An Energy-Efficient Heterogeneous Data Gathering for Sensor-Based Internet of Things

Gaurav Tripathi[1*], Vishal Krishna Singh[1] and Brijesh Kumar Chaurasia[2]

[1]Wireless Communications and Analytics Research Lab, Department of Computer Science, Indian Institute of Information Technology, CG City, Lucknow, 226002, Uttar Pradesh, India.
[2]Department of Computer Science, Pranveer Singh Insitute of Technology, Kanpur, 209305, Uttar Pradesh, India.

*Corresponding author(s). E-mail(s): gt50@live.in;
Contributing authors: vashukrishna@gmail.com;
brijeshchaurasia@ieee.org;

Abstract

Data gathering in the Sensor-Based Internet of Things is done in the midst of many constraints such as high in-network transmissions, uneven load distribution, high energy depletion, and data heterogeneity. Interestingly, compressed sensing based solutions for heterogeneous data gathering have been widely used to resolve these issues but remain unexplored for Sensor Based Internet of Things. Therefore, with the aim of optimizing the in-network transmissions and achieving uniform load distribution in the network, this work presents a novel compressed sensing based algorithm for heterogeneous data gathering in sensor-based Internet of Things. A random markov model is used to obtain co-relation-based segregation of the region of interest, followed by a novel compressed sensing based sampling and data gathering scheme. Simulation results are obtained for two different scenarios by varying the sink position with respect to the region of interest. Comparative analysis, with state of the art methods, proves the efficacy of the proposed scheme over existing methods where the proposed scheme achieves an improved performance of **81%** and **44%** for network lifetime and average energy consumption respectively.

Keywords: Compressed Sensing, In-Network Transmission, Internet of Things, Load Distribution, Machine Learning

1 Introduction

The Sensor-Based Internet of Things (SBIoT) is an interconnected network of real-world objects equipped with heterogeneous sensors, connected to the internet using heterogeneous protocols to achieve a common goal. Every connected object in SBIoT is equipped with multiple sensors collecting heterogeneous streaming data. Since the number of connected objects in SBIoT is huge and almost every object is equipped with one or more sensors, gathering data in real-time, the data being gathered in SBIoT is huge in size. Gartner predicted that the amount of data generated by connected SBIoT will be 79.4 Zettabytes (ZB) by the end of the year 2025. This massive heterogeneous data results in issues such as interoperability, scalability, network congestion, uneven energy consumption, high in-network transmissions, reduced network lifetime, etc.

Considering the high energy dissipation, in sending and receiving wireless communications, several schemes have been proposed to minimize the in-network transmissions in wireless networks such as SBIoT [1]. Compressed Sensing (CS) as one of the popular data gathering scheme has gained the desired attention in recent years because of several associated advantages with the technique such as uniform load distribution, reduced in-network transmissions, etc [2]. According to CS theory, a modest collection of linear projections is adequate for almost flawless reconstruction of a compressible signal, saving energy and extending network lifetime. Use of CS based gathering in SBIoT remains widely unexplored as most of the state-of-the-art approaches, consider alternate solutions for optimizing network lifetime.

In state-of-the-art method proposed in [3], a balanced tree-based distributed data gathering algorithm (DGABT) is proposed to address the issue of network lifetime. However, this approach lacks the uniformity in load distribution and imposes extra overhead on the execution time. In yet another approach based on random walk data gathering, proposed in [4], the information related to the construction tree is broadcasted using a sink node. The algorithm guarantees the shortest hop-to-hop distance, however lacks in controlling the nodes at the time of spanning tree construction. The tree is thus, poorly balanced which leads to the uneven load distribution in the network. Another tree-based algorithm, named as Local Minimum Spanning Tree

(LMST) is proposed in [5] to provide an improved random routing-based data gathering scheme where the process of contribution tree is adjusted using a trade-off between energy and distance. The algorithm is able to reduce energy consumption to a certain extent. However, the high data transfer latency imposes increased communication overhead.

A robust data collection tree (RDCT) construction algorithm is proposed in [6] focusing on load distribution with optimized lexicographic rate assignment. A work proposed in [7] titled MaxLeach proves to be more energy-efficient than the traditional leach protocol and a demand-based data gathering scheme is proposed in [8] with machine learning (ML) based multi-query optimization. On-demand data collection strategy reduces the communication overhead on the network thereby improving the network lifetime. Integrated data and energy gathering (iDEG) framework in [9] and Wavelet transform for energy efficient data gathering and matrix completion (WEED-MC) in [10] are compressed sensing based data gathering schemes. While the focus of iDEG is on reducing the energy loss and complexity of the algorithm, WEED-MC focuses on the recovery of the compressed data. A careful analysis of the work reveals the fact that although iDEG is able to address the issues of time-complexity and energy loss, the scope of the work is limited only to homogeneous data. WEED-MC on the other hand achieves better accuracy and a high recovery rate but the energy consumption is not addressed in the work. Another work proposed in [11] aims at achieving an improved lifetime of the network by accurate detection of important events in Wireless Sensor Networks (WSN). Since the data generated in an SBIoT network is primarily heterogeneous, one of the major constraints of this work is its limited consideration for homogeneous data only [12].

This work, therefore, aims at obtaining an energy-efficient data gathering strategy for heterogeneous data collection in SBIoT. The following are the novel contributions of this work:

1. A Markov distribution based correlated demarcation of the region of interest (RoI).
2. A compressive data gathering scheme to obtain random sampling of the network data.
3. A theoretical foundation is provided to analyze the performance of the proposed method with the aim to tide over the gap between CS and ML.
4. The validation of the effectiveness of the proposed scheme through extensive simulations and evaluation of the performance of heterogeneous data gathering and its impact on energy consumption is also presented

.

The rest of the paper is organized as follows: in Section-2, a thorough review of related work is discussed followed by Section-3, where a detailed problem formulation is presented. The Section-4, presents the overall system environment while the Section-5 presents the proposed approach to solve the addressed problem. The Section-6, presents the considered model for the energy consumption. In Section-7, a comprehensive mathematical analysis of the proposed method is presented, followed by the Simulation setup in Section-8. Finally, the results are presented in Section-9, followed by concluding comments in Section-10.

2 Literature Review

The performance of a SBIoT network depends on various factors such as, load balancing, energy efficient data gathering, co-relation based demarkation of RoI and many more. Existing work, in this domain, essentiates the need to address these factors through various seminal works.

Load balancing makes it easier to distribute workload across resources in order to continue providing services even if a service component fails. For instance the work proposed in [13] presents an efficient use of resources by provisioning and de-provisioning instances of applications. The authors in [14] proposed a model which binds the Bacterial Forging, Genetic, and an optimized Bee colony algorithms together. The scheme is able to achieve an improvement over the traditional load balancing algorithms, but it fails to address issues of Quality of Service (QoS) and security.

A fog-based architecture is proposed in [15], that supports time-bound Internet of Things (IoT) applications. One of the significant works of the past years is the work proposed in [16] where two protocols are proposed for dynamic environment of wireless body area network. The first protocol aims to obtain application specific QoS, and second aims to identify an optimum value by optimizing all QoS affecting factors. The study in [17] proposes a multithreaded architecture on lock-free data structures for streaming overlays or congested links that enhances performance by avoiding thread synchronisation.

In [18], a MAC layer uplink solution is designed to address the issue of energy-aware data collection in IoT based surveillance application for remote area. The narrowband Internet of Things (NB-IoT) scheduled access scheme specified by 3GPP has been implemented. The proposed methodology is limited to traditional algorithmic solution and the network does not learn from its previous states. A similar work, proposed in [19], supports the data gathering with the help of mobile edge computing. An enhanced Sort-Tile-Recursive tree based network division method is used for the clustering of edge nodes. The method is able to decrease the energy depletion. The work proposed in this is not much suited for densely deployed network, also in multi-edge networks it does not provide request based discovery of biased IoT nodes.

The work presented in [20] proposed a secure and energy-aware data gathering, CS method is integrated with public key and trust towards achieving an energy efficient load balancing in a secure channel. Although, the network lifetime is less than other data gathering schemes, the major advantage of the method is its ability to deliver a secure transmission method. The authors in [21] proposed a fuzzy based clustering method based on hop count. The approach first focuses on selection of Cluster Head (CH) on the basis of its weight using an energy aware gravitational search algorithm (GSA) followed by cluster formation. A Fuzzy Multi-criteria Clustering, hierarchical routing protocol with Energy-aware routing is presented in [22] with the aim of improving the network lifetime. The authors in [23] proposed an Ant Colony Optimization based energy-aware multipath routing protocol for WSN, which consists of three phases; namely link knowledge based Neighbour Discovery, Packet Transmission based on moving average method that is exponentially weighted, and end-to-end assured packet delivery.

Corelation based demarkation of RoI has been explored widely for optimizing the performance of sensor network. Markov process is known to be used as a mathematical model to analyse intricate technological and informational (cyber-physical) systems. The capacity to construct predictively controlled models in a cyber-physical system and subsets over time on the basis of the statistical data or the findings of observations is one of Markov processes' key benefits. A mathematical model of IoT ecosystem is proposed in [24] to address the subject system information process. A discrete Markov chain is reduced from simple queuing system to manage the system resources based on geometric distribution. The description of the studied system's operational process in the context of hostile cyberspace is presented in [25] using a Markov approach. One of this work's limitations is its insufficient attention to the parametric space for the extreme RIS operation's efficiency coefficient value. In order to operate a cyber-physical system, Markov model based two functional security metrics were described in [26]. This is further demonstrated by the general Vulnerability Priority Rating (VPR) system, where it is possible to effectively estimate the parameters of the model using sparse empirical data.

3 Problem Description

The perceptual layer of IoT is responsible for sensing the data from the real world and sending the converted digital data to the upper layers. Since the sensors are deployed in a common area, the data is almost similar, and the changes recorded are almost identical. Additionally, the transmission of this data results in a congested IoT network, with uneven load distribution, resulting in uneven energy consumption and costing early shutdown of the SBIoT network. Moreover, the data thus obtained is unreliable and unpredictable because of its heterogeneous nature. To represent the situation mathematically consider if D_i is the difference (rise or fall) recorded at instance i from its earlier instance $i-1$, and S_i denotes the signal reading at instance i.

$$D_i = S_i - S_{i-1}$$

By applying the above formula in an ideal environment, it was observed that in most of the cases the difference is negligible, for example from 100 readings we obtained 99 values of D_i out of which only 13 times the value obtained was more than 10.

The SBIoT network is equipped with heterogeneous sensors producing data in heterogeneous formats and the handling of such data is very complex. The data thus obtained can be defined as follows:

$$\left(D^{\alpha_i}_{\beta_j}\right)_t$$

Where D denotes the gathered data from a sensor α_i of a SBIoT node β_j at a time t.

The average of all the data gathered in the duration of the network lifetime is given by:

$$S_{avg} = \sum_{t=1}^{n} \left(S_{avg}\right)_t \tag{1}$$

Such that,α_i is i^{th} sensor installed in SBIoT node, β_j is j^{th} SBIoT node deployed in the area, and n is ever-increasing time.

However, clustering of heterogeneous data is an open research problem because the heterogeneous nature of the network results in frequently changing data formats.

Consider a set of n SBIoT nodes $\{SN_1, \ldots SN_n\}$ deployed in a sub-region $\Omega \subseteq R^2$. The position of each sensor $SN_i^{pos} \in \Omega^2$ is considered to be known. Five heterogeneous sensors are equipped in each SBIoT node. The pre-known positioning of SBIoT nodes is a typical assumption in the literature.

Finally, SN_{W_i} is a weight calculated on the basis of the remaining energy and the correlation structure of the i^{th} node in the objective function and SN_{E_i} is the energy dissipated by the $\mathrm{i^{th}}$ node to for the completion of the transmission of its data, written as:

$$\begin{aligned} E_{S_i} &= R_{Tx} \times \sum_{i=1}^{n} x_{i,sink} T^{com}_{sink,i} \\ &= \frac{R_{Tx} Pkt_i}{\sum_{i=1}^{n} x_{c,k} \left\langle R\left(SN_i, SN_{sink}\right)\right\rangle^{R_i^{\max}}_{R_i^{\min}}} \end{aligned} \tag{2}$$

where R_{Tx} is the transmission power of node and $T^{com}_{sink,i} = \frac{M_i}{\langle R(SN_i, SN_{sink})\rangle^{R_i^{\max}}_{R_i^{\min}}}$ is the average travel time of the data packet of amount Pkt_i from node to the sink node SN_{sink}.

Then the optimization problem is defined as follows:

$$\text{Minimize } (E_D + \sum_{i=1}^{n} P_i E_{S_i}) \tag{3}$$

$$(X_i \in \Omega_{1 \le i \le n})$$

$$(x_{i,sink} \in 0, 1_{1 \le i \le n})$$

$$(y_{i,i'} \in 0, 1 \begin{array}{c} 1 \le i \le n \\ 0 \le i \le n \end{array})$$

$$\text{subject to } (E_{S_i} = E_i^{\max}, 1 \le i \le n)$$

Considering the above scenario for CS framework, let $x = (x_1, x_2, \ldots, x_m)^T$ denote the data sample measurement by all SBIoT nodes in one data collection round. Assume that in some domain Ψ, i.e., $x = \Psi\Theta$, signal x is k-sparse, where $\Theta = (\theta_1, \theta_2, \ldots, \theta_m)^T$ is the vector of transform coefficient in Ψ. Let the measurements gathered by the CH from n SBIoT nodes are denoted by $y = (y_1, y_2, \ldots, y_n)^T$. Then, via a random measurement matrix Φ, y may be thought of as a compressed form of x, here, $y = \Phi x$, where Φ is a $n \times m$ matrix with $n \ll m$. The proposed data gathering strategy is used to characterise the measurement matrix Φ, where every row denotes a certain node's vector in a cluster. Resulting to at most one nonzero element in each row of Φ.

4 System Model

A randomly deployed SBIoT network is simulated for a park area of $200m. \times 200m$. The RoI is considered with areas with dense tree plantation, a water body, an open field, and a garden with small plants. A homogeneous demarcation of RoI is obtained by feeding δ and λ as the probability of varying terrain and probability of variation in the observation respectively to a standard Markov distribution model. The RoI is divided into 4×4 clusters with 1 cluster head each. Every deployed node is capable of reporting 5 observations namely temperature, pressure, smoke, wind speed, and humidity.

The general assumption is that the observations reported from a cluster have high co-relation and the probability of change of terrain (as well as physical phenomena in the cluster) is very less. For example, the cluster with a high density of tree plantation is likely to report less temperature and wind speed as compared to the open field. Similarly, the cluster surrounding the water may report high humidity levels as compared to the other regions of the RoI. A model scenario reflecting the simulated scenario is shown in Figure 1.

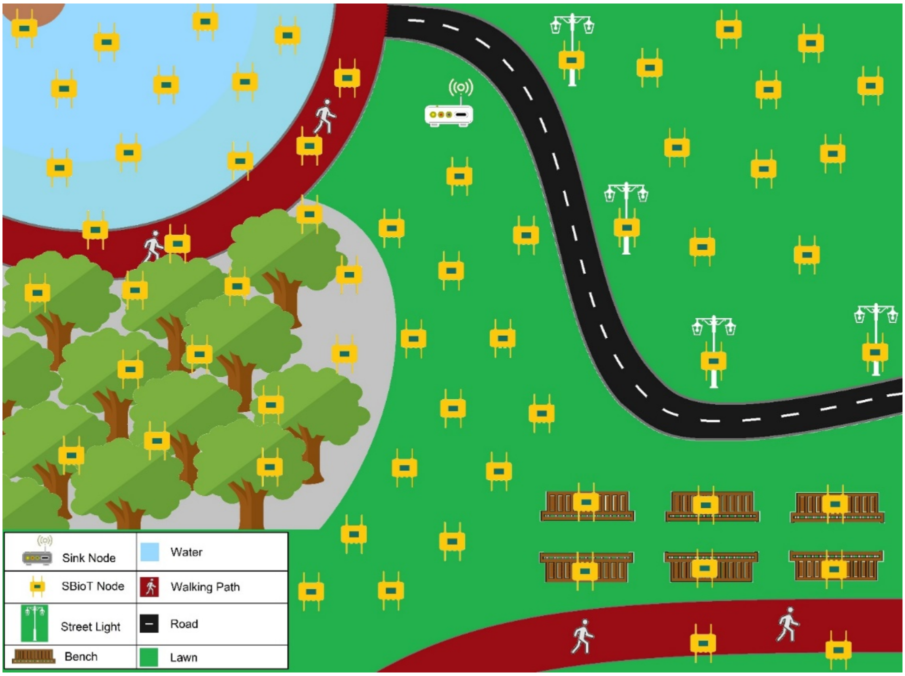

Fig. 1 Region of Interest (RoI)

5 Proposed Methodology

In case of an event, the nodes sense and report the events to the CH. Since the observations received at CH are for five heterogeneous parameters, the CH stores the observations in n dimension tensor. For every observation received at the CH, a pseudo-random sampling is performed by choosing the first weighted sample randomly, followed by $\alpha = \{\mathrm{L}+(\mathrm{C})\}^{\mathrm{th}}$ weighted sample being chosen next and $\alpha + (C)^{\mathrm{th}}$ where L is the index of the randomly chosen sample, and C is a constant. This pseudo-random sampling allows the proposed approach to overcome the drawbacks of accuracy in small-scale networks. Thus, for every reported observation 3 random samples are forwarded to the base station for processing, resulting in the formation of a storage data matrix. The data matrix is recovered using ℓ_1-magic resulting in a significant reduction in the in-network transmissions and without any significant loss of data.

5.1 Cluster Formation and CH Election

The stepwise details of the proposed method is discussed below:

1. **Markov model for co-related demarcation**
   The probability of varying terrain δ and the probability of variation λ is fed for homogeneous demarcation of RoI. The clusters are formed based on this demarcation. Then the nodes are randomly deployed.

2. **CH Election**
   Calculate the initial weights of each node based on its remaining energy and correlation structure. In each cluster, the node with the highest weight is elected as the CH for that particular round. The ML-based Boosting ensemble method is used for weight-based CH election in every round.
3. **Data Transmission**
   Calculate the shortest path between all CHs to the sink node using the Dijkstra algorithm. Here only the CH nodes and sink nodes are considered as vertices and the distances between these vertices are considered as edges. Transmit the data packets from each node towards the sink from the shortest path. The data transmission is done based on Algorithm-2.

5.2 Data Compression and Data Sampling

The step-wise details of the proposed method is discussed below:

1. **Data Compression at node level**
   For each node calculate the average of data gathered in three recent rounds by that node and save this average data to the data packet of that node. The data packet is shown in Figure 2 (a).
2. **Tensor Formation**
   CH saves the heterogeneous data received from different sensors to a n-dimension Tensor as shown in Figure 2 (b).
3. **Data Sampling**
   10 recent frames are marked with numbers from 0 to 9. L = a random number between 0 and 9.
   $\alpha = (\mathrm{L}+3) \mod 10$
   $\beta = (\alpha+3) \mod 10$
   Out of ten recent data frames transmit only L^{th}, α^{th}, and β^{th} data frames to the sink node. Drop all the other data frames. Data frames are shown in Figure 2 (c).

6 Energy Consumption Model

The average data rate for the communication between an SBIoT node SN_i and the sink node SN_{sink} is denoted by $R(SN_i, SN_{sink})$ defined as:

$$R\left(SN_i, SN_{sink}\right) = B_{log_2}\left(1 + \frac{P_r}{PL_{SN-sink}\left(SN_i, SN_{sink}\right) N_0}\right) \quad (4)$$

Where $PL_{SN-sink}$ is the average channel path loss, SBIoT node to sink probabilistic path loss model is considered. The average path loss between an SBIoT node SN_i and the sink SN_{sink} is expressed as:

Algorithm 1 Cluster formation and CH Election

Deploy n SBIoT nodes randomly in the area $x_m \times y_m$ square meters
Deploy sink in a given position according to simulation setup
Initialize the energy constants as given in Table 1
Divide the RoI into clusters using Markov model demarcation
for $iteration = 1, 2, \ldots, N$ **do**
    assign a cluster number, based on the coordinates of $iteration^{th}$ node
end for
while Number of Dead Nodes $< n$ **do**
    reset the *Activity status* and *Correlation count* of all SBIoT nodes.
    calculate $SN_{C_i} = Count(N_{i.cluster} \cap N_{i.range})$
    if round == 1 **then**
        assign SN_{W_i} to node i based on SN_{C_i} and SN_{E_i}
    else
        update SN_{W_i} of all the nodes using Boosting Ensemble Method.
    end if
    for all clusters **do**
        CH is elected based on SN_{W_i}
    end for
    if alive_nodes > 3 **then**
        for i in range (n/3) **do**
            assign $SN_{Act_i} = 1$; here i is chosen randomly
        end for
    end if
    for all nodes **do**
        Calculate the shortest path from i^{th} node to sink node via CHs
        Calculate the distance between node and its next hops
    end for
    for all nodes **do**
        if SN_i is not CH and $SN_{Act_i} == 1$ **then**
            SN_{Act_i} is allowed to transmit the data
            $E_{Tx_i} = (E_{elec} \times Pkt_i) + (E_{amp} \times Pkt_i \times (SN_{nextHop_i})^2)$
            $SN_{E_i} = SN_{E_i} - E_{Tx_i}$
            $E_{Rx_{i.CH}} = (E_{elec} \times E_{DA}) \times (Pkt_i)$
            $SN_{E_{i.CH}} = SN_{E_i} - E_{Rx_i}$
            $SN_{stream_{i.CH}} = SN_{stream_{i.CH}} + Pkt_i$
        else
            $E_{Tx_i} = (E_{elec} + E_{DA}) + (E_{amp} \times SN_{stream_i} \times (SN_{nextHop_i})^2)$
            $SN_{E_i} = SN_{E_i} - E_{Tx_i}$
        end if
    end for
end while

Algorithm 2 Data Compression and Data Sampling

while Network is not dead **do**
    Gather data using sensors equipped in SBIoT nodes
    for all nodes **do**
        D_i^{avg} = Average of data gathered in 3 recent rounds by node i
        TransmitD_i^{avg} as a data packet to the respective Cluster Head
        of node i ▷ Data Packet is shown in Figure 2 (a)
    end for
    for i in range(10) **do**
        Store received data packet to the $n \times m \times l$ tensor of the Cluster
        in its respective position ▷ Tensor is shown in Figure 2 (b)
    end for
    $dataFrame_i = Tensor_i$ ▷ Data Frames are shown in Figure 2 (c)
    L = Random number between 0 and 9
    α = (L+3) mod 10
    β = (α+3) mod 10
    Forward the L$^{\text{th}}$ data frame to Sink
    Forward the α^{th} data frame to Sink
    Forward the β^{th} data frame to Sink
    Clear buffer
end while

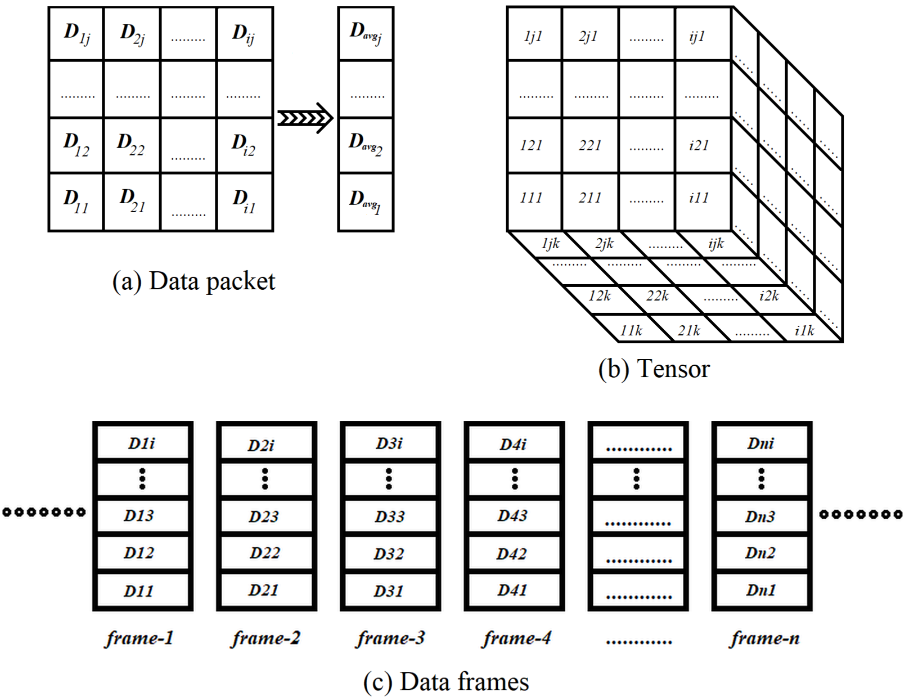

Fig. 2 Data in proposed methodology

$$\begin{aligned}&PL_{SN-sink}\left(SN_i, SN_{sink}\right) = \\ &P_{LoS}\left(SN_i, SN_{sink}\right) PL_{LoS}\left(SN_i, SN_{sink}\right) + \\ &\left[1 - P_{LoS}\left(SN_i, SN_{sink}\right)\right] PL_{NLoS}\left(SN_i, SN_{sink}\right)\end{aligned} \tag{5}$$

Where $P_{LoS}\left(SN_i, SN_{sink}\right)$ is the line-of-sight (LoS) probability of between a SBIoT node SN_i and the sink node SN_{sink}. This probability is based on surroundings as well as the terrain's altitude. As presented in [27], it may be expressed as follows:

$$\begin{aligned}&P_{LoS}\left(SN_i, SN_{sink}\right) = \\ &\frac{1}{1 + \alpha\left(\exp\left(-\beta\left(\theta\left(SN_i, SN_{sink}\right) - \alpha\right)\right)\right)}\end{aligned} \tag{6}$$

Where $\theta\left(SN_i, SN_{sink}\right)$ is the elevation angle of the sink with regards to the SBIoT node SN_i when α and β are environmental variables such as the proportion of water area to the RoI, the number of trees per square meter, and the distribution of heights of trees. An empirical technique is developed in [27] to compute these parameters as the function of features of the urban area. Finally, for LoS and non-line-of-sight (NLoS) area, $PL_{LoS}\left(SN_i, SN_{sink}\right)$ and $PL_{NLoS}\left(SN_i, SN_{sink}\right)$ are the average path losses, expressed as:

$$\begin{aligned}&PL_{LoS}\left(SN_i, SN_{sink}\right) = \\ &10\eta log_{10}\left(\frac{4\pi f_c}{c}||SN_i - SN_{sink}||_2\right) + \xi_{LoS} \\ &PL_{NLoS}\left(SN_i, SN_{sink}\right) = \\ &10\eta log_{10}\left(\frac{4\pi f_c}{c}||SN_i - SN_{sink}||_2\right) + \xi_{NLoS}\end{aligned} \tag{7}$$

where the free-space path loss is represented by the first component, where η represents the exponent of the path loss, f_c representing the frequency of carrier, c representing the celerity of light, and $||V||_2$ representing the vector V's norm-2 (i.e., the euclidean distance between $SN_i$and SN_{sink} is represented by $||SN_i - SN_{sink}||_2$. The other component, on the other hand, indicates the mean value of excessive route loss (i.e., ξ_{LoS} is the mean value of excessive route loss in LoS and ξ_{NLoS} is the mean value of excessive route loss in NLoS).

7 Analytical Analysis

Based on the given network and system model the proposed scheme focuses on energy-efficient compressed sensing based data gathering. The proposed scheme aims to increase the lifetime of the network by addressing the energy consumption issues, uneven load distribution, in-network transmissions, etc. The following discussion highlights the mathematical analysis of the proposed approach in terms of energy consumption, data compression, and CS-based data gathering.

7.1 Data Compression

Assume that the data packet generated at the end node is of size $DRx_{pkt} = 1024bits$ and that the data received at the sink node is of size DRx_{sink}. Therefore, total data generated in a particular round i by the end node is:

$$D^i_{node} = n_i \times DRx_{pkt} \tag{8}$$

where n_i represents the count of SBIoT nodes alive in round i.

Now considering the first round when all nodes are alive, the data gathered is given by:

$$DRx^1_{node} = n_1 \times DRx_{pkt} \times t = 500 \times 1024 \times 10 = 62.5MB$$

And the total data gathered by the sensor nodes in a round i is given by:

$$DG^i_{node} = DRx^i_{node} \times t = 62.5 \times 10 = 625MB \tag{9}$$

The proposed algorithm allows 3 pseudo-randomly chosen samples to be transmitted at predefined intervals of 10 μs. Evidently, the sampling reduces the data at least by 70%. Consider that $DRx^i_{CH\_1}$ is the size of the data packet received by the first CH from nodes in round i, and $DTx^1_{node\_1}$ is the size of data transmitted from $node\_1$ in round i then:

$$DTx^i_{node\_1} = 30\% of (D_{pkt} \times t) = 30\% of (1024 \times 10) = 384B$$

$$DRx^i_{CH} = \sum_{x=0}^{n_{CH\_1}} DTx_x \tag{10}$$

Considering the average number of nodes in one cluster

$$n^i_{avg_CHi} = \frac{n}{16} = \frac{500}{16} = 31.25$$

The size of the packet at the CH is thus given by:

$$DRx^i_{CH} = \sum_{x=0}^{n^i_{avg_CHi}} 384 = 12000Bytes \approx 11.7MB \tag{11}$$

Finally, the received heterogeneous data is stored in a tensor and the first 10 frames are compressed using the following formula:

$$D_{ready} = \bigcup_{j=1}^{t} \left(\sum_{i=1}^{n_{CH}} DRx^{j}_{CH_i} \right) \tag{12}$$

By applying this formula to create the sending packet ready to be transmitted the data is compressed by 90%. The ratio between the size of the data received and the size of the data transmitted is 10:1. So the data sent by a CH to sink is,

$$DTx^{i}_{CH} = \frac{DRx^{i}_{CH}}{10} = \frac{11.7MB}{10} = 1.17MB$$

D^{i}_{sink} is the size of data received by the sink in round i.

$$DRx^{i}_{sink} = \sum_{x=0}^{n_{CH}} DTx^{x}_{CH} = \sum_{x=0}^{15} 1.17MB = 18.72MB \tag{13}$$

Therefore, the approximate data compression from the originally generated data at the end node is based on $DG^{i}_{node} = 625MB$ and $DRx^{i}_{sink} = 18.72MB$.

$$DataCompression = \left(\frac{DG^{i}_{node} - DRx^{i}_{sink}}{DG^{i}_{node}} \right) \times 100$$

$$\mathbf{DataCompression \approx 97\%} \tag{14}$$

7.2 Energy Consumption

The energy dissipation of SBIoT nodes is calculated two times in a round: one for normal nodes and one for CH nodes.

7.2.1 Energy consumption of normal node

$$E^{k+1}_{i} = E^{k}_{i} - E_{Tx} \tag{15}$$

Here E^{k}_{i} is the remaining energy of i^{th} node in a particular round k, and E_{Tx} is the energy dissipation in the transmission of the data packet [28].

$$E_{Tx} = (E_{elec} \times D_{pkt}) + \\ E_{amp} \times D_{pkt} \times (D(SN_i, CH_i))^2 \tag{16}$$

Here, E_{elec} is the Electronics energy dissipation per bit,
E_{amp} is the Amplification energy dissipation per bit,
$(D(SN_i, CH_i)$ is the Euclidean distance from sensor node SN_i to CH of node CH_i.

7.2.2 Energy consumption of CH node

$$E_{CH}^{k+1} = E_{CH}^{k} - \{(E_{Tx}) + (n \times E_{Rx})\}$$

Where E_{CH}^{k} is the remaining energy of CH in round k,
E_{Tx} is the energy dissipation in the transmission of data packets,
E_{Rx} is the energy dissipation in receiving the data packet from one node.

$$E_{Rx} = (E_{elec} + E_{DA}) \times D_{pkt}$$

$$E_{Tx} = \left[(E_{elec} + E_{DA}) + \left\{E_{amp} \times D(CH, hop_{next})^2\right\}\right] \\ \times (n \times D_{pkt})$$

Here, E_{elec} is the Electronics energy dissipation per bit,
E_{amp} is the Amplification energy dissipation per bit,

$D(CH, NextHop)$ is the Euclidean distance between the CH to the next hop (CH/Sink) in the shortest path towards the sink node [28].

7.3 CS for data gathering:

CS introduces a novel paradigm for compression and transmission of sparse signals or compressible signals in SBIoT. Consider the signal vector $S = (s_1, s_2, \ldots, s_n)^T$, which has $n \times 1$ elements. Let the signal S is represented as follows in a $n \times n$ orthogonal basis $\Psi = (\psi_1, \ldots, \psi_n)$:

$$S = \theta\Psi = \sum_{i=1}^{n} \theta_i \psi_i \tag{17}$$

where, in the basis Ψ, $\Theta = (\theta_1, \theta_2, \ldots, \theta_n)^T$ is the coefficient vector of S. Considering that no more than $k (k \ll n)$ nonzero elements are present in Θ, we call S is a k-sparse signal. Based on the theory of CS, the k-sparse signal S may be accurately reconstructed using $m = O(klog(n/k))$ random projections, with high probability, i.e. $x = \Phi x = \Phi\Psi\Theta$, here $\Phi\Psi$ must satisfy the restricted isometry property (RIP). It is proved that by solving a ℓ_1-minimization problem, the signal S may be reconstructed from z [10].

$$\min_{\Theta \in \Re^n} \parallel \Theta \parallel_{l_1} \; s.t. z = \Phi S, S = \Psi\Theta \tag{18}$$

However, in a real-world scenario the natural signals, such as temperature, humidity, light, etc. are not perfectly k-sparse as shown in equation 18. Because It is possible that it has k big transform coefficients and the rest are smaller. Moreover, the power-law degradation of these transform coefficients is common, these signals are known as power-law decay signals. Based on the theory of CS, high probability power-law decay signals may be ideally retrieved from $m = O(klog\frac{n}{k})$ random observations. The proof below (based on the work presented in [29]) highlights the efficacy of the proposed CS-based data gathering.

Theorem 1: If a random variable A_{ij} is derived from a random matrix A, then $\exists(x) \ni (|A_{ij}| \leq x)$ with high probability, and $A_{ij} \sim Sub(x^2)$, where x is constant.

Proof: Based on Markov inequality:

$$\mathbf{P}(A_{ij}^2 > x^2) < \frac{E(A_{ij}^2)}{x^2} \tag{19}$$

where x is constant. such that $x \leq 1$ because $A_i A_i^T = 1$. Since $E(A_{ij}^2) = (\frac{1}{n})$, we get

$$\mathbf{P}(A_{ij}^2 > x^2) < \frac{1}{x^2 n}$$

$$\text{Thus, } \mathbf{P}(|A_{ij}| \leq x) \geq 1 - \frac{1}{x^2 n} \tag{20}$$

So, if n is sufficient, then with the probability approaching 1, $|A_{ij}| \leq x$ is satisfied. If $|A_{ij}|$ is a bounded random variable with a zero mean, then $A_{ij} \sim Sub(x^2)$.

With this, the next objective is to prove that if $A_{ij} \sim Sub(x^2)$, then A satisfies the RIP, also the k-sparse signal from measurements m may be reconstructed.

Lemma 1: When $\delta \in (0, 1)$ is fixed. Let $A = \Phi\Psi$ is $m \times n$ random matrix having A_{ij} entries and satisfying $Sub(x^2)$. When $m = O(k \log(\frac{n}{k}))$ then $\forall(S)$: RIP has been satisfied by A with a high probability i.e.,

$(1-\delta)\|S\|_2^2 \leq \|A_S\|_2^2 \leq (1+\delta)\|S\|_2^2$, where S is a k-sparse signal in n-dimension space.

Proof: As demonstrated in *Theorem 1*, $A_{ji} \sim Sub(x^2)$, then, for any $S \in \mathbb{R}^N$, $< A_i, S > \sim Sub(x^2 \|S\|_2^2)$. Assuming that $Y = [Y_1, Y_2, \ldots, Y_m]$ and $E((< A_i, S >)^2) = \sigma^2$, where $Y_i =< A_i, S >$ for $i = 1, \ldots, m$, then,

$$E(\|Y\|_2^2) = \sum_{i=1}^{m} E\left(Y_i^2\right) = m\sigma^2 \tag{21}$$

If the number of data gathering rounds is more than the threshold, the probability that a random node v_i is elected a CH at round r is $\pi(v_i) = \deg(v_i)/2|\varepsilon|) = \Theta(1/n)$. Furthermore, the sub-Gaussian bounded arbitrary variable is denoted by Y_i. According to Markov's inequality,

$$\begin{aligned}
&\mathbf{P}(|Y|_2^2 \geq \beta m\sigma^2) = \\
&\mathbf{P}(\ \exp\left(|Y|_2^2\right) \geq \exp\left(\beta m\sigma^2\right)) \leq \frac{E((\exp(|Y|_2^2))}{\exp(\beta m\sigma^2)} \\
&= \frac{\sum\limits_{v_1=1}^{n} \sum\limits_{v_2 \neq v_1}^{n} \cdots \sum\limits_{v_m \neq v_1 \ldots v_{m-1}}^{n} \exp\left(Y_1^2\right) \ldots \exp(Y_m^2) P_{jr}}{\exp(\beta m\sigma^2)} \\
&= \frac{\sum\limits_{v_i=1}^{n} \pi(v_i) \exp(Y_1^2) \exp(Y_2^2) \ldots \exp(Y_m^2)}{\exp(\beta m\sigma^2)} \\
&\leq \frac{n \max(deg(v_i)) \exp(c_2^2 m)}{2|\varepsilon| \exp(\beta m\sigma^2)} \\
&= x_1 \exp\left(-m\left(\beta\sigma^2 - c_2^2\right)\right) \\
&= x_1 \exp\left(-m\sigma^2 X^*\right)
\end{aligned} \tag{22}$$

where $x1, x2$ are constants and $X^* = \beta - (x_2^2/\sigma^2)$. By normalization, $A' = \sqrt{(n/m)}(A_1, A_2, \ldots, A_m)^T$. Then, $E(\|Y\|_2^2)$ is computed as follows:

$$E\left(\|Y\|_2^2\right) = E\left(\sum_{i=1}^{m}\left(A_i^{'}S\right)^2\right)$$

$$= \sum_{i=1}^{m} E\left(\sum_{j=1}^{n}\left(A_{ij}^{'}S_j\right)\right)^2$$

$$= 2\sum_{i=1}^{m}\left(\left(A_{ij}^{\prime 2}\right)\|S\|_2^2\frac{E}{2}+\sum_{j=1}^{n}\sum_{k\neq j}\left(A_{ij}^{'}\right)E\left(A_{ik}^{'}\right)x_jx_k\right) \tag{23}$$

The final equation is based on the assumption that $E\left(A_{ij}^{'}\right) = \sqrt(n/m)E(A_{ij}) = 0$ and $E\left(A_{ij}^{'}\right)^2 = (n/m)E\left(A_{ij}^2\right) = (1/m)$ are true. As shown above, $E\left(\|Y\|_2^2\right) = m\sigma^2$. By putting $\beta = 1-\delta$, equation (22) could be written as

$$\mathbf{P}\left(|Y|_2^2 \geq E\left(|Y|_2^2\right)(1-\delta)\right) \leq x_1 e^{m\delta^2 X^*}$$

Then we have

$$\mathbf{P}\left(|AS|_2^2 \geq (1-\delta)|S|_2^2\right) \leq x_1 e^{m\delta^2 X^*}$$

Now, by putting $\alpha = 1 + \delta$,

$$\mathbf{P}\left(|AS|_2^2 \leq |S|_2^2(1+\delta)\right) \leq x_1 e^{m\delta^2 X^*} \tag{24}$$

There are (n,k) possible signal sets S for a k-sparse signal S. We get $(m,k) \leq (en/k)^k$ from Sterling's approximation. Therefore, we have the probability exceeding $1-2(en/k)^k$.

$$x_1 e^{-m\delta^2 X^*} = 1 - 2x_1 e^{-m\delta^2 X^* + klog(n/k)+k}$$

such that

$$(1-\delta)\|S\|_2^2 \leq \|S\|_2^2 \leq (1+\delta)\|S\|_2^2 \tag{25}$$

So, $m = O(klog(n/k))$ can be chosen with a probability approaching 1, for A satisfying the RIP.

8 Simulation Setup

A randomly deployed SBIoT network was simulated in python for collecting heterogeneous observations in an area of $\{200m \times 200m\}$. The performance of the proposed algorithm was evaluated on several parameters with two different sink positions. For a fair comparative analysis, the network parameters were replicated from the approaches used for comparative analysis. Heterogeneous random events were generated using the python library SimPy. The details of the simulation parameters are listed in Table 1.

Table 1 Simulation Parameters

Parameter	**Value**
The initial energy of a node	5 $Jule$
Number of nodes deployed	100-500
Number of sensors per node	5
Region of Interest (RoI)	200×200 $meter^2$
Sink node position (Scenario 1)	(250×250)
Sink node position (Scenario 2)	(1×1)
Transmission range	25, 30, 35, 40 $meters$
Data Packet size	100 $bytes$
E_{elec}	50 $nanojules$
E_{amp}	50 $nanojules$
E_{DA}	50 $nanojules$
Data transfer speed	250 $kbps$

9 Results

A comprehensive comparative analysis of the proposed method is presented under two different sink positions:

1. Sink outside the RoI.
2. Sink inside the RoI.

For both the scenarios, the parameters were replicated and results were considered from Data Gathering Algorithm Based on Spanning Tree (DGABT) [3], Random walk algorithm for data gathering (RANDOM) [4], Local Minimum Spanning Tree (LMST) [5], and Robust Data Collection Tree Construction (RDCT) [6]. For further validation, a comparative analysis with the scheme proposed in [7] MaxLeach, is also presented.

9.1 Sink outside the RoI (Scenario 1)

The proposed approach is evaluated on the following parameters:

1. **Network lifetime with varying transmission range:**
   The stability period (rounds of data collection until first dead) is considered as the network lifetime throughout this work. As evident from the Fig. 3, the proposed approach is able to achieve the longest stability period as compared to the other state of the art algorithms proposed in [3–6]. The varying transmission range hardly impacts on the network lifetime for the proposed approach because of uniform load distribution obtained by incorporating CS for data gathering. In addition to the uniform load distribution, the co-relation based clustering plays a significant role in optimizing the routing structure. Optimized cluster result in availability of close neighbour for routing the data, thereby making the impact of transmission range, minimal on the network lifetime. A similar observation can be made for the

other approaches, where a small improvement is seen in the network lifetime as the transmission range is increased. The improvement might be the result of reduced in-network transmission due to high transmission range of the nodes in consideration.

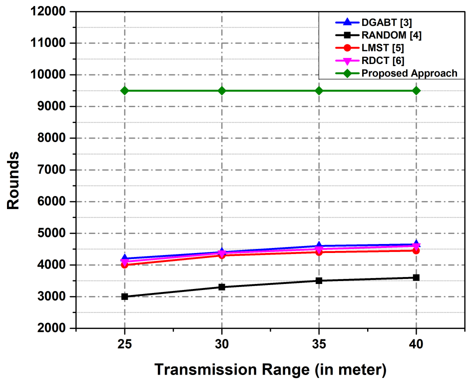

Fig. 3 Impact of transmission range on network lifetime (Scenario 1)

2. **Network lifetime with sink outside the RoI:**
   The Fig. 4 shows that the lifetime of the network, with the proposed approach, is highest as compared to the methods proposed in [3–6]. A careful observation of the facts makes this result a straightforward deduction, such as:
   (a) The proposed scheme uses a pseudo random sampling method to minimize the data load significantly.
   (b) Use of CS based data gathering allows the distribution of the load evenly over the network.
   (c) Use of ML optimizes the clustering and routing process which in term results in reduced energy consumption.
   (d) The use of tensor optimizes the storage overhead in the network significantly.
3. **Network lifetime with varying node density:**
   As the number of nodes in the network increases, not only the in-network transmissions increases, but also the data redundancy, because similar observations are reported by multiple nodes in close proximity of the event. Area remaining the same, increasing the node density affects the network lifetime adversely and the same can be seen in the Fig. 5. Although the proposed approach outperforms the other approaches proposed in [3–6], the impact of increased in-network transmissions and redundant observations is seen in the overall network lifetime where the first dead with 100 nodes is reported at around 9500 rounds of data collection and with 500 nodes the lifetime is reduced resulting in the first dead to be reported at around 8700 rounds of data collection.

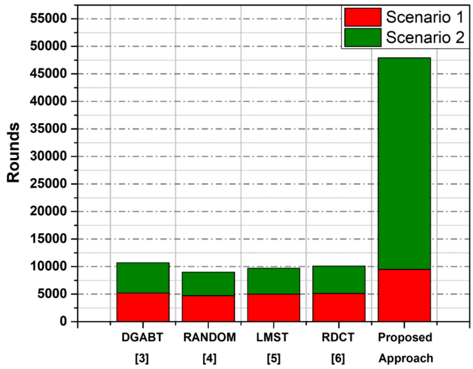

Fig. 4 Impact of sink location on network lifetime

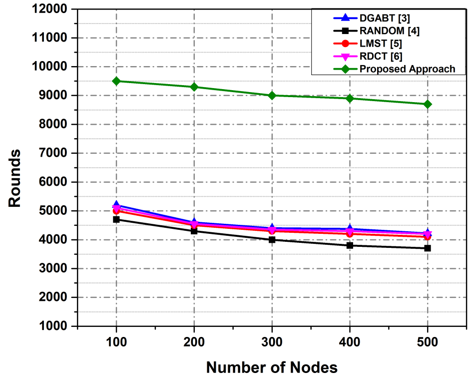

Fig. 5 Impact of Node density on lifetime (Scenario 1)

4. **Total energy consumption during lifetime with varying node density:**
   The proposed scheme is able to outperform the approaches proposed in [3–6] in terms of average energy consumption owing to the uniform load distribution, reduced in-network transmissions and reduced network overhead. As shown in the Fig. 6, while the average energy consumption of the proposed scheme is reported at 508 Jules with 500 nodes, the other methods struggle significantly to match the performance as the average energy consumption is reported at 700 in [3], 1300 in [4], 850 in [5], 700 in [6] for 500 nodes.

9.2 Sink inside the RoI (Scenario 2)

With the sink inside the RoI, the comparative analysis of the proposed scheme is presented on the following parameters:

1. **Network lifetime with varying transmission range:**
   The impact of sink position is clearly reflected in the network lifetime

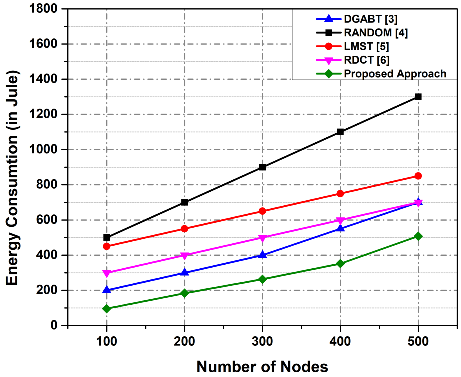

Fig. 6 Total energy consumption during lifetime (Scenario 1)

and is shown in Fig. 7 where the network lifetime is improved significantly to around 38000 rounds as compared to 9500 rounds when the sink is placed outside the RoI. Additionally, the network lifetime, in scenario 2, is not affected (almost) with varying transmission range as compared to results obtained for scenario 1. Both of these observations conclusively prove the effective uniform load distribution and reduced in-network transmissions in scenario 2 where there is a clear improvement because of the processing center being present in the center. Availability of neighbours and reduced transmission energy may be considered as the parameters directly responsible for this improvement.

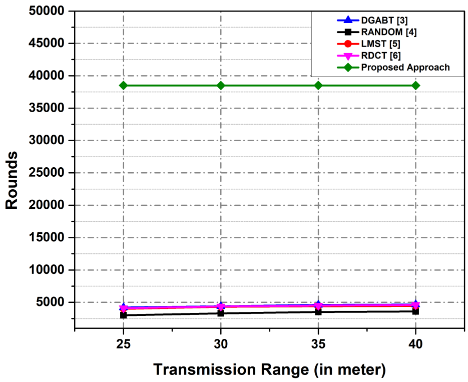

Fig. 7 Impact of Transfer range on network lifetime (Scenario 2)

2. **Network lifetime with sink inside the RoI:**
   Relocating the sink within the RoI allows significant reduction in the data transmission cost which is in addition to the uniform load distribution and effective clustering and CH election strategies of the proposed approach. The results, as shown in the Fig. 4, prove that the proposed scheme is able

to outperform the approaches proposed in [3–6] and achieves almost 7.5 times improved network lifetime.

3. **Network lifetime with varying node density:**
As the number of nodes are increased in the RoI, the probability of an event being reported by multiple nodes also increases. This in turn imposes overhead in terms of in-network transmissions and energy consumption. The same is reflected in Fig. 8 where a slight dip in the network lifetime is observed as the number of nodes is increased from 100 to 500. However the ML based adaptive CH election and CS based data gathering of the proposed scheme allows uniform load distribution and optimised routing resulting in significantly better network lifetime as compared to the approaches proposed in [3–6].

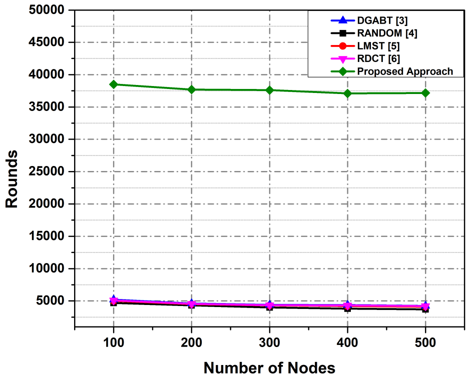

Fig. 8 Impact of Node density on lifetime (Scenario 2)

4. **Network lifetime with total energy consumption during lifetime with varying node density:**
A careful analysis of the results shown in Fig. 9 (scenario 2) and Fig. 6 (scenario 1) reveals the fact that the total energy consumption during the lifetime of the network is higher when the sink is placed within the RoI (scenario 2) as compared to the case when sink is placed outside the RoI (scenario 1). This result is a straightforward deduction from the fact that the first dead in scenario 2 is reported much later as compared to the first dead in scenario 1, thereby enabling more number of data collection rounds and hence higher energy consumption during during the lifetime.

9.3 Validation analysis:

To further validate the proposed scheme, a comparative analysis on two additional parameters is presented. The parameters from the approach in [3] are replicated and a detailed analysis is presented on the following to validate the claims of load distribution of the proposed method. The performance of the proposed algorithm is evaluated on two additional parameters which are:

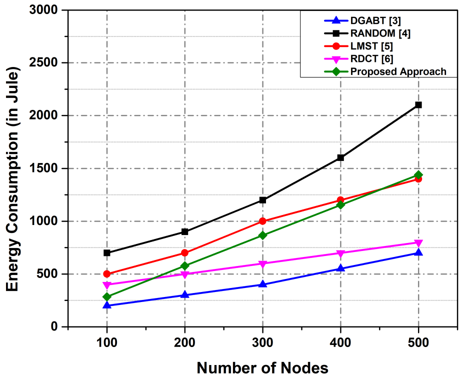

Fig. 9 Total energy consumption during lifetime (Scenario 2)

1. **Load distribution (Scenario 1 and 2):**
   As evident from Fig. 10, the proposed scheme is able to achieve better load distribution in case where the sink is placed within the RoI i.e. scenario 2. After the first node is reported dead at around 37500 rounds of data collection, a steep degradation is seen in the number of alive nodes. The steep degradation confirms the uniform load distribution claimed by the proposed approach. Although, the load is evenly distributed in scenario 1 as well, but the sink position being outside the RoI, imposes extra communication overhead over boundary nodes and the same is reflected in the results shown in Fig. 10. The figure proves the efficiency of the proposed scheme over the algorithm proposed in [7] where the first dead is reported only after 35000 rounds.

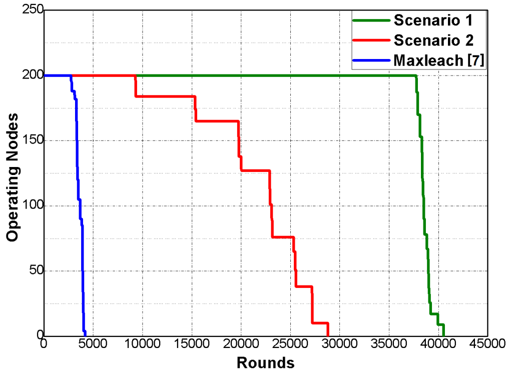

Fig. 10 Load distribution

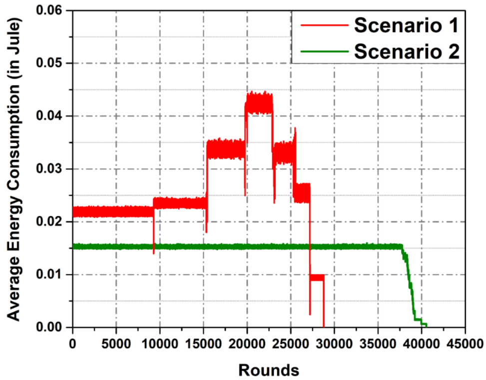

Fig. 11 Impact of Sink location on average energy consumption

2. **Average energy consumption (Scenario 1 and 2):**
   The average energy consumption per round for scenario 1 and scenario 2 is presented in Fig. 11. The uniform load distribution and increased network lifetime, achieved by the proposed scheme, is verified by the results of per node average energy consumption. The average per node energy consumption is uniform and even the highest value is at 0.015 μ Jule, which in turn becomes the reason of improved network lifetime. As evident from Fig. 11), the per node average energy consumption is uniform, validates the claim of uniform energy consumption by the proposed scheme.

A detailed summary of all the comparative outcomes is presented in table 2. A careful analysis of the table 2 validates that the proposed scheme is able to outperform the state-of-the-art methods proposed in [3–7].

Table 2 Comparative Analysis (Network size: 100, Area: 100x100, Trans. Range: 25 Meter)

Approach	**Network Lifetime (Round)**	**Energy Consumption (Jule)**	**First Node Dead**	**Middle Node Dead**
DGABT [3]	5200	200	5200	NA
RANDOM [4]	4700	700	4700	NA
LMST [5]	5000	500	5000	NA
RDCT [6]	5100	400	5100	NA
MaxLeach [7]	2754	NA	2754	3823
Proposed Approach	9422	122	9422	21817

10 Conclusion

An energy efficient approach for efficient heterogeneous data gathering in SBIoT applications is proposed in this work. Markov model, ML, and CS are used to optimise the classification, data transmission and in-network compression respectively, to enhance the performance of a SBIoT network. Detailed analytical and simulation results prove the dexterity of the proposed scheme over existing methods proposed in [3–7].

The use of ML and CS based data gathering allows significant reduction in the network overhead, resulting in 81% improvement in the network lifetime and 44% reduction in average energy consumption in scenario 1. Further, the proposed scheme is able to cope up with varying node density and transmission range and significantly outperforms the existing methods.

In future, the work may be extended to address the issues associated with mobile sink and security concerns imposed due to the same.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest (COI) Statement

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Kashani, M.H., Madanipour, M., Nikravan, M., Asghari, P., Mahdipour, E.: A systematic review of iot in healthcare: Applications, techniques, and trends. Journal of Network and Computer Applications **192**, 103164 (2021)

[2] Donoho, D.L.: Compressed sensing. IEEE Transactions on information theory **52**(4), 1289–1306 (2006)

[3] Dong, S., Sarem, M., Zhou, W.: Distributed data gathering algorithm based on spanning tree. IEEE Systems Journal **15**(1), 289–296 (2020)

[4] Zheng, H., Yang, F., Tian, X., Gan, X., Wang, X., Xiao, S.: Data gathering with compressive sensing in wireless sensor networks: A random walk based approach. IEEE Transactions on Parallel and Distributed Systems **26**(1), 35–44 (2014)

[5] Tan, H.O., Korpeoglu, I., Stojmenovi, I.: Computing localized power-efficient data aggregation trees for sensor networks. IEEE Transactions on Parallel and Distributed Systems **22**(3), 489–500 (2010)

[6] Liu, R.-S., Chen, Y.-C.: Robust data collection for energy-harvesting wireless sensor networks. Computer Networks **167**, 107025 (2020)

[7] Suman, J., Shyamala, K., Roja, G.: Improving network lifetime in wsn's based on maximum residual energy. In: 2021 2nd International Conference for Emerging Technology (INCET), pp. 1–5 (2021). IEEE

[8] Hülsmann, J., Traub, J., Markl, V.: Demand-based sensor data gathering with multi-query optimization. Proceedings of the VLDB Endowment **13**(12), 2801–2804 (2020)

[9] Jain, N., Bohara, V.A., Gupta, A.: ideg: Integrated data and energy gathering framework for practical wireless sensor networks using compressive sensing. IEEE Sensors Journal **19**(3), 1040–1051 (2018)

[10] Singh, V.K., Nathani, B., Kumar, M.: Weed-mc: Wavelet transform for energy efficient data gathering and matrix completion. IEEE Transactions on Parallel and Distributed Systems **31**(5), 1066–1073 (2019)

[11] Singh, V.K., Kumar, M., Verma, S.: Accurate detection of important events in wsns. IEEE Systems Journal **13**(1), 248–257 (2017)

[12] Shivhare, A., Singh, V.K., Kumar, M.: Anticomplementary triangles for efficient coverage in sensor network-based iot. IEEE Systems Journal **14**(4), 4854–4863 (2020)

[13] Kashani, M.H., Mahdipour, E.: Load balancing algorithms in fog computing: A systematic review. IEEE Transactions on Services Computing (2022)

[14] Velpula, P., Pamula, R.: Ebgo: an optimal load balancing algorithm, a solution for existing tribulation to balance the load efficiently on cloud servers. Multimedia Tools and Applications, 1–23 (2021)

[15] Haghi Kashani, M., Rahmani, A.M., Jafari Navimipour, N.: Quality of service-aware approaches in fog computing. International Journal of Communication Systems **33**(8), 4340 (2020)

[16] Kathuria, M., Gambhir, S.: Reliable packet transmission in wban with dynamic and optimized qos using multi-objective lion cooperative hunt optimizer. Multimedia Tools and Applications **80**(7), 10533–10576 (2021)

[17] Tommasi, F., De Luca, V., Melle, C.: Qos monitoring in real-time streaming overlays based on lock-free data structures. Multimedia Tools and Applications **80**(14), 20929–20970 (2021)

[18] Bhattacharjee, D., Acharya, T., Chakravarty, S.: Energy efficient data gathering in iot networks with heterogeneous traffic for remote area surveillance applications: a cross layer approach. IEEE Transactions on Green Communications and Networking **5**(3), 1165–1178 (2021)

[19] Ren, D., Li, X., Zhou, Z.: Energy-efficient sensory data gathering in iot networks with mobile edge computing. Peer-to-Peer Networking and Applications **14**(6), 3959–3970 (2021)

[20] Salim, A., Osamy, W., Aziz, A., Khedr, A.M.: Seedgt: Secure and energy efficient data gathering technique for iot applications based wsns. Journal of Network and Computer Applications **202**, 103353 (2022)

[21] Tomar, M.S., Shukla, P.K.: Energy efficient gravitational search algorithm and fuzzy based clustering with hop count based routing for wireless sensor network. Multimedia Tools and Applications **78**(19), 27849–27870 (2019)

[22] Mehta, D., Saxena, S.: Hierarchical wsn protocol with fuzzy multi-criteria clustering and bio-inspired energy-efficient routing (fmcb-er). Multimedia Tools and Applications, 1–34 (2020)

[23] Nayyar, A., Singh, R.: Ieemarp-a novel energy efficient multipath routing protocol based on ant colony optimization (aco) for dynamic sensor networks. Multimedia Tools and Applications **79**(47), 35221–35252 (2020)

[24] Kovtun, V., Izonin, I., Gregus, M.: Formalization of the metric of parameters for quality evaluation of the subject-system interaction session in the 5g-iot ecosystem. Alexandria Engineering Journal **61**(10), 7941–7952 (2022)

[25] Kovtun, V., Izonin, I., Greguš, M.: Model of information system communication in aggressive cyberspace: Reliability, functional safety, economics. IEEE Access **10**, 31494–31502 (2022)

[26] Kovtun, V., Izonin, I., Gregus, M.: The functional safety assessment of cyber-physical system operation process described by markov chain. Scientific Reports **12**(1), 1–13 (2022)

[27] Al-Hourani, A., Kandeepan, S., Lardner, S.: Optimal lap altitude for maximum coverage. IEEE Wireless Communications Letters **3**(6), 569–572 (2014)

[28] Singh, V.K., Kumar, M.: A compressed sensing approach to resolve the energy hole problem in large scale wsns. Wireless Personal Communications **99**(1), 185–201 (2018)

[29] Zheng, H., Guo, W., Xiong, N.: A kernel-based compressive sensing approach for mobile data gathering in wireless sensor network systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems **48**(12), 2315–2327 (2017)