A lifetime-enhancing cooperative data gathering and relaying algorithm for cluster-based wireless sensor networks

G Pius Agbulu1, G Joselin Retna Kumar1 and A Vimala Juliet2

Abstract
Despite unique energy-saving dispositions of cluster-based routing protocols, clustered wireless sensor networks with static sinks typically have problems of unbalanced energy consumptions, as the cluster head nodes around the sink are typically loaded with traffic from upper levels of clusters. This results in reduced lifetimes of the nodes and deterioration of other crucial performances. Meanwhile, it has been inferred from current literature that dedicated relay cooperation in cluster-based wireless sensor networks guarantees longer lifetime of the nodes and more improved performance. Therefore, to attain further enhanced performance among the current schemes, a lifetime-enhancing cooperative data gathering and relaying algorithm for cluster-based wireless sensor networks is proposed in this article. The proposed lifetime-enhancing cooperative data gathering and relaying algorithm shares the nodes into clusters using a hybrid K-means clustering algorithm that combines K-means clustering and Huffman coding algorithms. It makes full use of dedicated relay cooperative multi-hop communication with network coding mechanisms to achieve reduced data propagation cost from the various cluster sections to the central base station. The relay node selection is framed as a NP-hard problem, with regard to communication distances and residual energy metrics. Furthermore, to resolve the problem, a gradient descent algorithm is proposed. Simulation results endorse the proposed scheme to outperform related schemes in terms of latency, lifetime, and energy consumption and delivery rates.

Keywords
Cluster-based, cooperative, relay node, RLNC, LCDGRA, WSN

Date received: 11 June 2019; accepted: 17 December 2019
Handling Editor: Zhong Shen

Introduction
A wireless sensor network (WSN) is a collection of sensor nodes distributed within a particular area to observe some physical conditions and gather the observation at a base station (BS).1 The nodes are typically battery-operated, and they have inherent limitations in computation and communication capacities. In most monitoring applications, it is often pricey and unattainable to interchange the node’s batteries since the networks are formed in harsh regions with hundreds to thousands of nodes.2 As a result, the sensor nodes are expected to operate cooperatively for extended periods without any battery back-ups.3 Conventional WSN applications produce sizable quantities of data. Transporting such large amounts of data from the nodes to the BS will deplete the deficient energy reserves of the nodes timelier and reduce their lifetime.

1Department of Electronics and Instrumentation Engineering, SRM Institute of Science and Technology, Chennai, India
2Department of Electronics and Control/Instrumentation Engineering, SRM Institute of Science and Technology, Chennai, India

Corresponding author:
G Pius Agbulu, Department of Electronics and Instrumentation Engineering, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Kancheepuram, Chennai 603203, Tamil Nadu, India. Email: As5721@srmist.edu.in
Moreover, since nodes within a close range may sense identical events, it is incompetent to transport the sensed readings directly from the nodes to the BS.\textsuperscript{4} Therefore, it is very significant to consolidate the sensed observations from the sensor nodes into valuable information at the intermediate nodes to guarantee considerably reduced overheads of energy consumption and data redundancies. Clustering technology provides an effective method to reduce the communication energy consumption of the sensor nodes and enhance their lifetime.\textsuperscript{5–7} Clustering helps to simplify the network structure and evade direct transmission among the nodes and BS. It facilitates the self-organization of nodes into clusters by specific rules, such that each cluster head (CH) can fuse the data from their members and filter them to eradicate redundancies. This aggregation strategy reduces the traffic loads and communication energy consumption of the network.

The low-energy adaptive clustering hierarchy (LEACH) protocol is one of the well-known classical routing schemes based on clustering.\textsuperscript{8} LEACH protocol runs in two phases: setup phase and steady phase. During the setup phase, CHs are designated and made known to other nodes. In the steady phase, the nodes report their data to the heads in fixed slots. Although LEACH has some drawbacks such as poor CH node allocation, various improvements have been made on the LEACH protocol over the recent years.\textsuperscript{2,9} Typically, cluster-based sensor networks with static sink are faced with overheads of uneven power consumption and low success rates. This is because CH nodes around the sink are usually loaded with traffic, as they serve as relay nodes to transfer the data packets from the upper levels to the BS and thereby deplete their energy faster. Once a CH node dies at a fade zone surface, the other crucial performance metrics of the network such as throughput drop hastily. A possible solution to ensure success rate and longer lifetime of the nodes is by assigning the relay nodes discretely from various clusters. In this way, the CHs can have relay cooperation to transmit their accumulated packets to the destination and achieve longer lifetimes.

The scalable efficient clustering hierarchical (SEACH) routing is one among the classical cluster-based routing that is based on dedicated-relay cooperation.\textsuperscript{10} Though the data deliveries model of SEACH protocol is time-driven, which makes it more appropriate for WSN applications with periodic communication requirements. In the literature, SEACH protocol is revealed to encourage more energy efficient and scalable communication well above the other conventional time-driven cluster-based routing protocol. In general, relay cooperation in cluster-based WSNs facilitates improved total channel capacities and guarantees data deliveries from the cluster sections even when there is a loss of connectivity or failure of a CH node. It also allows more utilization of WSN broadcast nature and efficient cooperative communication with network coding (NC).\textsuperscript{11} Mainly, NC communications in WSN offer many benefits ranging from improved throughputs with reduced latency to enhanced network lifetime. It enables linear combinations of many packets from various nodes at a shared channel.\textsuperscript{2} At the receiver nodes, the coded packet can be recovered upon computations of a few logical or mathematical formulae.

However, one of the significant challenges of cooperative relaying techniques in WSN is the selection of the appropriate relay nodes.\textsuperscript{12,13} In WSN, relay communication typically increases the traffic activates of the relay nodes. This causes the relay or cooperator nodes to deplete their energy quicker than the other nodes.\textsuperscript{14} Once a relay node fails, a coverage hole appears in that area and other nodes may have no links to reach their destination. When the relay nodes are isolated or far apart, more energy is expended from long-distance communications. Moreover, NC operations in WSN necessitate good connectivity and extra energy resources at the coding links to successfully receive and process the packet flows from other shared mediums.\textsuperscript{15} Therefore, when the coverage and residual energy of the cooperator nodes or coding nodes are poor, the coding gain drops adversely.

Motivated by these aforementioned problems, this article proposes a lifetime-enhancing cooperative data gathering and relaying algorithm (LCDGRA) that combines full advantages of clustering technology with multi-hop cooperative relay and NC communication. The proposed LCDGRA operates in three phases and mainly focuses on resolving the routing problem of the cluster-based WSN through enhancing the power consumption and network lifetime, packet latency, and data receiving rate metrics.

Our new contributions in this article are the following:

1. We propose a new cluster-based event-driven scheme called LCDGRA. The proposed scheme can be applied in various event-driven monitoring applications to strengthen the routing performances in terms of network lifetime, energy efficiency, data delivery rate, and latency.
2. We propose a centralized hybrid-clustering scheme that combines K-means clustering and Huffman entropy coding design. The hybrid-clustering scheme is meant to secure the coverage and residual energy conditions of the CH nodes as well as minimize the energy consumption of the nodes at the setup phase.
3. We frame the relay node assignment as an NP-hard problem in terms of residual energy, and placements of the relay nodes. Also, to resolve the NP-hard problem, we propose a gradient descent heuristic-based algorithm. This scheme...
can be applied in various cluster-based networks to enhance the relay node selection and achieve higher transmission success rates.

Subsequent sections of this article are as follows: section “Related works” reviews some of the significantly related works. The network communication and energy models are discussed in section “Proposed system model.” Section “Design flow of LCDGRA” adequately describes the design phases of the proposed system. Section “Simulation results” presents and discusses the comparative performance evaluation results, while section “Conclusion” gives the closing comments.

Related works

In recent years, significant focus has been given to routing protocols in WSN due to the differences in network architectures and routing stipulations in various sensing applications. In WSN, routing is needed to establish reliable communication paths between nodes and their BSs, and it significantly influences the power consumption of the nodes. Recently, many energy-efficient cluster-based routing schemes have been applied in various WSNs.

The LEACH-centralized (LEACH-C) protocol is an improved version of LEACH. This scheme is meant to improve the CH allocation in LEACH by using a centralized control algorithm to distribute the CH nodes throughout the system. In LEACH-C, the BS computes the average residual energy of the nodes across the network and only allows the nodes with residual battery powers equivalent to the computed average residual energy to work as CH in each round. The LEACH with fixed clusters (LEACH-F) is another centralized variant of LEACH. In LEACH-F, the setup phase is not required at all rounds, only the CH position rotates among the nodes. The data fusion oriented clustered routing protocol based on LEACH (DF-LEACH) is another variant of LEACH protocol. DF-LEACH aims to facilitate both global energy efficiency and reduced energy dissipation at the CHs. In this scheme, data fusion is performed in a hop-by-hop mode among the CHs, before transmission to the BS.

The threshold sensitive energy-efficient sensor network (TEEN) protocol was proposed for WSNs as an event-driven cluster-based protocol. TEEN employs two limits: a soft and a hard threshold. The hard threshold is aimed to decrease total data transport. Therefore, this limit enables the nodes to communicate their sensed data only when their observations correspond to the attribute of interest events. The soft threshold is intended to decrease total data transport by enabling the nodes to disregard data deliveries when variations in their observations are minor, or there are no variations. The adaptive periodic TEEN (APTEEN) protocol is a hybrid alternative of TEEN that merges the benefits of reactive routing and proactive routing systems. In APTEEN, the nodes can react to changes in their monitoring environment and as well perform periodic data deliveries.

Roy and Das introduced the cluster-based event-driven routing protocol (CERP). In CERP, the nodes form clusters utilizing an event-based clustering system, whereby the clustering takes place only when there are event occurrences in the field. In the clustering phase, every node estimates a competing value concerning the transceiving and aggregation energy. Eventually, each node having the maximum competing value in the respective clusters is elected as CH and introduced to the members. The hybrid energy-efficient distributed clustering protocol (HEED) was proposed to support various time-driven WSNs by Younis and Filmy. In HEED, initial probabilities of sensor nodes to serve as CH depend on their residual energy and proximities to other neighbors.

The K-means and Dijkstra’s algorithm-based routing (KDUCR) protocol is an energy-aware routing scheme based on K-means clustering algorithm. The main aim of K-means clustering in WSN is to select random points in the sensing area and allocate the nodes to their nearest points to establish K-sections of cluster. The algorithm computes the centroid for each group repeatedly until the centers converge. This algorithm minimizes the intricacies associated with developing clusters at various levels in WSN. KDUCR protocol uses Dijkstra’s shortest path algorithm to determine the shortest routes with sufficient residual-energy to dispatch the aggregated data from each CH node in the sink.

The secured scalable energy-efficient clustering hierarchical (S-SEACH) is an improved version of SEACH. S-SEACH uses similar relay node assignment scheme as SEACH and mainly aims to achieve both improved energy efficiency and protection against intrusion. Recently, one time-driven cluster-based scheme with relay cooperation was proposed for WSNs by Wu et al. It is intended to satisfy the basic needs of network connectivity in large-scale WSN farmland services. Ahmed and Samreen proposed the cluster chain-based relay nodes assignment protocol (CCBRNA). It merges the advantages of relay cooperative communication with clustering and chain-based routing mechanisms.

The mode of linear NC was first recommended for WSN by Ahsledwe et al. The authors showed that a node could combine many packets of other nodes by logical or mathematical formulae. The coding mode is of two types, namely the intra-section NC and the intersection NC. The intra-section NC is intended to improve data delivery rates in lossy networks, while the
intersection NC is intended to improve the throughputs of lossless sensor networks. Meanwhile, both coding methods are described generally as binary (XOR) and random linear coding (RLNC). In the literature, there are specific recent works carried out on NC in WSN.29

Yin et al.28 proposed a multi-hop routing scheme that combines both NC and compressive sensing (CS) methods. It is proposed to facilitate energy-saving and reliable data reporting multi-hop WSNs. Sun et al.29 introduced the NC-WSN protocol to promote data delivery reliability, effective load balancing, and improved response time in various WSNs. Migabo et al.30 proposed the cooperative and adaptive NC for gradient-based routing (CoAdNC-GBR) protocol to support efficient NC and data communication in WSNs. CoAdNC-GBR estimates the total amount of nodes in the neighborhoods periodically to implement NC processes and uses the sink nodes’ address to deliver the coded packets. However, CoAdNC-GBR is more suitable for query-based applications since it considers an on-demand gradient-based routing system, in which queries from the data operator actuates the data communication phases of the network.

Hence, considering the works above, a new LCDGRA for cluster-based WSNs is proposed.

**Proposed system model**

**Network model**

In this study, we consider the sensor network model depicted in Figure 1. $G = (V, E)$ signifies the directed graph of the sensor network. $V$ signifies the vertexes, which include sets of nodes scattered randomly in the sensing area and a central BS placed at the end of the area. $E$ denotes the set of edges or links. Based on the functions of the different nodes in the sensing area, each node belongs to one of the following types: normal node (NN), a relay node (RN), and CH.

This study aims to develop an improved event-based clustering routing scheme, characterized by low latency, low-energy consumption, extended longevity, and high data receiving rates. Therefore, considering the sensor network model above, a new event-driven cluster-based routing scheme called LCDGRA is presented in this work. The proposed system is designed acknowledging existing WSN event-driven cluster-based routing design, including the assumptions as follows:

1. The BS and sensor nodes are static after placements.
2. There is only one BS in the network that is at the far end of the network.
3. An external source powers the BS.
4. The sensor network is uniform, and the initial energy of all nodes is the same.

**Energy consumption model**

Figure 2 shows the considered radio model in this work.2 Both free space $d^2$ and multi-path $d^4$ channel models are considered, including path loss $P(i,j)_{loss}$. The path loss is based on the communication distance $d$ and received signal strength (RSSI) of two nodes, $i$ and $j$, which can be defined as follows

$$P(i,j)_{loss} = \begin{cases} d^2, & d \leq d_o \\ d^4, & d > d_o \end{cases}$$

where $d_o$ is the communication threshold distance. For communication in range of the threshold distance $d \leq d_o$, the path loss is equivalent to $d^2$; else when $d > d_o$ the path loss corresponds to $d^4$. Meanwhile the threshold distance can be defined using equation (2)
do = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \quad (2)

where \( \varepsilon_{mp} \) and \( \varepsilon_{fs} \) are the receiving and transmitting amplifier’s attributes in free space and multipath communication. The energy required in transmitting \( k \) length of packets over a distance \( d \) is estimated using

\[
E_{TX}(k,d) = \begin{cases} 
 k \times E_{elec} + k \times \varepsilon_{fs} \times d^2, & d \leq d_o \\
 k \times E_{elec} + k \times \varepsilon_{mp} \times d^4, & d > d_o 
\end{cases}
\quad (3)
\]

where \( E_{elec} \) is the energy diffusion per \( k \) bit packet of the transmitter or receiver, which is usually influenced by digital coding, filtering, and modulation factors. The energy needed to receive the \( k \) bits quantity packets by the receiver is estimated using equation (4)

\[
E_{Rx}(k) = k \times E_{elec} \quad (4)
\]

The energy for transceiving and processing \( k \) quantity of packets per round is estimated as follows

\[
E_{Total} = \sum_{i=1}^{N} E_{RX}(k) + E_{DAN} + E_{TX}(k,d) \quad (5)
\]

where \( i = 1, 2, \ldots, N \), \( N \) is the total nodes involved in the communication round, and \( E_{DAN} \) is the energy required for aggregation and random linear coding of the packets per hop from the source region to the destination.

### Design flow of LCDGRA

This section describes the design methods of the proposed LCDGRA, which functions in three phases:

1. Network initialization and clustering
2. Relay nodes selection
3. Data aggregation and reporting

### Network initialization and clustering

In this phase, the network is initialized and the nodes are grouped into few clusters as described below.

**Network initialization.** Here the network is initialized. The central BS initiates this with initialization messages conveyed to all the nodes in the network. Then, all the nodes respond to this request with initialization responses. The response messages from the nodes contain their residual energy and current locations in the sensing region. Through the processes described above, the network is initialized.

**Clustering based on hybrid K-means.** In this point, the nodes are grouped into clusters. A hybrid K-means clustering system is proposed in this work to group the nodes into \( K \) number of clusters effectively. The hybrid approach merges the advantages of K-means clustering algorithm and Huffman coding algorithm. It is meant to optimize the sensor nodes’ transmission distances and energy usage during the clustering phase. Hence, the clustering and CH decisions are carried out by the central BS as described in the stages below.

**Stage 1: cluster formation.** In this stage, the nodes are divided into \( K \) units of clusters. Initially, the optimal total \( K \)-points of clusters is estimated as follows

\[
K_{OPT} = \sqrt{\frac{N}{2\pi}} \times A \times \sqrt{d_o^2} \quad (6)
\]

where \( D(N, S) \) is the Euclidian distances of the nodes (\( N \)) to the BS, \( d_o \) is the communication threshold as specified in equation (2), and \( A \) is the area of the network.

Once the total amount of \( K \)-points is computed, the nodes are assigned to their closest cluster-centroids. The distance between the nodes and the cluster centroids is expressed as follows

\[
D(j) = \sum_{i=1}^{N} \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \quad (7)
\]

where \( i = 1, 2, \ldots, N \), \( N \) is the total nodes, and \( X_i \) and \( Y_i \) are the \( X \) coordinates of the nodes and cluster centroids, respectively, while \( Y_i \) and \( Y_j \) are their \( Y \) coordinates.

Finally, new centroids are computed for every cluster until the centroid points become fixed.
Stage 2: appointment of CHs. In this stage, the CHs are elected. Initially, a competing value is calculated for each node in all the clusters as follows

\[ N_{\text{comp}} = \frac{E_{\text{resi}}}{\sum_{i=1}^{N} d_{\text{node}}^2 + k(E_{\text{rs}} + E_{\text{rx}}) + E_{\text{DAL}}} \]  

(8)

where \( i = 1, 2, \ldots, N \), \( N \) denotes the total members of that cluster, \( d_{\text{node}} \) is the distance from the node to the cluster members, \( E_{\text{resi}} \) is the residual of the node, \( E_{\text{rs}} \) is the energy needed for the reception of \( k \)-bit packets of the cluster members, \( E_{\text{rx}} \) is the energy needed to send the packets to the nearest relay nodes, and \( E_{\text{DAL}} \) is the required energy for in-network aggregation and random linear coding of the packets.

Once the contending value of every node is estimated, the value for each node is multiplied with a random value in the range of 0–1, to determine their respective probabilities. The probabilities obtained for every node is later summed-up to one and organized in descending form. After that, a code is formed for each of the nodes by the Huffman coding algorithm to evaluate their weights. Finally, the node that has the smallest weight in the different clusters is chosen as the CH node and declared to the cluster members.

The reader can get more information about the Huffman coding algorithm in the work of Mehfuz et al.32 In each round, different CHs are chosen in all clusters to facilitate load distribution until every node depletes within the sensing area.

Relay node selection

In this stage, the relay node selection takes place. Relay nodes are appointed in the various cluster regions to cooperatively help the CHs to deliver their aggregated data to the BS. Similar to the CH election, the election of relay nodes is carried out by the central BS. However, it is a familiar phenomenon in WSN that nodes utilize higher energy in data deliveries as opposed to receiving and processing. Hence, the central BS is bound to opt for relay nodes that have higher residual energy and shortest communication distances with adequate coverage. This study examines this condition as an NP-hard problem and proposes an effective gradient descent algorithm to solve this problem.

Gradient descent algorithm is an incremental first order iterative method intended to provide optimal weights that effectively reduce a cost function. The algorithm finds the local minimum weight by taking steps reciprocal to the negative gradient of the cost function, as it searches for the best weight. Typically, the weight update operation is given by \( W_N = W_E + (\alpha \times \Delta f/\Delta W) \) where \( f \) is the cost function, \( W_N \) and \( W_E \) are the new and existing sets of weight, \( W \) is the vector coefficient of the weights, and \( \alpha \) is the step length or learning rate. The step length needs to be selected carefully as a large value of step length, will overshoot the iteration. As well, when the step length is too small, the iteration to reach the local minimal will increase. The most preferred values are in the range of 0.01–0.3. Therefore, the step length is taken as 0.003, which is reasonably not too large or too small. More details about gradient descent optimization can be gotten from the works of Dong and Zhou.33

In the proposed system, the following are defined as the constrained functions for decent gradient optimization:

**Constraint** \( f_a \). The first constraint function \( f_a \) is defined as the ratio of average residual energy of the relay nodes to that of the normal nodes. Minimizing this constraint implies that high residual energy nodes are to become relay nodes. The average residual energy of the nodes can be determined using

\[ E_N = \frac{\sum_{i=1}^{N} E_{ri}}{N} \]  

(9)

where \( i = 1, 2, \ldots, N \), \( N \) is the number of nodes, and \( E_{ri} \) is the residual energy of each node under consideration. Furthermore, the average residual energy of the normal nodes is denoted with \( E_{\text{NN}} \), while that of the relay nodes is represented as \( E_{\text{RN}} \).

**Constraint** \( f_b \). The second constraint function \( f_b \) is defined as the ratio of the average distances of the normal nodes to the central BS, to that of the relay nodes to the BS. Minimizing this constraint implies that well-placed nodes with minimal communication distances to the central BS and good coverage, are to become relay nodes.

The average distance between the nodes and the central BS can be determined using

\[ D_N = \frac{\sum_{i=1}^{N} \sqrt{(X_{Ni} - X_{BS})^2 + (Y_{Ni} - Y_{BS})^2}}{N} \]  

(10)

where \( i = 1, 2, \ldots, N \), \( N \) is the total number of nodes, and \( X_{Ni} \) and \( Y_{Ni} \) are the \( X \) coordinates of the nodes and the central BS, while \( Y_{Ni} \) and \( Y_{BS} \) represent their \( Y \) coordinates, respectively. Likewise, the average distance from the normal nodes to the BS is represented by \( d_{\text{NN}} \), while \( d_{\text{RN}} \) is the average distance from the relay nodes to BS.

Hence, based on the above-constrained functions, as per \( f_a \) and \( f_b \), the cost function for optimization is formulated as follows

\[ f_{(\alpha, \beta)} = \beta \left( \frac{E_{\text{RN}}}{E_{\text{NN}}} \right) + \gamma \left( \frac{d_{\text{NN}}}{d_{\text{RN}}} \right) \]  

(11)
where \( f(a,b) \) is the cost function under minimization, and \( \beta \) and \( \gamma \) are control limits \( 0 \leq \beta \leq 1 \) and \( 0 \leq \gamma \leq 1 \), which are fixed usually in advance. Through minimizing the cost function expressed in equation (11), it is expected that the nodes with minimal path losses between the CHs, higher residual energy, shortest transmission distances, and adequate coverage are to become relay nodes. The nodes’ locations can be reached with the localization utilities available in work done by Luo and Chen.  

Algorithm 1 gives a sufficient description of the proposed relay node selection scheme. Once the algorithm is computed for the various nodes, the BS learns the optimal cost for electing the relay. Once it selects the relay nodes based on the above preferences, it notifies the selected nodes with a declaration message, which indicates their current functions as relay nodes. Afterwards, the relay nodes declare their information to the network with short notification messages, which has their residual energy, cluster identity (ID), relay node ID, and a header that reveals the message as a mere notification. Once each CH or relay node gets a join notification, it examines the message for membership suitability by analyzing the communication distance and signal strength of the sender. After the relay node or CH has confirmed the membership suitability, they notify the relay node of interest with a join message. The join message contains the ID of the preferred relay node and the sender ID that indicates the duty of the node as either relay node or a CH node; this is how the routing information that is needed for the cooperatively relaying operations is updated across the network. Once a relay node is about to ON or OFF its radio unit during the operation, it declares such transition to other relay and CH nodes having it as their relay members. This proposed relay node selection scheme can be applied in WSN to ensure transmission reliability and improvement in relay node allocation.

### Data aggregation and reporting

After the CH and relay nodes are elected, the nodes transit into idle states awaiting events. Besides, an event represents a variation in the sensed value that exceeds specific sensing thresholds. Followed by the event occurrence in the sensing field, the stages of data aggregation and reporting are performed. It is well known in WSN, that the amount of data communications has a significant influence on energy consumption of the network. Therefore, it is essential to reduce the number of deliveries to guarantee considerably less energy consumption. In the proposed scheme, to ensure minimal energy consumption, random linear coding is executed per hop from the source to the destination.

Once an event occurs, the cluster members of that particular region, report their sensed data to the CH node. Upon receiving the sensory data of the members, the CH node organizes the variations in the readings above specific threshold into \( N \) blocks of data packets \( P_i = [1,2,\ldots,P_N] \) based on their identities. After that, the packets are assigned coding vectors \( C_i = [1,2,\ldots,C_N] \) from Galois field \( (2^8) \). The packets are later coded together by blending them linearly as follows

\[
P_{\text{coded}} = \sum_{i=1}^{N} P_i C_i
\]

where \( i = 1,2,\ldots,N \), \( P_{\text{coded}} \) is the encoded packet, \( P_i \) denotes the source packets, and \( C_i \) denotes the coding vectors. After the coding process, the coded packet is sent from the source CH to the nearest hop relay nodes. Reconstruction of the coded packet \( P_{\text{coded}} \) into the original packets is based on the obtained packets at the destination. Initially, Gaussian eradication is carried out. The header message is setup into an \( n \times n \) matrix, and later into a reduced row-echelon (ref). Finally, the original packets are rebuilt by working out a few sequences of the underlying linear equations. Figure 3 shows the complete flowchart of the proposed LCDGRA.

### Simulation results

In this section, the performance of the proposed LCDGRA scheme is assessed using MATLAB 2018b simulations. The analysis was conducted with 100 nodes scattered randomly over an \((X=100 \text{ m}, Y=100 \text{ m})\) area including one sink node placed at \((X=100 \text{ m}, Y=50 \text{ m})\) far end of the area. Other key parameters applied in the simulation analysis are presented in Table 1.
In this work, an efficient hybrid K-means clustering scheme combining K-means algorithm with Huffman coding design has been applied to organize the nodes into K quantity of clusters. In addition, an efficient gradient descent heuristic-based algorithm has been applied to find the most appropriate relay nodes to serve the CHs report their aggregated data to the destination cooperatively. Figure 4 plots the convergence of the gradient decent cost function as per equation (11). It can be observed from Figure 4 that the cost function converges successively at about 31 iterations.

Figure 5 compares the latency of the network when using each of the schemes. The average latency presents the time expended from when data are disseminated from the sender to the arrival time at the destination. It includes delays in data queuing, data propagation, and processing. It is visible from the simulation results presented in Figure 5 that the latency of the proposed LCDGRA routing system is 18% lesser than that of CERP. It is also noticeable that the latency of CERP is 16% lesser than TEEN, when examining the network latency.

Figure 6 shows the data-receiving rate of the network. The packet-receiving rate represents the ratio of the sent quantity of data to the amount of data well received at the destination. It reveals the packet loss rates and link qualities of the sensor network. From the simulation results depicted in Figure 6, it is evident that the proposed scheme keeps a stable, even, and high receiving rate in the range of 1–0.98, which is 0.82–0.72.
higher than that of CERP protocol. For TEEN protocol, it is observed that the receiving rate drops entirely at a later stage of network operation, and the receiving rate is merely about 0.99-0.98, which is considerably less than that of LCDGRA and CERP.

Figure 7 analyzes the average energy consumption in the course of data aggregation. The average energy consumption in each round is estimated using equation (13), where $N$ denotes the total nodes, $E$ is the energy consumption per round, $R$ is the overall round, and $E_{\text{res}}$ is the nodes’ residual energy upon completion of each $R$th round. From the simulation result revealed in Figure 7, it is obvious that the energy consumption of the proposed LCDGRA is 21% less than that of CERP protocol, and 37% less than that of TEEN.

\[
E = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{E_{\text{res}}}{R} \right)
\]  

Figure 8 compares the lifetime of the network until half of the nodes deplete. The network lifetime is regarded as the period from the network initialization, to when half of the nodes deplete over the network. From Figure 8, it is clear that for the proposed...
LCDGRA, half of the nodes deplete around 3500, while in CERP and TEEN, half of the nodes depletes around 2600 and 2100, respectively.

The above simulation results confirm that the proposed LCDGRA truly merges the benefits of random linear coding and cooperative multi-hop communication with the best relay nodes in a cluster-based topology. Furthermore, the proposed LCDGRA outperforms both the protocols, TEEN and CERP, on the following rationales: in TEEN, the nodes wait for their assigned time slots to report their data to CH nodes. Thus, in scenarios where sensor nodes have no data to report, they are still constrained to waste their limited energy resources. Besides, the CH nodes also keep their radio units turned on always, to collect the sensed data of their cluster members. Therefore, during network operations, the CHs spend their energy quicker, which cause them to fail faster and thereby reducing the data receiving rates. In CERP, some of the nodes are isolated such that they have no links to other nodes due to variations in the clusters created at every round initiation. Therefore, the nodes fail faster, which creates more packet loss.

These are in contrast to the proposed LCDGRA, where the nodes are clustered adopting an efficient hybrid K-means clustering method that considers the communication distance and energy. Moreover, dedicated cooperating relay nodes that have good placements and energy resources consistently assist the CH nodes to communicate their aggregated data. In addition to that, the payload packets are coded at each hop from the source to the destination with random linear NC. These results prove that the proposed scheme does not only decrease sensor network latency and energy consumption but improves the throughput and longevity as well. Hence, it is attestable that the proposed scheme guarantees to meet the essentials for more energy efficiency, coherent and timely event reporting in WSNs.

**Conclusion**

In this article, a new event-driven cluster-based routing algorithm called LCDGRA was presented. The proposed LCDGRA routing method is simple and consists of three main phases. In Phase 1, the nodes are grouped into K clusters and the CHs are allocated in every cluster by a hybrid K-means clustering that combines both K-means clustering and Huffman coding algorithms. In Phase 2, rather than dedicating the relay nodes task to the CH nodes, relay nodes are assigned from the non-CH nodes to perform the data delivery tasks. Therefore, the CHs have sets of cooperating relay nodes that are dedicated to obtain and communicate their aggregated data to the destination. The relay node election is expressed as an NP-hard problem in studies of residual energies and communication distances of the nodes. Also, an efficient gradient descent heuristic-based scheme is proposed to solve the NP-hard challenge. In the final phase, the aggregated packets from the event region are coded with random linear coding and relayed in multi-hops to the central BS cooperatively.

The performance of the proposed LCDGRA has been assessed with other classical event-driven cluster-based protocol, including TEEN and CERP, through simulation. The simulation results confirm that the proposed LCDGRA significantly outperforms both TEEN and CERP routing protocols in terms of reduced energy consumption with prolonged lifetime and increased data delivery rates with reduced latency.

**Acknowledgements**

The authors thank the Directorate of Research of SRM Institute of Science and Technology, Chennai, for creating practicable settings to perform this research. The authors also extend profound gratitude to the reviewers of this article for their highly significant recommendations.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

G Pius Agbulu [https://orcid.org/0000-0002-4961-1721](https://orcid.org/0000-0002-4961-1721)
References

1. Yoo SE, Cheng AM, Chong PK, et al. Technological advances in wireless sensor networks enabling the diverse Internet of Things applications. *Int J Distr Sen Netw* 2018; 14(3): 155014771876322.

2. Nayak P and Devulapalli A. A fuzzy logic-based clustering algorithm for WSN to extend the network lifetime. *IEEE Sens J* 2016; 16(1): 137–144.

3. Lee HC, Ke H, Fang YM, et al. Open source wireless sensor system for long-term monitoring of slope movement. *IEEE Trans Instrum Meas* 2017; 66(4): 767–776.

4. Nikolic G, Nikolic T, Stojcev M, et al. Battery capacity estimation of a wireless sensor node. In: *IEEE microelectronics (MIEL) international conference*, Niš, Serbia, 9–11 October 2017. New York: IEEE.

5. Sasirekha S and Swaminathan S. Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network. *J Commun Netw* 2017; 19(4): 392–401.

6. Thakkar A and Kotecha K. Cluster head election for energy and delay constraint applications of wireless sensor network. *IEEE Sens J* 2014; 14(8): 2658–2664.

7. Kaur J, Randhawa S and Jain S. A novel energy efficient cluster head selection method for wireless sensor networks. *Int J Microw Wirel Tech* 2018; 8(2): 37–51.

8. Heinzelman WR, Chandrakasan A and Balakrishnan H. Energy-efficient communication protocol for wireless micro-sensor networks. In: *IEEE annual Hawaii international conference on system sciences*, Maui, HI, 4–7 January 2000. New York: IEEE.

9. Heinzelman WR, Chandrakasan A and Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans Wirel Commun* 2002; 1(4): 660–670.

10. Tarhani M, Kavian YS and Siavoshi S. SEECH: scalable energy efficient clustering hierarchy protocol in wireless sensor networks. *IEEE Sens J* 2014; 14(11): 3944–3954.

11. Ahlswede R, Cai N, Li SYR, et al. Network information flow. *IEEE Trans Inf Theory* 2000; 46(3): 1204–1216.

12. Amdouni I, Adjih C and Plesse T. Network coding in military wireless ad hoc and sensor networks: experimentation with GardiNet. In: *IEEE international conference on military communications and information systems*, Cracow, 18–19 May 2015. New York: IEEE.

13. Sahayaraj JM and Ganasekaran J. Relay node selection with energy efficient routing using hidden Markov model in wireless sensor networks. *Int J Netw Viri Org* 2018; 19(2–4): 176–186.

14. Nitesh K and Jana PK. Relay node placement algorithm in wireless sensor network. In: *IEEE conference on advance computing*, Gurgaon, India 21–22 February 2014. New York: IEEE.

15. Amanowicz M and Krygier J. On applicability of network coding technique for 6LoWPAN-based sensor networks. *Sensors* 2018; 18(6): 1718.

16. Bazzi HS, Haidar AM and Bilal A. Classification of routing protocols in wireless sensor network. In: *IEEE international conference on industrial technology*, Niagara Falls, ON, Canada, 17–19 March 2015. New York: IEEE.

17. Mehta D and Saxena S. A comparative analysis of energy efficient hierarchical routing protocols for wireless sensor networks. In: *25th ACM conference on computer and communications security*, Toronto, ON, Canada 15–19 October 2018. New York: ACM.

18. Chen L and Zhao B-H. Data fusion oriented routing protocol based on leach. *J Beijing Univ Posts and Telecommun* 2009; 32: 71–74.

19. Zhixiang D and Bensheng Q. Three-layered routing protocol for WSN based on LEACH algorithm. In: *IET conference on wireless and mobile sensor network*, Shanghai, China, 12–14 December 2017. London: IET.

20. Manjeshwar A and Agarwal DP. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: *IEEE proceedings of international parallel & distributed processing symposium*, San Francisco, CA, 23–27 April 2001. New York: IEEE.

21. Manjeshwar A and Agarwal DP. APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In: *IEEE proceedings of international parallel and distributed processing symposium*, Ft. Lauderdale, FL, 15–19 April 2002, pp. 195–202. New York: IEEE.

22. Roy S and Das AK. Cluster based event driven routing protocol (CERP) for wireless sensor network. *Int J Comput Appl* 2014; 88: 6–11.

23. Younis O and Filmy S. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans Mob Comput* 2004; 3(4): 366–379.

24. Qiong ZY and Rui LY. A routing protocol for wireless sensor networks using K-means and Dijkstra algorithm. *Int J Adv Media Commun* 2016; 6(2–4): 109–121.

25. Sandhya R and Sengottaiyan N. S—SEECH secured—scalable energy efficient clustering hierarchy protocol for wireless sensor network. In: *IEEE international conference data mining advanced computing*, Ernakulam, India, 6–18 March 2016. New York: IEEE.

26. Wu H, Zhu H and Miao Y. An energy efficient cluster-head rotation and relay node selection scheme for farm-land heterogeneous wireless sensor networks. *Wirel Pers Commun* 2018; 101(3): 1639–1655.

27. Ahmed G and Samreen A. Cluster chain based relay nodes assignment. *Int J Comput Trends Technol* 2014; 7(2): 106–114.

28. Yin J, Yang Y, Wang L, et al. A reliable data transmission scheme based on compressed sensing and network coding for multi-hop-relay wireless sensor networks. *J Comput Elect Eng* 2016; 56: 366–384.

29. Sun B, Gui C, Song Y, et al. A novel network coding and multi-path routing approach for wireless sensor network. *J Wirel Pers Commun* 2013; 77(1): 87–99.

30. Migabo ME, Olwal TO, Djouani K, et al. Cooperative and adaptive network coding for gradient-based routing in wireless sensor networks with multiple sinks. *J Comp Netw Commun* 2017; 2017: 5301462.

31. Gachhada O and Acharya N. K-means based energy aware clustering algorithm in wireless sensor network. *Int J Sci Eng Res* 2014; 5(5): 156–161.

32. Mehfuz S, Tiwari U, Rathore A, et al. A Huffman based lossless compression algorithm for wireless sensor networks. In: *IEEE international conference on innovative
applications of computational intelligence on power, energy and controls with their impact on humanity, Ghaziabad, India, 28–29 November 2014. New York: IEEE.

33. Dong X and Zhou ZD-X. Learning gradients by a gradient descent algorithm. *J Math Anal Appl* 2008; 341(2): 1018–1027.

34. Luo RC and Chen O. Mobile sensor node deployment and asynchronous power management for wireless sensor networks. *IEEE Trans Ind Elec* 2012; 59(5): 2377–2385.