Wireless Personal Communications (2019) 109:1311–1331
https://doi.org/10.1007/s11277-019-06614-5

Clustering and Compressive Data Gathering in Wireless Sensor Network

Utkarsha S. Pacharaney[1] · **Rajiv Kumar Gupta**[2]

Published online: 18 May 2019

Abstract

In wireless sensor network (WSN) redundant data gathering and transmission occurs due to dense deployment. Recently compressive sensing (CS) has attracted considerable attention for efficient data gathering in WSN. CS can reduce data transmission but the total number of transmissions for data collection is high. To alleviate this, hybrid of CS and raw data collection is proposed and integrated with clustering. Clusters used in this integration reduce the number of CS transmissions, but do not reduce the number of transmissions. In a cluster amount of transmission depends on the number of transmitting nodes and their location in data gathering, hence the way in which nodes are clustered together can significantly effect on the number of transmissions in cluster and overall transmissions in network. When density of sensor nodes in a network is high, we can take advantage of their inherent spatial correlation to reduce the number of transmissions. Motivated by this, we propose a novel base station (BS) assisted cluster, spatially correlated, to reduce the number of transmission in a CS-based clustered WSN. Different from other spatially correlated clusters, in this cluster only CH senses, gathers data in the correlated region, and then transmits compressively sensed measurements to BS without incurring any intra-cluster communication cost. In addition, the clusters so formed, convert a randomly deployed network into a structured one i.e. when several clusters are grouped together they form a hexagonal topology (proved to have a high success rate in cellular network). The proposed system makes WSN transmission efficient by reducing number of transmissions in the network and number of data transmission at the CH using clustering and CS. Also energy consumption is reduced and network lifetime is prolonged.

Keywords Wireless sensor networks · Cluster · Compressive sensing · Correlation

✉ Utkarsha S. Pacharaney
utk21pac76@gmail.com

Rajiv Kumar Gupta
rajivmind@gmail.com

[1] Department of Electronics Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India

[2] Department of Electronics and Telecommunication Engineering, Terna Engineering College, Nerul, India

1 Introduction

Wireless sensor network (WSN) is an assembly of densely scattered sensor nodes, whose primary objective is to gather data for which they are deployed in an ad-hoc fashion. Due to dense deployment, redundant data gathering and transmission take place in WSN. Data transmission contributes majority of energy consumption in a battery-operated network [1], which is traditionally reduced using compression. Existing approaches adopt in-network data compression such as entropy coding, principal component analysis or transform coding [2–4]. However, these include significant computations and control overheads that are often not suitable for sensor networks. Compressive sensing (CS) is a novel paradigm that senses the data in its compressed form and is used effectively for gathering data in a WSN. By leveraging ideas from CS, we can enable more economic use of sensing resource in an energy-constrained network.

A brief idea of CS is as follows: Consider a signal $f = \left[f_1, f_2, \ldots\ldots\ldots\ldots f_N\right]$ denoting a set of sensor readings from N nodes. In sensor networks without CS technique, the network needs to acquire all N samples of the signal f leading to huge amount of data traffic. If data f is k -sparse then we can obtain $Y = \left[Y_1, Y_2, \ldots\ldots\ldots\ldots Y_M\right]$, where $M \ll N$ but includes most of the information of f, by multiplying it with compression matrix $\emptyset$. Also f can be recovered from Y by solving l_1 optimization problem i.e. $\min_{z \in \mathbb{R}^n} ||Z||_{l_1}$ subject to $Y = \emptyset \Psi Z$ where Ψ represents a proper basis such that $f = \Psi Z$. While powerful, the CS theory is mainly designed to exploit intra-signal structure at a single node by collecting fewer measurements than the original sensor data [5] i.e. each node transmits M weighted samples for a set of N data items, where $M \geq ck \log\left(\frac{N}{k}\right)$. If we apply plain CS, every node will be forced to generate M samples leading to higher data transmissions in the early stage of network, which are unnecessary. Therefore, appropriate way of applying CS is to start CS coding when the outgoing samples are not lesser than M, otherwise raw data transmission is used. This hybrid CS framework induces two types of traffic into the network i.e. encoded and raw. Hybrid CS is further integrated with tree and cluster- based hierarchical method of data collection. Cluster-based data collection methods have more advantages [6] than non-cluster-based methods; hence CS is integrated with clustering. But these clusters are randomly formed with all the nodes sending data to the Cluster head (CH) which further aggregates it and transmits to BS by multi hopping it through other CHs. Thus, we have intra-cluster communication between the CH and member nodes and inter-cluster communication between CHs of different clusters.

Let WSN consist of N sensor nodes with C clusters. Let ith cluster be formed by S_i number of sensor nodes and

$$\sum_{i=1}^{C} S_i = N \tag{1}$$

In each cluster one nodes is CH and $(S_i - 1)$ nodes are member nodes. These member nodes transmit their data to CH in the transmission phase and then CH performs CS [7]. Therefore, number of transmissions in the network is

$$\sum_{i=1}^{C} S_i - 1 + \sum_{i=1}^{C} M_i \tag{2}$$

First term corresponds to the number of transmissions of $(S_i - 1)$ nodes to CH in all C clusters and the second term corresponds to CS performed by CH.

If each cluster contains the same number of nodes S and each CH perform CS to M number of measurements then the communication load of the whole network is

$$C * (S - 1) + C * M = N + C(M - 1) \tag{3}$$

To reduce the communication load of the network, we propose a cluster with number of transmissions equal to $C * M$, which are very less. Existing methods with CS reduce number of CS transmissions, but do not reduce the overall number of transmissions in the network. The total amount of transmissions in a cluster depends on the number of transmitting nodes available and their location in data gathering. The ways in which nodes cluster together, can affect the number of transmissions in a clustered network significantly.

When the density of sensor nodes is high, clustering framework can be used to exploit spatial correlation in a dense network. The clusters used in the integration of CS and clustering neglect the inherent spatial correlation in a densely deployed WSN. Hence, we propose a base station (BS) assisted cluster, which is spatially correlated. Different from other spatially correlated clusters, this cluster only senses using CH, gathers data in the correlated region, and then transmits the compressively sensed measurements to the BS without incurring any intra-cluster communication cost. When several clusters are grouped together, they form a hexagonal topology (which is proven to have high success rate in cellular network) hence converting randomly deployed network into a structured one. The proposed system makes WSN transmission efficient by reducing the number of transmissions in the network and number of data transmissions at the CH by using clustering and CS, hence reducing energy consumption and prolonging network lifetime.

The main contribution of this paper is to

1. Present a transmission-efficient data gathering scheme with the integration of CS and a spatially correlated cluster.
2. A cluster with radius equal to the sensing range of the CH is constructed considering spatial correlation of sensory information between sensor nodes geographically close to each other.
3. Unlike the conventional clustering approach used with CS, in which CH collects data from its members and then transmits it to the BS, we save on the intra-cluster communication cost with only the CH sensing and gathering data in the correlated region and then transmitting to the BS.
4. Existing methods stochastically elect CH and form a cluster around it. Our CH selection is at an appropriate location and hence, the correlated cluster formed by this method when grouped together, forms a hexagonal topology which is widely accepted in wireless networks.

The organization of this paper as follows: Section II reviews the related work. Section 3 presents a correlation model, which gives the relation between correlation and sensing range of the sensors. The proposed spatially correlated cluster is described in Sect. 4. While a detailed energy consumption analysis of cluster formation and data transmission is carried out in Sect. 5. We discuss results in Sect. 6. Finally, we conclude this work in Sect. 7.

2 Related Work

In general, for minimizing expensive data transmission and increasing the lifetime of WSN, data compression is one of the best techniques. With this aim, Duarte et al. [8] have overviewed and detailed an array of proposed compression methods. In particular, CS is a new approach to simultaneously sense and compress that promises to reduce the sampling and computational cost [9]. It is used very economically in an energy constrained sensor network. The first practical implementation of CS called Compressive Data Gathering (CDG) was carried out by Luo [10], which used a tree-based aggregation. Each node transmits M data packets for a set of N data items and the total number of transmissions for collecting data from N nodes is MN, which is significantly high for a large-scale network. Thus, observation is that CS can reduce the number of data transmissions but the overall number of transmissions for data collection is high. To alleviate this problem, a hybrid version of CS and raw data collection was proposed by Luo [11] and applied by Xiang [12] in a tree based aggregation. Since clustering, has many advantages over tree based method, integration of clustering and CS is proposed in literature. In [7], Xie et al. combined hybrid CS with clustering, that is, inside the cluster data gathering is carried out without CS and between CHs data gathering is carried out with CS, and the optimal size of cluster is found analytically that lead to minimum number of transmissions. Minh et al. [13] also calculated the optimal number of clusters and proved that consumed power is a decreasing function of number of clusters i.e. more clusters results in more power saving. Dan et al. [14] found that in hybrid CS, each CH performs M transmissions for the intracluster data gathering and the number of transmission increases in a large-scale network. He proposed a new hybrid CS method with hexagonal cluster structure where within a cluster, sensor nodes transmit collected data directly to the CHs through single hop without CS then CHs transmit data to neighboring CHs without CS if the number of packets in each CH is smaller than the measurement M. If packet are larger than M, CS is used for data gathering. In [15], Lan et al. proposed a Compressibility Based Cluster Algorithm (CBCA) that enables less data transmissions than the random clustering method used in the existing integration of CS and clustering. In CBCA, the network topology is converted first into a logical chain similar to the concept used in PEGASIS and then spatial correlation of the cluster nodes readings is employed for CS. To keep cost of transmission low for each measurement, [16] has obtained measurements from clusters, sourced from adjacent sensors, using spatially localized sparse projections. Aiming to take advantage of spatial data correlation in a dense WSN, cluster structure can be reorganized with nodes to exploit the spatial correlation in network to reduce number of data transmissions and energy consumption. Various clustering algorithms have been proposed in the survey [17], but very little work is present to group sensor nodes with similar readings in the same cluster. Vuran et al. [18] have pointed out that in a dense WSN, significant energy saving is possible by allowing less number of nodes to transmit information instead of redundant ones. A dynamic cluster is formed when nodes need to access transmission media with radius of cluster equal to r_{corr}. The CHs, known as representative nodes, are chosen based on distortion constraint on reconstructed signal. Also, the imbalance in the density of nodes is studied through clustering and under-sampling in [19]. In literature, the judgment of spatial correlation is based on geographic distance of sensors [20], tolerance error of different sensor reading [21] or area of overlap between sensors [22], and the combination of error tolerance range and spatial correlation range [23] to form correlated cluster. Selection of the CH is very crucial in these clusters since they represent the data feature of the entire group [24]. Apart from ignoring the cost of learning the correlation, drawbacks of these correlated clusters are that there is no uniform measure on tolerance error range or distance between sensors, abundant communication overhead for spatial clustering, more number of iterations to select the CH and above all, the construction of cluster takes several rounds of message exchange and computation of correlations in these clusters.

From all the reported work, it is clear that existing CS based data gathering integrated with clustering reduces the cost of CS transmissions, but number of transmissions in the cluster and hence in the network are not reduced. The clusters formed in this integration neglect the inherent spatial correlation in densely deployed WSN. We can exploit spatial correlation in cluster while reducing number of transmissions. Also spatially correlated clusters in literature are elastic; i.e. they have no specific or strict requirement on similarity measured between nodes and distance between them. In the light of these drawbacks, we propose a novel clustering procedure taking into consideration, the spatial correlation of sensory information between nodes located geographically in the sensing range of the CH, and hence form a cluster with radius equal to the sensing range of the CH. These correlated clusters have same type of information so only CH senses, gathers data and sends compressively sensed information to BS. This proposed scheme significantly reduces overall number of transmissions in the network while prolonging network lifetime.

3 Correlation Based on Sensing Range of Node

The proposed correlated clusters are based on spatial location of sensor nodes hence we present data correlation between them using correlation model. The correlation between sensory data of nodes related to spatial correlation between them is estimated based on sensory coverage of nodes. In addition, their reading association character can describe the correlation of different sensor nodes, which is covariance. The covariance between two measured values from node n_i and n_j at location R_i and R_j respectively can be expressed as [25]

$$Cov\{R_i, R_j\} = \sigma_s^2 K_\vartheta(d) \tag{4}$$

where σ_s^2 = variance of sample observation from sensor nodes, $Cov(.)$ = mathematical covariance, $K_\vartheta(.)$ = denotes correlation function.

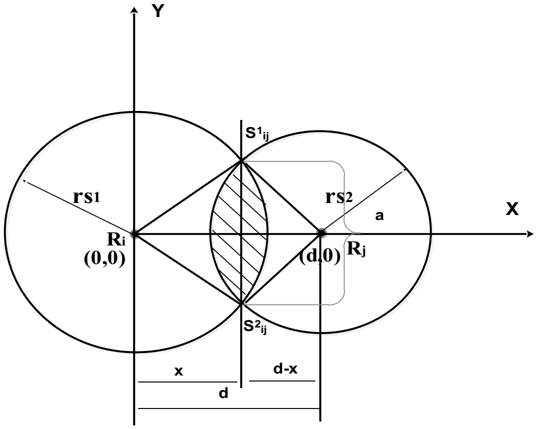
3.1 Correlation Model

Assume two sensor nodes with sensing range rs_1 and rs_2 located d distance apart as shown in Fig. 1.

Using geometry to set up the correlation model with meaning of symbols explained as follows:

R_i and R_j denote location of node n_i and n_j of disk with radius rs_1 and rs_2 respectively.
A_i Area R_i denoting area of R_i
A_j Area R_j denoting area of R_j
R_i^j Region delimitate by the perpendicular bisector of R_i and R_j and belongs to R_i
A_i^j Area R_i^j denoting the area of R_i^j
A_j^i Area R_j^i denoting the area of R_j^i
R denotes the sensing region

Fig. 1 Correlation model



A denotes the area of R
d distance between R_i and R_j
S^1_{ij} and S^2_{ij} are intersection points of the two nodes
a length of the common chord joining S^1_{ij} and S^2_{ij}

$$a = \frac{1}{d}\sqrt{4d^2 rs_1^2 - \left(d^2 - rs_2^2 + rs_1^2\right)} \tag{5}$$

x is the distance between R_i and chord a

$$x = \frac{d^2 - rs_2^2 + rs_1^2}{2d} \tag{6}$$

If $d < rs_1 + rs_2$ then R_i overlaps with R_j and correlation is defined as

$$K_\vartheta(d) = \frac{A_i^j + A_j^i}{A} \tag{7}$$

$A_i^j + A_j^i = A^{int}$ is the area of the asymmetry lens which intersect the sensing range of two sensor node and is calculated by using the formula of circle segment of radius R' and triangle height d'

$$A^{int}\left(R', d'\right) = R'^2 \arccos\left(\frac{d'}{R'}\right) - d'\sqrt{R'^2 - d'^2} \tag{8}$$

Hence

$$A^{int} = A\left(rs_1, x\right) + A\left(rs_2, d - x\right) \tag{9}$$

$$A^{int} = rs_1^2 \arccos\left(\frac{d^2 + rs_1^2 - rs_2^2}{2drs_2}\right) rs_2^2 \arccos\left(\frac{d^2 + rs_1^2 - rs_2^2}{2drs_1}\right) - \frac{1}{2}\sqrt{4d^2 rs_1 - \left(d^2 - rs_2 + rs_1\right)^2} \tag{10}$$

Hence, from (7) we obtain

$$K_{\vartheta}(d) = \frac{A^{int}}{A} \tag{11}$$

For the case when the sensing range of the two nodes is same i.e. $rs_1 = rs_2 = r$

$$K_{\vartheta}(d) = \frac{2}{\pi} \arccos\left(\frac{d}{2r}\right) - \frac{d}{\pi r^2} \tag{12}$$

It is observe from (12) that the correlation is a function of distance and the sensing range of the sensors. That is two nodes lying within the sensing range of each other are correlated. In addition for any value, $1 < d < rs_1 + rs_2$, correlation exists between nodes with correlation being zero for $d = rs_1 + rs_2$ while for $d = 0$ correlation is maximum. Taking this into consideration we construct a correlated cluster with the radius of the cluster equal to the sensing range of the CH as discussed in the next section

4 Spatially Correlated Cluster

The assumptions made of the proposed cluster are as follows:

i. Sensor nodes know the geographic location via attached GPS or other localization techniques.
ii. The sensing range r_s is half the transmission range r_t i.e. $r_t = 2r_s$ [26].
iii. Sensed information is highly correlated.
iv. The nodes are capable to adjust their transmission power to the desired recipient within distance r_t.

Clustering involves selection of cluster head, cluster formation and data transmission. In this work, clusters are formed only once in the network that is membership of the nodes is fixed in a cluster while the role of CH changes in each round. A round is defined as the time when the energy of CH reaches a pre-defined threshold (E_{th}).

4.1 Selection of Cluster Head

In our proposed cluster, CH represents information for the entire cluster hence its selection is very crucial for spatially correlated clusters. Using 'leader first' protocol, where CH is selected first and then a cluster is formed around it, selection procedure of CH is described in this section. After deployment of sensor nodes, they inform their location and residual energy to the BS. Considering BS assisted cluster formation, BS selects CH with highest residual energy and most appropriate location [27]. Cluster head selection process is described mathematically by a parameter W_{CH} and expressed as:

$$W_{CH} = w_1 R_{CH}^{AL} + w_2 R_{CH}^{E} \tag{13}$$

where R_{CH}^{AL} is the location and R_{CH}^{E} is the energy factor of CH while w_1 *and* w_2 indicate the contribution of both the parameters in the expression of W_{CH}.In order to obtain W_{CH}, R_{CH}^{AL} and R_{CH}^{E} are to be computed. R_{CH}^{AL} is evaluated at the intersection of distance $2r_s$ and 60^0 angle line as illustrated in Fig. 2.

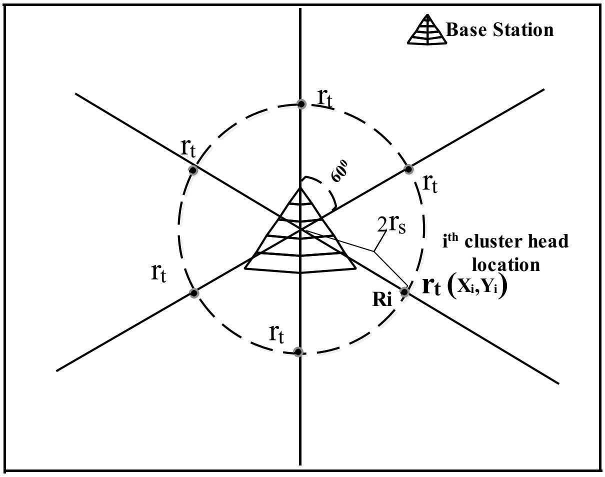


Fig. 2 Location of cluster head

Let the location be denoted by coordinate (x_i, y_i) at R_i. R^E_{CH} is measured by taking ratio of average residual energy of CH and average residual energy of non-CHs. Thus, BS calculates W_{CH} and geocasts the tuple (R^{AL}_{CH}, R^E_{CH}). All nodes receive this tuple but the one with closest value announced in the message is elected as CH and forms a cluster around. These selected CHs can expand the network by selecting CHs at twice their sensing range thus achieving scalability. Cluster formation around the BS, explained in the next section.

4.2 Cluster Formation

Considering, spatial correlation of sensory information between sensor nodes geographically close to each other, we construct a cluster with radius equal to the sensing range of the CH. Hence, after the CHs election they announce their availability as a CH within the area of its sensing radius r_s by transmitting announcement packets to form a cluster. Nodes associate to CHs via sending packet, which includes its ID, location information and residual energy. These nodes remain permanent members of the cluster and will become CH once in their lifetime. Figure 3 illustrates cluster formation according to the disk-sensing model [28]. The average number of nodes in a correlated cluster will depend on the sensing range r_s and node density in the area.

Lemma 1.1 *Choosing the CH at the intersection of distance $2r_s$ and 60° angle lines by the BS and forming a cluster with radius equal to r_s, when a triangle is circumscribed by this cluster an equilateral triangle is obtained and grouping six neighboring equilateral triangle together forms a hexagon. (Proof is given in Appendix)*

Although the deployment of sensor nodes is random in the network the clusters so formed result in a structured hexagonal topology as shown in Fig. 4. Hexagonal topology is already a success in cellular network providing maximum coverage area [29] and has the advantage of providing better connectivity. CHs are connect with each other as they are at distance of twice the sensing range i.e. transmission range, from each other. Only CHs

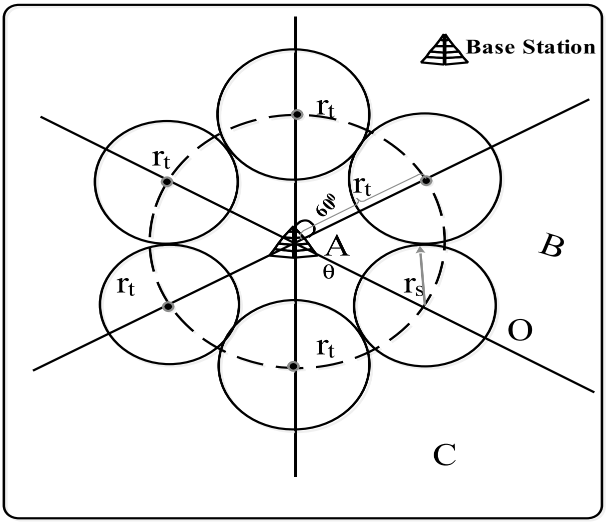


Fig. 3 Formation of cluster

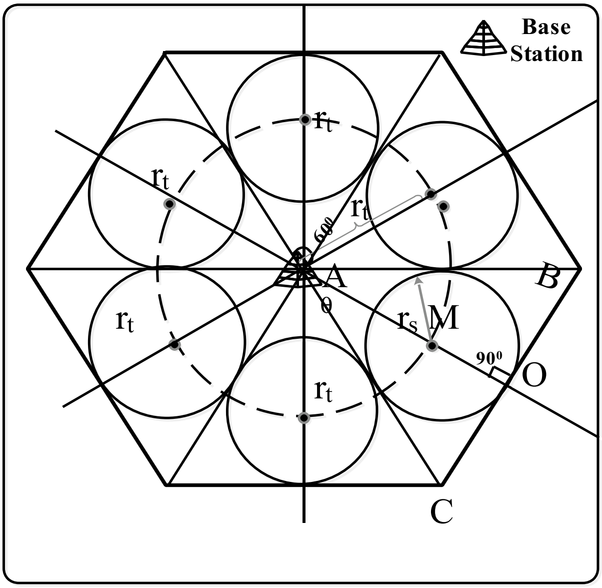


Fig. 4 Formation of hexagon

transmit data to the BS and other nodes sleep, the sleeping schedule of nodes is explained in the next section.

4.3 Sleep Scheduling

When CH becomes the coordinator of cluster, it sends a beacon frame to schedule sleep mode of cluster members [30]. All sensor nodes in the cluster are asleep and only CH is awake. It senses, compresses data and sends it to BS. Energy is thus conserved by data compression, efficient control of transmission power and efficient sleep scheduling of the cluster nodes. In the payload field of the beacon the sleep schedule is distributed by the CH, which is not same for all the nodes, it depends on the node that become the CH for the next round and subsequent rounds. The Cluster head determines the next CH of the cluster based on distance between them.

4.4 Data Transmission

After clustering, CS-based data gathering is carried out by CH and CS measurement are transmitted to the BS.

Let CH_i denotes the CH of the cluster i, and f_i represents the sensor reading at CH_i. CH_i has N_i readings which can be denoted as.

$$f_i = \left[f_1, f_2, \dots \dots \dots \dots f_{N_i}\right] \tag{14}$$

The CH multiplies it by a random matrix $\emptyset_i$ and then sends the product Y_i to the BS.The BS collects each measurement from one cluster at a time, and a Block Diagonal Matrix (BDM) as a sensing matrix is build.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix}_{M\times 1} = \begin{bmatrix} \emptyset_1 & 0 & 0\,0 \\ 0 & \emptyset_2 & 0\,0 \\ 0 & 0 & \ddots\ 0 \\ 0\,0 & 0 & \emptyset_M \end{bmatrix}_{M\times N} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N_i} \end{bmatrix}_{N\times 1} \tag{15}$$

This BDM with only one nonzero entry in each row and column is the sparsest measurement matrix.

Finally, the BS receives, $Y = \bigcup_{i=1}^{C} Y_i$, compressed information of all clusters at the BS. The original data can be reconstructed from Y by using l_1 minimization

5 Energy Consumption Analysis

The objective of this work is to prolong the lifetime of the WSN network, consisting of battery operated sensor nodes. In this section, we compute the energy consumed in the formation of the cluster and during the data transmission. We first begin with the energy consumption model used in this work. Figure 5 describes radio energy dissipation model used in wireless sensor network [31].

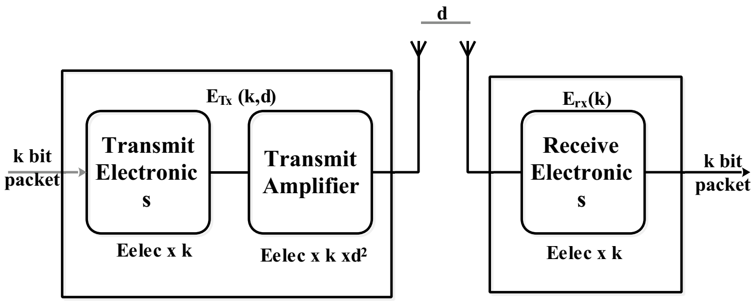


Fig. 5 Radio energy dissipation model

To transmit k bit at a distance d, the energy expended is given by

$$E_{Tx}(k,d) = \begin{cases} \{E_{elec}k + \mathcal{E}_{fs}kd^2 & d \le d_0 \\ \{E_{elec}k + \mathcal{E}_{mp}kd^4 & d > d_0 \end{cases} \tag{16}$$

where E_{elec} is the electronics energy and $\mathcal{E}_{fs}d^2$ or $\mathcal{E}_{mp}d^4$ are amplifier energy, which depends on the distance from the receiver.

To receive k bits the energy expended by the receiver is:

$$E_{Rx}(k) = E_{elec}k \tag{17}$$

$$E_{Total}(k) = E_{TX}(k,d) + E_{Rx}(k) \tag{18}$$

5.1 Energy Consumption in Cluster Formation

The energy consumption for first six clusters formed around the BS is as below:

Let the number of sensor nodes be N.

After deployment of nodes, they transmits their location to the BS using l_0 control packets. The energy transmitted is given by

$$E_{TX} = \left(l_0 E_{elec}k + l_0 \mathcal{E}_{fs}kd^2\right)N \tag{19}$$

After receiving this information, the BS calculates the CH location in all six directions and geocast the information. All nodes receive this and energy consumed by nodes is,

$$E_{RX} = Nk \tag{20}$$

Initially CH candidadtes are only six nodes. The region where CH candidacy is expected reduces to r_s around the location mentioned in the message after the geocast by BS. Let the nodes in these regions be P_j where $\bigcup_{j=1}^{6} P_j = N$.

The calculations are carried out with P_1 nodes for a cluster C_1. The candidate nodes compete with each other to become CH. While competing, each of the P_1 nodes transmits data while $\left(P_1 - 1\right)$ nodes will receive data by other nodes in the region r_s.

The total energy consumed is

$$E_{CH-cand} = \pi r_s^2 P_1 l_0 \left(E_{elec}k + k\varepsilon_{fs} r_s^2\right) + \left(\pi r_s^2 P_1 - 1\right) l_0 E_{Rx}(k) \tag{21}$$

The node with the highest threshold given in Eq. 13 is elected as CH. Upon being selected, each CH announces its role with a CH announcement packet that is received only by the nodes inside its sensing range r_s.

$$E_{annou} = E_{TX}\left(k, r_s^2\right) + \left(\pi r_s^2 P_1 - 1\right) E_{Rx}(k) \tag{22}$$

Each member node needs to send a control packet l_0 to associate with a CH that then replies back with an association message which is given by

$$E_{assoc} = E_{TX}\left(k, r_s^2\right)\left(\pi r_s^2 P_1 - 1\right) l_0 + \left(\pi r_s^2 P_1 - 1\right) l_0 E_{Rx}(k)) \tag{23}$$

Finally, the CH sends beacon and distributes sleep schedule to each member node. The energy spent in transmitting beacons is given as:

Let the average rate of control packet generation be *avgoh* and the transmission time of a control packet *ttpkt*, P_{tx} is the power consumption in transmitting one packet. Thus the amount of energy spent due to overhead in time interval [0,t] can be expressed as [32]

$$E_{overhead}(t) = t\left(avgoh.ttpkt\left(P_{tx} + P_{rx}\right)\right) \tag{24}$$

Hence, total energy consumed in the cluster formation is:

$$E_{Energy-cluster-formation} = E_{TX} + E_{RX} + E_{CH-cand} + 6 * E_{annou} + 6 * E_{assoc} + 6 * E_{overhead}(t) \tag{25}$$

5.2 Energy Consumption in Data Communication

In clustering, the communication cost incurred during data transmission includes the communication cost by each member of the cluster to send data to the CH i.e. E_{intra} and the communication cost incurred by the CH while sending data to the BS i.e. E_{inter}.

In this work only CH are the main contributors to energy consumption and we have no E_{intra}, thus we only analyze the energy consumption of the CH.

The total energy consumption of the CH comprises: sensing energy cost E_{sm}, data processing energy consumption E_{DP}, Data transmission energy consumption E_{TR-hop}.

Therefore, energy consumption at the CH is

$$E_{CH} = E_{sm} + E_{DP-CS} + E_{TR-hop} \tag{26}$$

We only consider one cluster in the analysis. According to [33]

$$E_{sm} = V_{dc} I_i T_i \tag{27}$$

In general, sensor *i,* has sensing energy consumption as given in Eq. 27, where I_i is the current drawn by the sensor *i* and T_i is the time required for obtaining a single sample from sensor *i.*

The number of operations for signal processing determines the energy consumption of CPU. In other words, energy consumption of the data processing scales with the operation during the process of signal processing. At the CH, we acquire *M* measurements through compressed projections on CH. So the energy consumption of CH for data processing includes data compression cost (E_{DP-CS}) including that of data reading and writing.

The *M* random measurements are acquired through a linear compressed projection on *S* sensory data. The linear compressed projection is a matrix multiplication operation that is the process of multiplying the $M \times S$ measurement matrix by an S-dimensional data vector to get an *M*-dimensional vector. The additions and multiplications executed are $M \times (S-1)$ and $M \times S$ respectively. So the total energy consumption of CH for data processing can be expressed as a sum:

$$E_{DP-CS} = S\xi_{mrd} + MS\xi_{mul} + M(S-1)\xi_{add} + M\xi_{mwr} \tag{28}$$

where $\xi_{add} and \xi_{mul}$ are add and multiply operations and $\xi_{mrd} and \xi_{mwr}$ are memory read and write operation in the CPU of the sensor node [34].

On the CH, we apply CS and send *M* measurements to the sink. The energy consumed by the CH in sending measurements to the BS via *H* multihop is

$$E_{TR-hop} = \left(E_{elec} k + \mathcal{E}_{fs} k d^2\right) MH \tag{29}$$

6 Results and Performance Evaluation

In this section, we evaluate the performance of the proposed system, which makes the WSN transmission efficient by reducing the number of data transmissions at CH by applying CS and reducing number of transmissions in the network by our proposed spatially correlated cluster.

Firstly, we evaluate performance of the proposed clustering method with and without applying CS. Three metrics are used to evaluate the performance of both the methods:

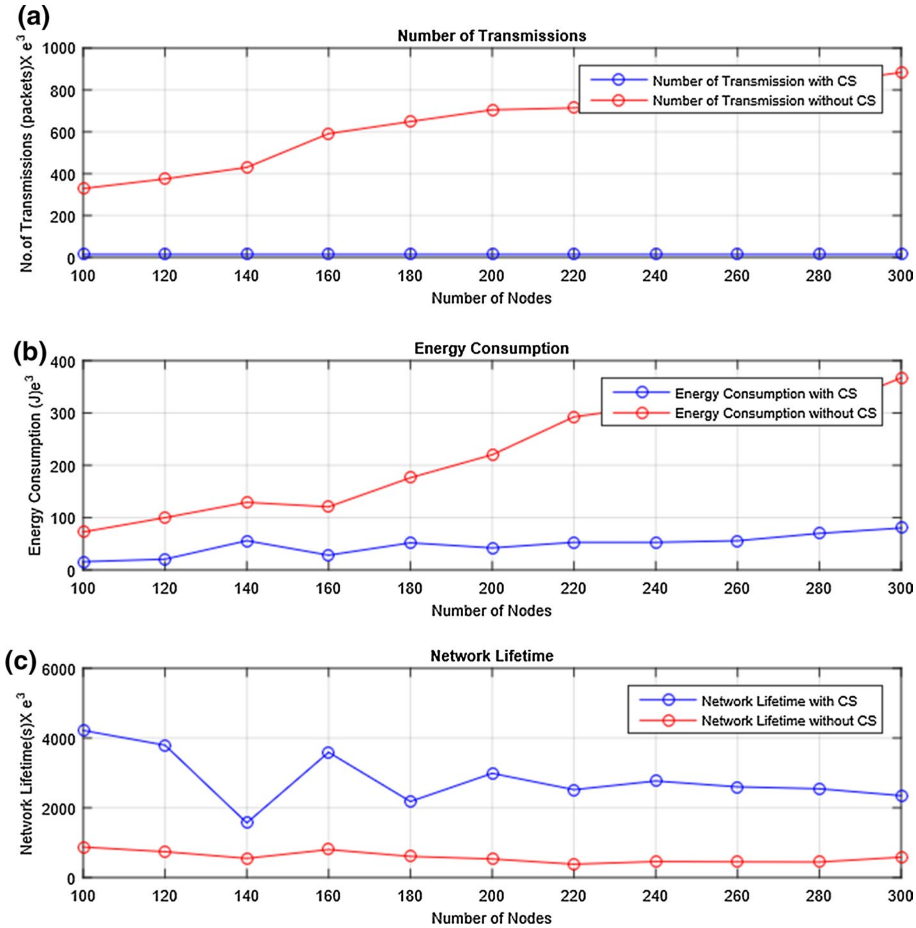


Fig. 6 Comparison of clustering with and without CS

number of transmission which are required to collect data from sensors to the base station, the energy consumed according to the energy model discussed in Sect. 6 and network lifetime. Figure 6a shows the number of transmissions for both the methods and it is evident that the transmissions for data gathering for proposed method is significantly smaller than that of clustering without CS. In clustering without CS method, all the nodes participate in transmissions as against our method where solely cluster head is gathering and transmitting data on behalf of all the member nodes in the cluster. Since transmissions are less the energy consumption is reduced and the lifetime of the network is prolonged as seen in the Fig. 6b, c respectively.

After the selection of CH by BS, correlated clusters are formed around it and then CS is applied. We have only six clusters around the BS in the beginning so only six CHs are transmitting data. But the network can expand if the selected CHs further selects CHs at twice their sensing range. In this case we get a tiered network in which the first tier will have six CHs, the second tier has twelve CHs i.e. the number of CHs will be $6 \times i$ where $i =$ tier number.

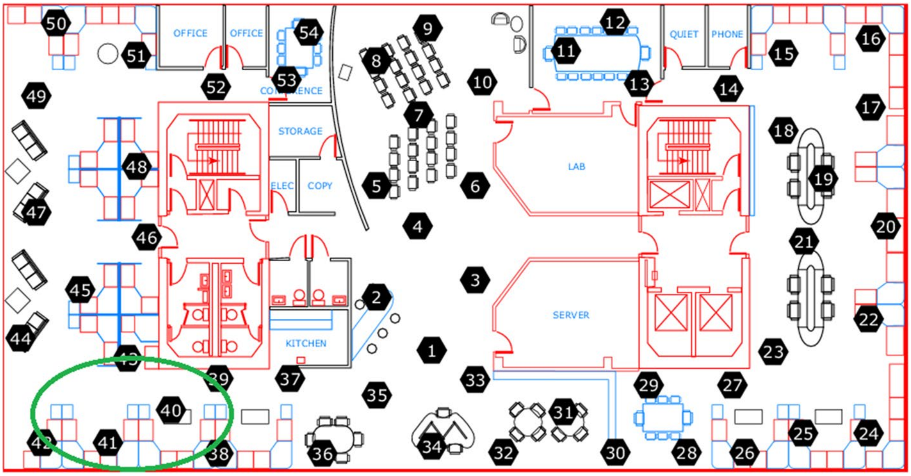


Fig. 7 Intel Berkeley lab

Table 1 Node coordinates

Node	(x location, y location)
38	(30.5, 31)
39	(30.5, 26)
40	(33.5, 28)
41	(36.5, 30)
42	(39.5, 30)
43	(35.5, 24)

The simulation for CS is developed in MATLAB R2015. The dataset used is the real data obtained by a WSN deployed at Intel Laboratory Berkeley [35]. This dataset contains temperature, humidity, light and voltage values periodically collected with 54 distributed Mica2Dot sensor nodes from 25th February–5th April 2004. Figure 7 shows the distribution of sensor nodes in Berkeley Lab

The proposed model is applied on a group of six nodes form Node38 ~ Node43 with nodes 40 acting as a CH, in the Berkeley lab as shown by a green circle in Fig. 7. To prove correlation of the spatial readings of nodes, we tabulate their spatial coordinates i.e. x and y coordinates of sensors (in meters relative to the upper right corner of the lab) in Table 1 and a graph of temperature vs time for the first two hundred readings in Fig. 8. Table I gives the spatial coordinates of sensor Node38 ~ Node43.

Figure 8 shows that the readings are highly correlated. Hence, instead of sending all the data to the sink we suppress correlated information and save the number of transmissions. Applying CS at the CH as proposed in our clustering method at node 40, while keeping node 38, 39, 41, 42 and 43 in the sleep mode, the original and reconstructed signal is depicted in Fig. 9.

Like any other compression technique, measuring the accuracy of the reconstruction is an important parameter. One of the most popular ways to do is by calculating the Root Mean Square Error (RMSE) value given by:-

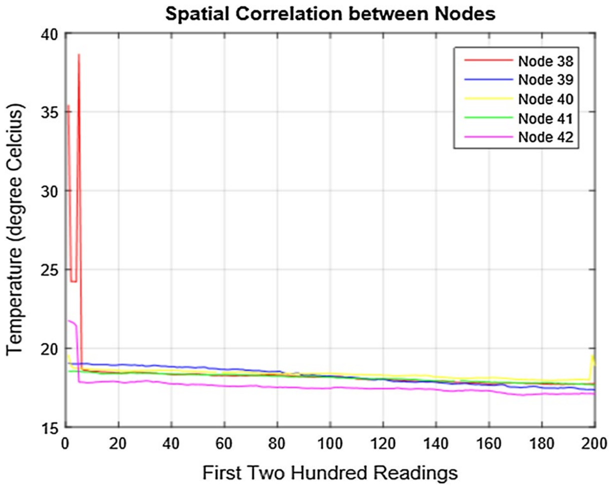


Fig. 8 Spatial correlation readings of Node 38–Node 43

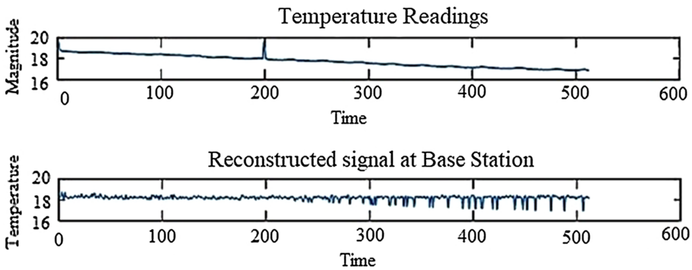


Fig. 9 Original and reconstructed signal at cluster head of correlated cluster

$$RMSE = \left\| \frac{s - \hat{s}_2}{s_2} \right\| \tag{30}$$

where s is the original signal, $\hat{s}$ is the approximated signal and $\|s\|_2 = \left(\sum_{i=1}^{n} \left| s_i \right|^2 \right)^{\frac{1}{2}}$ is the 2-norm or Euclidean length of s [36]. The RMSE calculated for our proposed method is RMSE = 0.041624

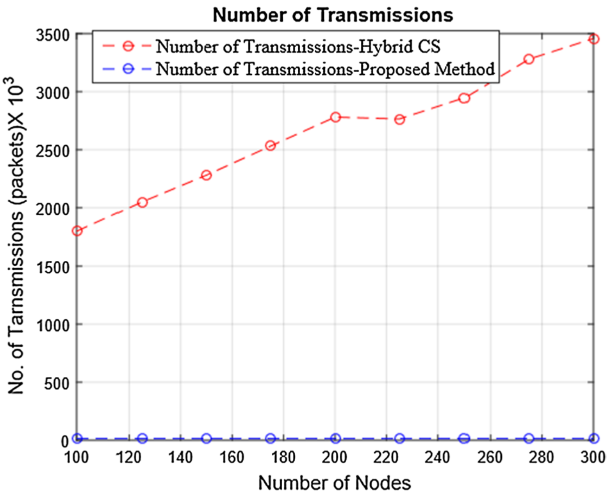
Comparision of the proposed clustering method is done with the one used in hybrid CS, where nodes first transmit data to CH and then CHs further transmit compressed data using CS to BS. Performance metrics such as, number of transmissions, energy consumption and network lifetime are compared with this method of clustering keeping the same scenario for both the methods, that is the number of nodes and energy model. We have performed the simulation of the proposed correlated cluster in NS-2 and tabulate the parameters in Table 2.

Figure 10 shows that method using hybrid CS with clustering have more number of transmissions as the density of nodes in network is increased whereas proposed method achieves reduced number of transmissions since only CHs are the nodes in the cluster that are transmitting and others being correlated are in the sleep mode. These sleeping members in the cluster further save energy consumption in the network. Thus, even if we

Table 2 Simulation Parameters

Simulator	Network simulator 2
Topology	Random
Interface type	Phy/WirelessPhy
Mac type	802.11(ad-hoc)
Queue type	Drop tail/priority queue
Queue length	100 packets
Antenna type	Omni Antenna
Propagation type	Two ray ground
Routing protocol	AODV
Transport agent	UDP
Application agent	CBR
Nerwork area	600 * 600
Number of nodes	100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300
Transmission range	60 m
Sensing range	30 m
Initial energy	50 J

Fig. 10 Number of transmission versus number of nodes



increase the node density in the network, these new nodes will passively be a part of the existing clusters. However, clusters used with hybrid CS will involve these new nodes in transmission process, thus increasing the energy consumption of the network. The plot in Fig. 11 shows the overall energy consumption with increased in number of nodes in proposed method and hybrid CS.

Network lifetime is defined as time when the first node runs out of energy. In the proposed method, CHs take turns to transmit data to the sink and change their role when the energy level reaches a threshold value E_{th}, which means not all the nodes are active at the same time. Hence, the overall lifetime of the network is prolonged. When density of node

Fig. 11 Energy consumption versus number of nodes

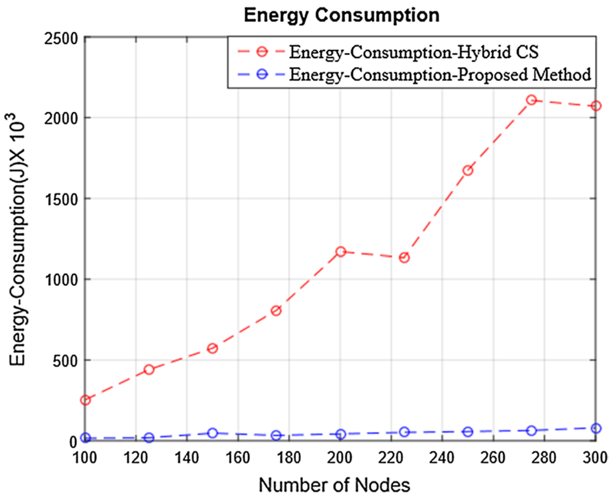
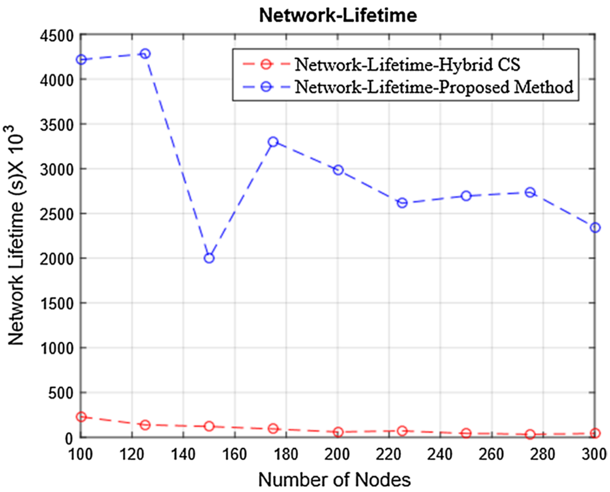


Fig. 12 Network lifetime versus number of nodes



is increased in the network there will be further improvement in the network lifetime as compared with the existing method as shown in Fig. 12.

Hence, we observe that taking advantage of spatial correlation we can save the number of transmission by only using CH to transmit data to the base station.

7 Conclusion

Redundant data generated by a dense WSN can be economically compressed with CS. Usually CS is integrated with clustering to gain its benefits neglecting the inherent spatial correlation in WSN. In this work, we exploit spatial correlation in the cluster which reduces the number of transmissions and making the network transmission efficient. Different from other clusters a spatially correlated BS assisted cluster, were correlation is estimated based on the sensing range of the CH, is proposed. The CH is selected, at the intersection of distance $2r_s$ and 60° angle line, that forms a cluster within its sensing range r_s. Six clusters when grouped together with this arrangement forms a hexagonal topology, which Lemma 1.1 proves. We schedule the member node in the cluster to sleep which reduces communication cost, and eliminates intra-cluster communication cost. NS-2 and MATLAB simulators demonstrate cluster formation and data communication. Comparing our work with existing methods of clustering, where member node transmit data to the CH and CH further transmits it to the base station, much higher transmission efficiency is achieved in this work. Along with saving on number of transmissions, we also conserve energy consumption in the network. This work executes a detailed energy analysis of cluster formation and data transmission. In spite of sending fewer measurements using CS at CH, the original signal is recovered at the base station with low RMSE value. Thus, we have integrated compressive sensing with spatially correlated cluster avoiding communication overhead and reducing the number of transmissions. The network lifetime is prolonged along with reduced energy consumption, which is the main bottleneck for a battery operated sensor.

Appendix

Consider triangle ABC with inscribed circle of radius $OM = r_s$ (Fig. 13).

$$AM = 2(OM)$$

In the right angle triangle *ANM,*

$$\textit{Therefore } \theta = \sin^{-1}\frac{1}{2} = 30°$$

Considering, ΔAOB, BC is tangent to the circle at O. Therefore ΔAOB is a right angled triangle at O and $\angle BAO = 30°$

Since $\angle A + \angle B + \angle O = 180°$

Therefore $\angle B = 60°$

Similarly $\angle C = 60°$ and $\angle A = 60°$

Since $\angle A = \angle B = \angle C = 60°$

Thus, the triangle is an equilateral triangle.

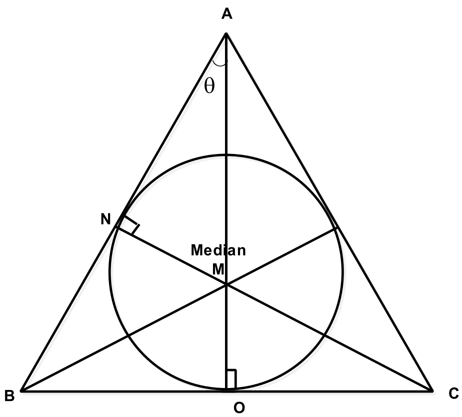


Fig. 13 Triangle ABC

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Communications of the ACM, 38*(4), 393–422.
2. Dabirmoghaddam, A., Ghaderi, M., & Williamson, C. (2010). Cluster-based correlated data gathering in wireless sensor networks. In *2010 IEEE international symposium on modeling, analysis and simulation of computer and telecommunication systems (MASCOTS)*. IEEE.
3. Rooshenas, A., Rabiee, H. R., Movaghar, A., & Naderi, M. Y. (2010, December). Reducing the data transmission in wireless sensor networks using the principal component analysis. In *2010 Sixth international conference on intelligent sensors, sensor networks and information processing (ISSNIP)* (pp. 133–138). IEEE.
4. Gastpar, M., Dragotti, P. L., & Vetterli, M. (2006). The distributed karhunen–loeve Transform. *IEEE Transactions on Information Theory, 52*(12), 5177–5196.
5. Candès, E. J., & Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine, 25,* 21.
6. Lin, H., Wang, L., & Kong, R. (2015). Energy efficient clustering protocol for large-scale sensor networks. *IEEE Sensors Journal, 15*(12), 7150–7160.
7. Xie, R., & Jia, X. (2014). Transmission-efficient clustering method for wireless sensor networks using compressive sensing. *IEEE Transactions on Parallel and Distributed Systems, 25*(3), 806–815.
8. Duarte, M. F., Shen, G., Ortega, A., & Baraniuk, R. G. (2012). Signal compression in wireless sensor networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370*(1958), 118–135.
9. Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory, 52*(4), 1289–1306.
10. Luo, C., Wu, F., Sun, J., & Chen, C. W. (2009, September). Compressive data gathering for large-scale wireless sensor networks. In *Proceedings of the 15th annual international conference on mobile computing and networking* (pp. 145–156). ACM.
11. Luo, J., Xiang, L. & Rosenberg, C. (2010). Does compressed sensing improve the throughput of wireless sensor networks? In *2010 IEEE international conference on communications (ICC)*. IEEE.
12. L. Xiang, J. Luo, & Vasilakos, A. (2011, June). Compressed data aggregation for energy efficient wireless sensor networks. In *Proceedings of the IEEE sensor, mesh and ad hoc communications and networks (SECON'11)* (pp. 46–54).
13. Nguyen, M. T., & Teague, K. A. (2014). Compressive sensing based data gathering in clustered wireless sensor networks. In *2014 IEEE international conference on distributed computing in sensor systems (DCOSS)*. IEEE.
14. Liu, D., Zhou, Q., Zhang, Z., & Liu, B. (2016). Cluster-based energy-efficient transmission using a new hybrid compressed sensing in WSN. In *2016 IEEE conference on computer communications workshops (INFOCOM WKSHPS)* (pp. 372–376).
15. Lan, K.-C., & Wei, M.-Z. (2017). A compressibility-based clustering algorithm for hierarchical compressive data gathering. *IEEE Sensors Journal, 17*(8), 2550–2562.
16. Lee, S., Pattem, S., Sathiamoorthy, M., Krishnamachari, B., & Ortega, A. (2009, July). Spatially-localized compressed sensing and routing in multi-hop sensor networks. In *International conference on geosensor networks* (pp. 11–20). Springer, Berlin.
17. Xiaoronga, C., Mingxuan, L., & Suc, L. (2012). Study on clustering of wireless sensor network in distribution network monitoring system. *Physics Procedia, 25,* 1689–1695.
18. Vuran, M. C., & Akyildiz, I. F. (2006). Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE/ACM Transactions on Networking, 14*(2), 316–329.
19. Deng, X., Zhong, W., Ren, J., Zeng, D., & Zhang, H. (2016). An imbalanced data classification method based on automatic clustering under-sampling. In *2016 IEEE 35th international performance computing and communications conference (IPCCC)* (pp. 1–8). IEEE.
20. Yuan, J., & Chen, H. (2009, Sep). The optimized clustering technique based on spatial-correlation in wireless sensor networks. In *Proceedings of the IEEE youth conference Information*, *telecommunication* and *computing YC-ICT* (pp. 411–414).
21. Liu, C., Wu, K., & Pei, J. (2007). An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. *IEEE Transactions on Parallel and Distributed Systems, 18*(7), 1010.
22. Shakya, R. K., Singh, Y. N., & Verma, N. K. (2013). Generic correlation model for wireless sensor network applications. *IET Wireless Sensor Systems, 3*(4), 266–276.

23. Liu, Z., Xing, W., Zeng, B., Wang, Y., & Lu, D. (2013, March). Distributed spatial correlation-based clustering for approximate data collection in WSNs. In *2013 IEEE 27th international conference on advanced information networking and applications (AINA)* (pp. 56–63). IEEE.
24. Ramesh, K., & Somasundaram, K. (2012). A comparative study of clusterhead selection algorithms in wireless sensor networks. arXiv preprint arXiv:1205.1673.
25. Vuran, M. C., Akan, Ö. B., & Akyildiz, I. F. (2004). Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks, 45*(3), 245–259.
26. Zhang, H., & Hou, J. C. (2005). Maintaining sensing coverage and connectivity in large sensor networks. *Ad Hoc and Sensor Wireless Networks, 1*(1-2), 89–124.
27. Papastergiou, S., Polemi, D., & Douligeris, C. (2009). SWEB: An advanced mobile residence certificate service. In *International conference on e-democracy* (pp. 421–430). Springer, Berlin.
28. Wang, B. (2010). *Coverage control in sensor networks. Computer communications and networks*. Berlin: Springer. https://doi.org/10.1007/978-1-84996-059-5_2.
29. Liaqat, M., et al. (2014). HEX clustering protocol for routing in wireless sensor network. In *2014 IEEE 28th international conference on advanced information networking and applications (AINA)*. IEEE.
30. Ouni, S., & Ayoub, Z. T. (2013). Predicting communication delay and energy consumption for IEEE 802.15.4/Zigbee wireless sensor networks. *International Journal of Computer Networks Communications, 5*(1), 141.
31. Heinzelman, W. B. (2000). *Application-specific protocol architectures for wireless networks*. Diss. Massachusetts Institute of Technology.
32. Yuan, F., Zhan, Y., & Wang, Y. (2014). Data density correlation degree clustering method for data aggregation in WSN. *IEEE Sensors Journal, 14*(4), 1089–1098.
33. Razzaque, M. A., & Dobson, S. (2014). Energy-efficient sensing in wireless sensor networks using compressed sensing. *Sensors, 14*(2), 2822–2859.
34. Zhu, L., Ci, B., Liu, Y., & Chen, Z. (2015). Data gathering in wireless sensor networks based on reshuffling cluster compressed sensing. *International Journal of Distributed Sensor Networks, 11*(11), 260913.
35. Intel Labs Berkeley Data. http://db.csul.mit.edu/www.select.cs.cmu.edu/data/labapp3/.
36. Quer, G., Masiero, R., Pillonetto, G., Rossi, M., & Zorzi, M. (2012). Sensing, compression, and recovery for WSNS: Sparse signal modeling and monitoring framework. *IEEE Transactions on Wireless Communications, 11*(10), 3447–3461.

Utkarsha S. Pacharaney received her M.E and B.E degree in Electronics and Telecommunication from the University of Mumbai in the year 2011 and 1999 respectively. She is currently working towards the Ph.D. degree from the University of Mumbai. Her research interest includes wireless communication and sensor networks. She has a teaching experience of over 16 years and currently working with Datta Meghe College of Engineering, Airoli.

Dr. Rajiv Kumar Gupta has received his Ph.D. from the department of Electrical Engineering, IIT Bombay, 2010. He has received his M.S. (By Research) Communication Engineering from IIT Delhi in the year 2011 and BTech in Electronics and Communication Engineering at NIT Kurukshetra in the year 1988. He is currently working as Professor in the department of Electronics and Telecommunication at Terna Engineering College. His research interest includes Antenna and RF communication circuits. He has published around 70 papers in conference and journals.