ represents Hermitian, and \underline{I} is the standard identity matrix. This presents the idea that individual basis vectors are orthogonal and have unit norm which can be written as:

$$\underline{\Psi}_k^H \underline{\Psi}_k = \underline{I} \Rightarrow \underline{\Psi}_m^H \underline{\Psi}_n = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}.$$

As we have the orthogonality condition and basis are unit norm, so we can write $a_k = \underline{\Psi}_k^H \underline{x}$. Consequently, error vector is

$$\begin{aligned}
 e^2(\underline{\Psi}_k | \underline{x}) &= |\underline{x} - \underline{\Psi}_k \underline{\Psi}_k^H \underline{x}|^2 \\
 &= \underline{x}^H (\underline{I} - \underline{\Psi}_k \underline{\Psi}_k^H)^2 \underline{x} \\
 &= \underline{x}^H (\underline{I} - \underline{\Psi}_k \underline{\Psi}_k^H) \underline{x} \\
 &= \underline{x}^H \underline{x} - (\underline{x}^H \underline{\Psi}_k)(\underline{\Psi}_k^H \underline{x})
 \end{aligned} \tag{7}$$

where $(\underline{I} - \underline{\Psi}_k \underline{\Psi}_k^H)$ is an idempotent matrix. So, our problem is to minimize the average value of error squared over \underline{x} , which is equivalent to maximizing the inner product $(\underline{x}^H \underline{\Psi}_k)(\underline{\Psi}_k^H \underline{x})$. Breaking this into columns of $\underline{\Psi}_k$, we can write,

$$\max_{\underline{\Psi}_p, p=1, \dots, k} E \left\{ \sum_{p=1}^k (\underline{x}^H \underline{\Psi}_p)(\underline{\Psi}_p^H \underline{x}) \right\} = \max_{\underline{\Psi}_p, p=1, \dots, k} \sum_{p=1}^k \underline{\Psi}_p^H \underline{R} \underline{\Psi}_p, \tag{8}$$

subject to the constraint: $\underline{\Psi}_m^H \underline{\Psi}_n = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$

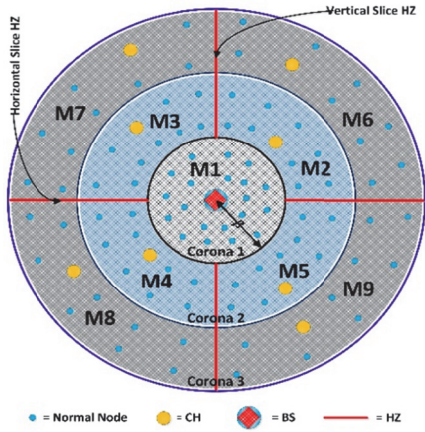


Fig. 2. Network model and nodes deployment.

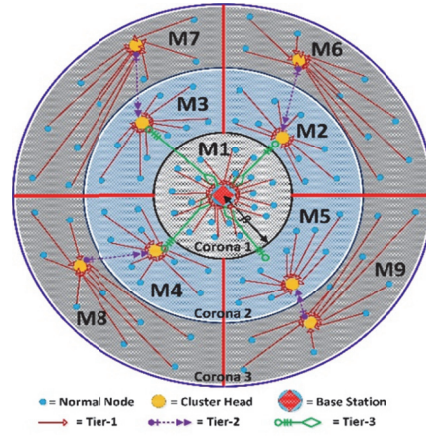


Fig. 3. Communication architecture.

where $\underline{R} = E\{\underline{x} \underline{x}^H\}$. The solution to this is to set $\underline{\Psi}_k$ be the Eigen vector of \underline{R} corresponding to p th largest Eigen value. The Eigen vector relation satisfies the relation $\underline{R} \underline{\Psi}_p = \lambda_p \underline{\Psi}_p$. Here, the mean squared error is $E \{e^2(\underline{\Psi}_k | \underline{x})\} = \sum_{p=k+1}^N \lambda_p$ and the individual principal components (PC) are uncorrelated. More specifically, we can write

$$E \{a_m^* \underline{\Psi}_m^H \underline{\Psi}_n a_n\} = \begin{cases} 0 & m \neq n \\ E\{|a_n|^2\} & m = n \end{cases}$$

because we are assuming that the data is zero-mean. Moreover, these individual components are also uncorrelated which is

$$E \{a_m^* a_n\} = \underline{\Psi}_m^H E \{ \underline{x} \underline{x}^H \} \underline{\Psi}_n = \underline{\Psi}_m^H \underline{R} \underline{\Psi}_n = \underline{\Psi}_m^H \underline{\Psi}_n \lambda_n = \begin{cases} 0 & m \neq n \\ \lambda_n & m = n \end{cases}$$

where $R\Psi_p = \lambda_p\Psi_p$ is being satisfied by Eigen vectors and orthogonality condition is invoked. If we have a zero-mean low rank real valued signal with added white noise given by $\underline{x} = \underline{s} + \underline{n}$. The noise have covariance matrix $E\{\underline{n}\underline{n}^T\} = \sigma^2 I$ and \underline{s} is exactly represented as $\underline{s} = \sum_p a_p \Psi_p = \underline{\Psi}_k \underline{a}_k$. So, estimating \underline{s} with $k < N$ principal components is:

$$\begin{aligned}\hat{\underline{s}} &= \underline{\Psi}_k \underline{a}_k = \underline{\Psi}_k \underline{\Psi}_k^T \underline{x} \\ E\{|\hat{\underline{s}} - \underline{s}|^2\} &= E\{|\underline{\Psi}_k \underline{\Psi}_k^T \underline{s} - \underline{s} + \underline{\Psi}_k \underline{\Psi}_k^T \underline{n}|^2\} \\ &= E\{|\underline{\Psi}_k \underline{\Psi}_k^T \underline{n}|^2\} = E\{|\underline{n}^T \underline{\Psi}_k \underline{\Psi}_k^T \underline{\Psi}_k \underline{\Psi}_k^T \underline{n}|^2\} \\ &= \sum_{p=1}^k E\{|\underline{\Psi}_p^T \underline{n}|^2\} = \sum_{p=1}^k \sigma^2 \\ E\{|\hat{\underline{s}} - \underline{s}|^2\} &= k\sigma^2.\end{aligned}\tag{9}$$

Comparing the above derived result to $E = \{|\underline{x} - \underline{s}|^2\} = E\{|\underline{n}|^2\} = N\sigma^2$, this PCA decomposition reduces the white noise by a factor of $\frac{N}{k}$.

We assume all M node collects T_n samples of Gaussian distribution data G_{1n} and G_{2n} where the global data matrices are G_1 and G_2 of all nodes. For each node, we define local covariance matrix as C_n and the global data covariance matrix as C . As we have,

$$G_1 G_2^T = \sum_{n=1}^M G_{1n} G_{2n}^T = \sum_{n=1}^M \sum_{i=1}^{T_n} \mathbf{g}_{i,1n} \mathbf{g}_{i,2n}^T.\tag{10}$$

Algorithm 1: Principal Component Analysis (PCA)

- Step 1:** Aggregate data for every node Di and make it zero mean
Step 2: Finding covariance matrix $C = M^{-1} Y^T Y$
Step 3: Calculate Eigen values λ_i and Eigen vectors q
Step 4: Using k Eigen vectors of C as principal components
Step 5: Computing compressed data set $Y' = [k] * Y^T$
-

Now the main task is to find a global covariance matrix C , instead of finding for each node individual, that can produce similar results as can be achieved through centralized PCA. We note that by applying the average consensus algorithm, we can achieve the results as can be achieved using PCA at BS, *i.e.*,

$$\begin{aligned}(G_{1n} G_{2n}^T)_c &:= \frac{1}{M} \sum_{n=1}^M G_{1n} G_{2n}^T = \frac{1}{M} \sum_{n=1}^M \sum_{i=1}^{T_n} \mathbf{g}_{i,1n} \mathbf{g}_{i,2n}^T = \frac{1}{M} G_{1n} G_{2n}^T \\ C &:= Y * Y^T = \sum_{n=1}^M Y_n Y_n^T, C_n := Y_n * Y_n^T = \sum_{i=1}^{T_n} y_{i,1n} y_{i,2n}^T \\ C_* &= (Y_n * Y_n^T)_c = \frac{1}{M} \sum_{n=1}^M Y_n Y_n^T = \frac{1}{M} \sum_{n=1}^M C_n = \frac{1}{M} C \Rightarrow C_* = \frac{1}{M} C,\end{aligned}\tag{11}$$

which shows that we can compute the global covariance matrix, by using some rounds of

consensus. This can lead to the performance same as finding PCA at BS but with the lowest energy usage possible. With this, we double the energy efficiency: 1) by compressing the data and transmitting using PCA, and 2) by finding global covariance matrix for PCA in a distributed manner rather than the inefficient centralized method. This ultimately results in a longer stability period and network lifespan. This should be noted that by increasing the consensus rounds, optimum results are achieved but the algorithm normally converges at 3-4 rounds. Algorithm 1 describes PCA while DPCA-CA is presented in Algorithm 2.

4.5 Selection of CHs

Selection of cluster heads is a very essential phase in any routing protocol. We are using fixed/static clusters as already stated, so number of clusters and cluster heads will not change in any round even till death of the network. In each region, except region 1, only one CH will be selected in every round. Therefore, eight CHs are nominated in each round. The selection criteria of CHs in our proposed model is based on maximum residual energy. Moreover, we used synchronized CH selection meaning that selection of cluster head between region M_2 and M_6 , M_3 and M_7 , M_4 and M_8 , and M_5 and M_9 will collaborate to find the best CH. For instance, if cluster head in region M_2 is selected on left side of center reference point of its region then cluster head in region M_6 will also be selected on left side of its respective reference point. This further decreases the communication distance, and helps in reducing energy and coverage holes ultimately prolonging the network lifetime.

Algorithm 2: Distributed PCA with Consensus Algorithm (DPCA-CA)

Step 1: Aggregate data for every node D_i and make it zero mean

Step 2: Finding covariance matrix $C_n = Y_n * Y_n^T$

Step 3: $C_* = (Y_n * Y_n^T)_c = \frac{1}{M} \sum_{n=1}^M C_n$

Step 4: Using k Eigen vectors of C_* as principal components

Step 5: Computing compressed data set $Y' = [k] * Y^T$

4.6 Proposed Protocol Operation

In our proposed system design, we are utilizing reactive and heterogeneous characteristic of a wireless network meaning that sensor nodes with different levels of initial energies transmit their information to BS and CHs based on crossing a pre-determined threshold. We have also used fixed/static clusters whereby dividing our network area in 9 sub-regions. The number of clusters and deployed sensor nodes will not change throughout network lifespan. As stated earlier, these regions are titled as M_1, M_2, \dots, M_9 as depicted in Fig. 3. After the network is initialized, sensor nodes start sensing desired attribute. The transmission of normal nodes is restricted and is subjected to if the pre-defined threshold has been crossed or not. Once the data is collected at CH node, it implements DPCA-CA for compression before transmission. Operation of our proposed method has been presented in Algorithm 3. The threshold aware scheme has been shown in this algorithm while the threshold values are defined in the following subsection.

Algorithm 3: Proposed Protocol Operation

Phase 1: Nodes deployment and network initialization

Phase 2: Record centers of all CH regions

Phase 3: Set nodes N_i of each M_i region

Phase 4: Compute residual energies of N_i nodes of M_i

Phase 5: Choose CHs with maximum energy

Phase 6: **if** $N_i.CH \subseteq O_s$ **then**
 if $N_i.CH \subseteq M_6$ **then**
 $N_i.CH.next.hopCH = \text{nearest}[CH.id \text{ of } M_{1...4}]$
 else $N_i.CH \subseteq M_7$
 $N_i.CH.next.hopCH = \text{nearest}[CH.id \text{ of } M_{1...4}]$
 else $N_i.CH \subseteq M_8$
 $N_i.CH.next.hopCH = \text{nearest}[CH.id \text{ of } M_{1...4}]$
 else $N_i.CH \subseteq M_9$
 $N_i.CH.next.hopCH = \text{nearest}[CH.id \text{ of } M_{1...4}]$
 end if
end if
Phase 7: Avoid coverage holes using coverage model
Phase 8: Call Algorithm 2 for dimensionality reduction
Phase 9: **if** sensed attribute $\geq HT$ **then**
 previous attribute = sensed attribute
 if first time HT crossed **then**
 Transmitter activate and transmit to CH
 else
 if previous attribute – sensed attribute $\geq ST$ **then**
 Transmitter activate and transmit to CH
 end if
 end if
end if

4.6.1 Hard Thresholding (HT)

It's the first value of threshold with which each normal sensor compares its sensed attribute. If any normal node crosses this threshold, it turns on the transmitter and reports to CH. CHs do not need to follow this.

4.6.2 Soft Thresholding (ST)

It's the second value of threshold with which sensor compares its value, only if the sensor has crossed hard threshold. It triggers the sensors to turn its transmitter on and report to CH. A higher value of ST may lead to lower power consumption and higher network lifetime but at the cost of lesser transmissions (which may involve loss of critical data), and vice versa.

4.7 Coverage Model

In our proposed protocol, target field is covering 100m×100m area. We formulate

deployed nodes in terms of set notation as $D = \{d_1, d_2, \dots, d_M\}$. Any node can be represented in terms of its coverage model as a sphere with its center at (m_i, n_i) and radius r_i . This is used to mitigate, to the maximum, coverage and in turn energy holes. A random variable X_i is used to associate the event when a pixel (m, n) is within the coverage range of any node d_i . Consequently, probability of occurrence of the event X_i , as symbolized by $P\{X_i\}$, can be written as the probability of coverage $P_{cov}(m, n, d_i)$. We can decompose it into a bi-valued function as follows:

$$P\{X_i\} = P_{cov}(m, n, d_i) = \begin{cases} 1 \rightarrow \text{if } (m - m_i)^2 + (n - n_i)^2 \leq r_i^2 \\ 0 \rightarrow \text{otherwise} \end{cases} \quad (12)$$

This means that any pixel (m, n) is within the coverage range of a node d_i if its distance to (m_i, n_i) is not larger than the radius r_i . Since event X_i is independent to others, r_i and r_j are not related, $i, j \in [1, M]$ and $i \neq j$. Consequently, we can conclude the following expressions:

$$\begin{cases} P\{\bar{X}_i\} = 1 - P\{X_i\} = 1 - P_{cov}(m, n, d_i) \\ P\{X_i \cup X_j\} = 1 - P\{\bar{X}_i \cap \bar{X}_j\} = 1 - P\{\bar{X}_i\} \cdot P\{\bar{X}_j\} \end{cases} \quad (13)$$

where \bar{X}_i represents the complement of X_i , meaning that d_i fails to cover pixel (m, n) . Moreover, the pixel (m, n) is covered if any of nodes in the set covers it meaning that coverage hole is avoided. Conversely, if none of the node covers pixel (m, n) then a coverage hole is formed. So, the probability that pixel (m, n) is within the coverage range of node set can be written as the union of X_i

$$\begin{aligned} P_{cov}(m, n, D) &= P\left\{\bigcup_{i=1}^M X_i\right\} = 1 - P\left\{\bigcap_{i=1}^M \bar{X}_i\right\} \\ P_{cov}(m, n, D) &= 1 - \prod_{i=1}^M (1 - P_{cov}(m, n, d_i)) \end{aligned} \quad (14)$$

We finally express coverage rate of node set $P_{cov}(D)$ as ratio of coverage area (Z) to total network field area ($100\text{m} \times 100\text{m}$),

$$P_{cov}(D) = \sum_{x=1}^{100} \sum_{y=1}^{100} P_{cov}(m, n, Z) / (100 \times 100). \quad (15)$$

5. EXPERIMENTS, RESULTS AND DISCUSSIONS

In this section, we discuss the simulation outcomes of our scheme. These results are then compared with prevailing traditional schemes based on stability period, network lifetime, instability period, number of packets sent and received, and consumption of energy. We have randomly deployed $M = 100, 1000$ and 10000 sensor nodes in the field area of $100\text{m} \times 100\text{m}$. This network area is then partitioned uniformly in 9 regions. BS is located at center of the sensor network. Sensors in region 1 have initial energies of $0.7 \times (1 + \alpha)J$ where the value of α is 0.5, whereas initial energy of sensors in other regions is $0.7J$.

It should be noted that there is no specific reason of why α is set to 0.5. This value is just to make the nodes of region 1 advanced from other nodes in terms of initial energies. For realistic results, average values are plotted by simulating the network 5 times. Initialization of parameters for radio model is shown in Table 1. To make our scheme a more practical one, uniform random model (URM) [13] is used to compute packets drop where packet drop probability is set to 0.3. Simulations are carried out over various datasets to show its consistency. For 1D data, Gaussian distributed dataset has been used, while for 2D data, various standard test images have been used from image processing database.

Fig. 4 demonstrates packets sent to BS by the sensor nodes for $M = 100$ nodes. As shown in this figure, DDR [8] transmits more number of packets as compared to our scheme. This is because packets are transmitted after crossing threshold value along with data compression done using DPCA-CA approach. Therefore, this smaller number of transmitted packets results in less energy consumption and improves the network lifetime. As shown in Fig. 4, network death in DDR occurs around 3500th round while in our scheme network has a lifespan till 6500th round.

The packet drop ratio of DDR vs. our proposed protocol is depicted in Fig. 5. As shown, packets drop in proposed protocol is less in comparison with DDR. The reason of this smaller packet drop is the threshold sensitive nature. Moreover, the use of URM and number of packets received on BS is illustrated in Fig. 6. It clearly shows that packets received on BS in M-BEHZAD are smaller in comparison with DDR. This again is associated with the threshold aware transmissions of our proposed scheme.

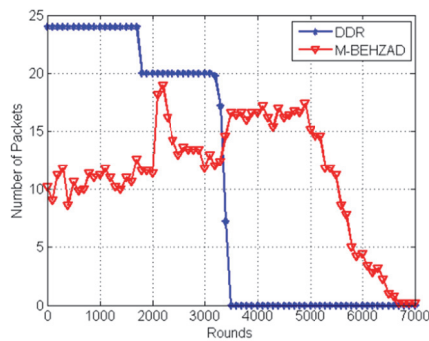


Fig. 4. Packets sent to BS.

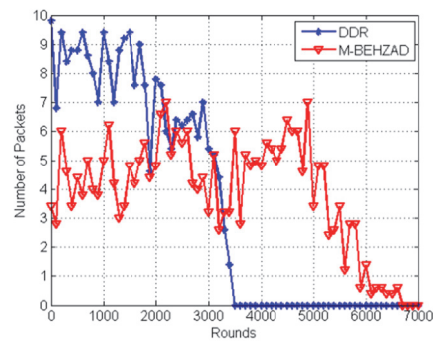


Fig. 5. Packets dropped comparison.

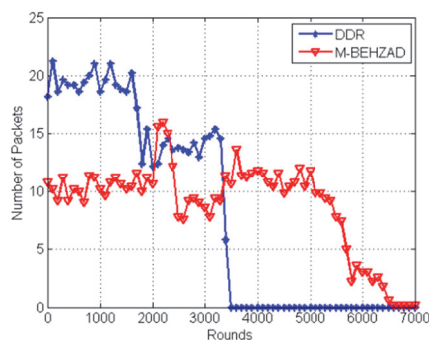


Fig. 6. Packets received on BS.

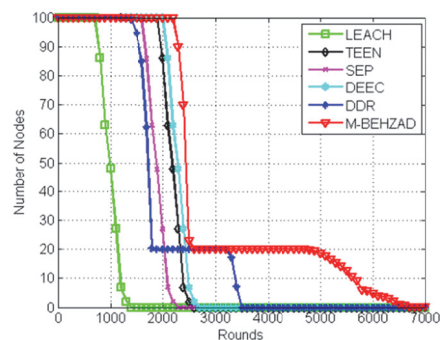


Fig. 7. Stability period - alive nodes.

The stability period comparison of our approach with other methods in literature is shown in Fig. 7. As can be seen from this figure, network in our approach lives 5100 rounds more than LEACH [1], 3000 rounds more than DDR [8], 5300 rounds more than TEEN [2], 2900 rounds more than TEECH-ME [17] and 1300 rounds more than THEEM [18]. Our proposed protocol has first node die time (FDT) of 2200 while in DDR and LEACH it is 1300 and 700 respectively. This shows that network in our scheme has a longer stability period. Moreover, our proposed protocol has all node die time (ADT) of 6500 while in DDR and LEACH it is 3500 and 1400 respectively. This longer stability period is because of threshold aware communication among nodes along with compression using DPCA-CA approach. Balanced energy consumption of sensors also helps to tackle the issue of energy and coverage holes. The comparison with other routing protocols shown in Fig. 7 clearly validates the efficiency of our proposed scheme.

Likewise, number of dead nodes in each round is shown in Fig. 8. It can be seen that first node die time of proposed scheme is 2200, whereas LEACH and DDR has FDT of 700 and 1300, respectively, which proves that our scheme is effective than traditional schemes in terms of network lifespan. Furthermore, we have also experimented the proposed method for 100, 1000 and 10,000 number of nodes and the network lifetime results have been summarized in Fig. 9. The figure clearly demonstrates the efficiency of our method over other methods for different number of nodes. It also validates the point that our algorithm uniformly outperforms other algorithms in various scenarios.

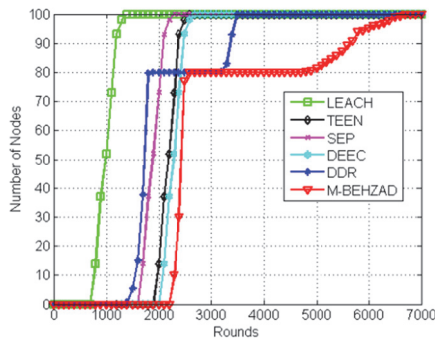


Fig. 8. Dead nodes.

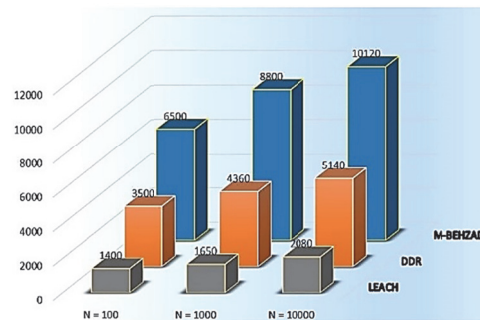


Fig. 9. Network lifetime.

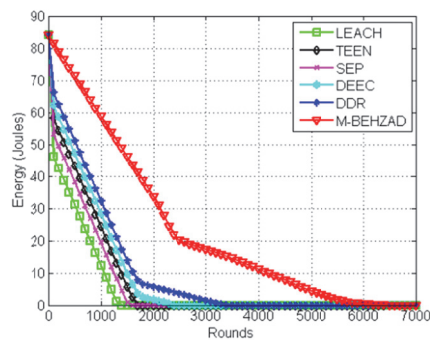


Fig. 10. Energy consumption.

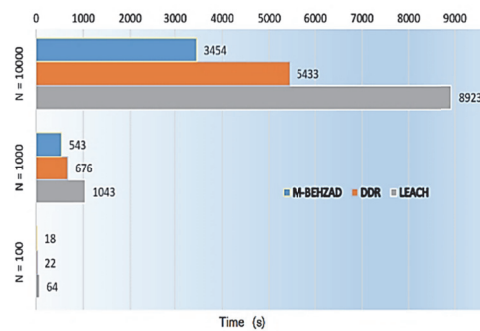


Fig. 11. Computational overhead comparison.

Energy consumption has been compared in Fig. 10, while Fig. 11 compares the computation overhead that the competing methods are utilizing for different number of nodes. It can be seen that our proposed protocol beats other schemes in terms of energy consumption as well as computational time (on a 2.20GHz Intel Core i7-3632QM machine). The lifetime comparison has also been summarized in Table 2. The efficient energy consumption of our scheme is associated with its robust design and efficient routing mechanism.

Table 2. Comparison of network lifetime with traditional algorithms.

Protocol	[1]	[8]	[2]	[17]	[13]	[18]	M-BEHZAD
Nature	Proactive	Proactive	Reactive	Reactive	Reactive	Reactive	Reactive
Lifetime	1400	3500	2500	3600	4400	5200	6500

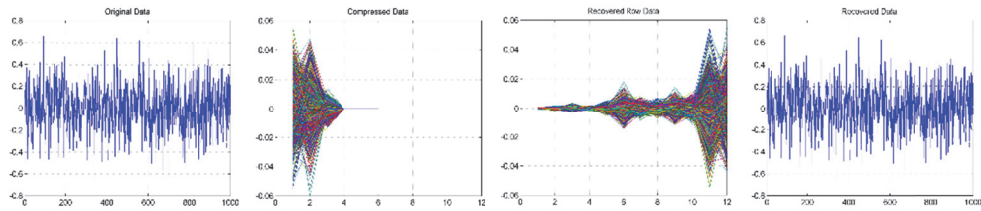


Fig. 12. Gaussian data compression and restoration.



Fig. 13. Compression and restoration of standard images (*Cameraman*, *Lena*, *Baboon* and *Field*).

M-BEHZAD is applied for 1D Gaussian and 2D/image data compression and recovery as shown in Figs. 12 and 13, respectively. As can be seen from these figures, the data is compressed and recovered successfully with minimum amount of energy consumption. Consequently, the proposed protocol has also proved itself as a better algorithm for image and data compression applications. The proposed algorithm has also been applied on both grey scale and colored images to validate its applicability. With increased number of consensus rounds, the convergence is guaranteed and usually the convergence is achieved in 3-4 rounds. The original images were compressed using various number of principle components k as shown in Fig. 13. It can be seen that increasing k increases performance gain. However, it is worth noting that at 25% of PC used, the image is recovered completely and this is the reason why PCA is widely adopted for image compression applications.

6. CONCLUSIONS

We propose M-BEHZAD to reduce frequent transmissions using threshold based communication architecture along implementation of DPCA-CA. The nodes in our scheme transmit reactively thereby reducing the energy consumption. Uniform random distribution with fixed clustering is utilized to extend the stability period. The criteria for selection of CHs is maximum residual energy. Another significant contribution is the implementation of 3-Tier design for avoiding energy holes. Results from comprehensive simulations using MATLAB are given and compared with traditional schemes. The results prove that M-BEHZAD outperforms other routing schemes in terms of energy utilization, network lifespan and stability period. We have also shown that our protocol is equally effective for grey scale and colored image compression and restoration purposes.

REFERENCES

1. W. Heinzelman *et al.*, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Vol. 8, 2000, pp. 8020.
2. A. Manjeshwar and D. P. Agrawal, "TEEN: A routing protocol for enhanced efficiency in wireless sensor networks," in *Proceedings of the 15th International on Parallel and Distributed Processing Symposium*, 2000, p. 189.
3. G. Smaragdakis *et al.*, "SEP: A stable election protocol for clustered heterogeneous wireless sensor networks," in *Proceedings of the 2nd International Workshop on Sensor and Actor Network Protocols and Applications*, 2004.
4. L. Qing, Q. Zhu, and M. Wang, "Design of a distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks," *Computer Communications*, Vol. 29, 2006, pp. 2230-2237.
5. M. Valcarcel, *et al.*, "Consensus-based distributed principal component analysis in wireless sensor networks," in *Proceedings of the 11th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, 2010, pp. 1-5.
6. A. Haider, *et al.*, "REECH-ME: Regional energy efficient cluster heads based on maximum energy routing protocol for WSNs," in *Proceedings of the 8th IEEE In-*

- ternational Conference on Broadband, Wireless Computing, Communication and Applications*, 2013, pp. 88-92.
7. N. Amjad, *et al.*, "DREEM-ME: Distributed regional energy efficient multi-hop routing protocol based on maximum energy in WSNs," in *Proceedings of the 8th IEEE International Conference on Broadband and Wireless Computing, Communication and Applications*, 2013, pp. 43-48.
 8. A. Ahmad, *et al.*, "Density controlled divide-and-rule scheme for energy efficient routing in wireless sensor networks," in *Proceedings of the 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering*, 2013, pp. 1-4.
 9. S. Lindsey and C. S. Raghavenda, "PEGASIS: power efficient gathering in sensor information systems," in *Proceedings of IEEE Aerospace Conference*, Vol. 3, 2002.
 10. O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, Vol. 3, 2004, pp. 366-379.
 11. P. K. Sahoo and W. C. Liao, "HORA: A distributed coverage hole repair algorithm for wireless sensor networks," *Transactions on Mobile Computing*, Vol. 14, 2014, pp. 1397-1410.
 12. Y. Gu, *et al.*, "ESWC: Efficient scheduling for the mobile sink in wireless sensor networks with delay constraint," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 24, 2013, pp. 1310-1320.
 13. M. Akbar, *et al.*, "Towards network lifetime maximization: Sink mobility aware multihop scalable hybrid energy efficient protocols for terrestrial WSNs," *International Journal of Distributed Sensor Networks*, 2015, Article No. 10.
 14. J. Jia, *et al.*, "Exploiting sensor redistribution for eliminating the energy hole problem in mobile sensor networks," *EURASIP Journal on Wireless Communications and Networking*, Vol. 68, 2012.
 15. W. Heinzelman, "Application-specific protocol architectures for wireless networks," Ph.D. Thesis, Massachusetts Institute of Technology, 2000.
 16. S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proceedings of IEEE International Conference on Computer Communications*, 2003.
 17. I. Khan, *et al.*, "TEECH-ME: Threshold sensitive energy efficient cluster heads based on maximum energy routing protocol for WSNs," in *Proceedings of the 9th IEEE International Conference on Broadband and Wireless Computing, Communication and Applications*, 2014, pp. 485-490.
 18. W. Raza, *et al.*, "THEEM: Threshold-sensitive energy efficient multi-hop routing protocol for WSNs," in *Proceedings of the 9th IEEE International Conference on Broadband and Wireless Computing, Communication and Applications*, 2014, pp. 88-92.
 19. M. Behzad and Y. Ge, "Performance optimization in wireless sensor networks: A novel collaborative compressed sensing approach," in *Proceedings of the 31st IEEE International Conference on Advanced Information Networking and Applications*, 2017, pp. 749-756.
 20. M. Behzad, *et al.*, "TSDDR: Threshold sensitive density controlled divide and rule routing protocol for wireless sensor networks," in *Proceedings of the 9th IEEE International Conference on Broadband and Wireless Computing, Communication and Applications*, 2014, pp. 78-83.



Muzammil Behzad received his B.S. degree in Electrical Engineering, with distinction, from COMSATS Institute of Information Technology (CIIT), Pakistan, and his M.S. degree in Electrical Engineering from King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. Presently, he is serving as Research Associate in CIIT, Islamabad, Pakistan. His research interests are oriented around the field of signal and image processing, wireless and digital communications, wireless sensor networks, and their applications.



Muhammad Sharjeel Javaid received his B.S. degree in Electronic Engineering from Ghulam Ishaq Khan University of Engineering Sciences and Technology (GIKI), Pakistan, and his M.S. degree in Electrical Engineering from King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He is currently a Lecturer in University of Hafr al Batin, Saudi Arabia. His current research interests include energy markets, communication protocols in power system and wireless power transfer.



Muhammad Ashraf Paracha received his B.S degree in Computer Science from Kohat University of Science and Technology (KUST), Pakistan, and his M.S. degree in Computer Engineering from Centre of Advance Studies in Engineering (CASE), Pakistan. He is currently Manager Software Development in Pakistan Telecommunication Company Ltd. (PTCL) Headquarters, Islamabad, Pakistan. His current research interests include information security, wireless sensor networks and transitional software behavior.



Salman Khan received his B.S. degree in Electrical Engineering from COMSATS Institute of Information Technology (CIIT) Islamabad, Pakistan. Previously, he has worked as Assistant Business Manager in Pakistan Telecommunication Company Ltd. (PTCL), Sialkot, Pakistan. Currently, he is working as a Security Consultant for low current systems in collaboration with Saudi Telecom Company (STC) and Al Rajhi Bank in Saudi Arabia. His research interests include analyzing communication protocols in sensor networks.