Alharbi *et al. J Wireless Com Network* (2021) 2021:46
https://doi.org/10.1186/s13638-021-01911-9

EURASIP Journal on Wireless Communications and Networking

RESEARCH

Open Access

Towards improved clustering and routing protocol for wireless sensor networks

Mohammad Ali Alharbi[1*], Mario Kolberg[1] and Muhammad Zeeshan[2]

*Correspondence: maa@cs.stir.ac.uk
[1] Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK
Full list of author information is available at the end of the article

Abstract

Wireless sensor network (WSN)-based Internet of Things (IoT) applications suffer from issues including limited battery capacity, frequent disconnections due to multi-hop communication and a shorter transmission range. Clustering and routing are treated separately in different solutions and, therefore, efficient solutions in terms of energy consumption and network lifetime could not be provided. This work focuses data collection from IoT-nodes distributed in an area and connected through WSN. We address two interlinked issues,clustering and routing, for large-scale IoT-based WSN and propose an improved clustering and routing protocol to jointly solve both of these issues. Improved clustering and routing provide area-based clustering derived from the transmission range of network nodes. During process of clustering, cluster-heads are selected in such a way that provide fail-over-proof routing. An efficient routing path is achieved by finding the minimal hop-count with the availability of alternate routing paths. The results are compared with state-of-the-art benchmark protocols. Theoretical and simulation results demonstrate reliable network topology, improved network lifetime, efficient node density management and improved overall network capacity.

Keywords: Clustering, Internet of Things, Routing, Wireless sensor networks

1 Introduction

Wireless sensor networks face many challenges, including network lifetime, connectivity, security, synchronization and energy efficiency. Network partitioning (also called the hot-spot problem) is one such challenge arising because of the limited battery capacity of nodes. Initially, the concept of clustering is used as a solution to overcome this problem. In clustering, the process of cluster-head (CH) selection is important to increase network lifetime. However, for a wider area, multi-hop communication is required due to the limited transmission range of nodes [1, 2]. Most of the literature to date addresses routing and clustering as separate issues in WSNs [1, 3–8]. Very few protocols consider multi-hop routing [1, 2]. In single-hop transmissions, few protocols consider a specific percentage of nodes as CHs [3]. In many key approaches, CHs are selected randomly [3] and thus, can deployed close to each other. Since these CHs are in the transmission range of each other, network throughput decreases due to the carrier sense multiple access/collision avoidance (CSMA/CA) mechanism [4]. Treating both routing and clustering as a single and unified problem is a solution to this issue [1]. The solution should be scalable, employ an efficient mechanism for data collection and selection of CHs and minimize cluster heads' communication with sensor nodes and base stations.

The steps followed in many existing protocols include cluster identification, cluster-head selection, synchronization, steady-state phase, network topology and route discovery, data aggregation and data transmission. While all these phases consume energy [5], only a fraction of energy is consumed for data transmission [6]. Since WSN nodes are energy-constrained, these steps need refinement to increase the overall network lifetime [7]. Another limitation of existing hardware devices is the limited packet buffer size [8]. The data packets are buffered in the network queue of CHs and start dropping if the data size exceeds the buffer limit of CHs.

In WSN, nodes are deployed in large areas to collect data of interest. These nodes are short-range and battery-powered devices and thus, require multi-hop communication to send the collected data to BS. Clustering is one of the solution but is considered from the perspective of single hop communication where each CH can transmit data directly to BS. However, for larger areas, routing along with clustering is also desired. Failure of any CH in routing path result in failure in data transmission. This phenomena is called hot-spot problem. This is because clustering and routing, although considered but separately [1, 2]. We propose a clustering and routing solution based on fixed-areas of clusters in which during CH selection, routing topology with fail-over scenario is also considered.

The solution proposed in this work treats clustering and routing as a unified problem. Fixed-area-based clusters are generated based on the location and transmission range of nodes. After identifying the clusters and CHs, a routing mechanism is proposed with the provision of a handling fail-over scenario. This technique reduces the energy consumption overhead imposed due to fully dynamic clustering and the cluster-head selection process. The main contributions of this work are as follows:

1. An energy-efficient clustering and routing solution for hot-spot problems in WSNs. This solution is based on the location of wireless sensor nodes and employs a simplified CH selection process that selects a CH based on the cumulative weights of the residual battery and node connectivity.
2. We also propose a routing algorithm that handles fail-over scenarios by providing alternative routing paths to any selected cluster head.
3. The proposed solution is for large-scale WSNs and suitable for multihop communication among devices deployed in a wide area.

The remainder of this paper is organized as follows: Sect. 2 elaborates related works in the domain of WSNs, Sect. 3 elaborates the proposed improved clustering and routing protocol for IoT-based WSNs, Sect. 4 describes the results and analysis, Sect. 5 summarizes the discussion, and Sect. 6 concludes this work.

2 Related work

Hot-spots are a well-known problem in WSNs and are described as a situation when WSNs are partitioned due to the energy depletion of nodes, and the network does not remain connected [2]. The researchers used two different techniques to solve this problem. One group considered only clustering and CH selection with the single-hop transmission. The CH is selected based on the residual battery of the sensor node. These protocols assume a fixed transmission range for all sensor nodes, including the sink, which is not a realistic assumption for multi-hop wireless networks [3, 9]. Most of the existing studies also treat clustering and routing as separate issues and cause hot-spot problems due to unbalanced energy distributions among the nodes in WSNs [1, 2]. The authors in [1, 2] argue that routing and clustering are interlinked issues and must not be treated separately.

Many studies focus clustering alone to optimize the energy in WSN nodes, such as Low Energy Adaptive Hierarchical Clustering (LEACH) [9] and Balanced energy efficient network integrated super heterogeneous (BEENISH) protocols [10], which consider the residual battery of nodes only for CH selection. Authors in [10] compare their results with Distributed Energy-Efficient Clustering (DEEC) algorithm [11], Developed DEEC (DDEEC) [12], and Enhanced DDEEC (EDDEEC) [13] protocols. The results indicate that the BEENISH [10] protocol has the highest data transmission rate with the largest number of nodes alive during the round. However, this protocol assumes a uniform distribution of nodes without random placement and dynamic clustering.

CH is selected based on two different strategies: The residual energy of cluster members (CMs) and the rotation of cluster head membership periodically among CMs [2]. Selection of a CH at distance from BS is one of the major reason in early die-out of farther nodes from the BS [10]. This issues is addressed in LEACH-eXtended Message Passing (LEACH-XMP) [14] that used clustering-based technique which considered parameters including the density of nodes, the distance between nodes and the residual energy of nodes as CH selection criteria.

One of the variants of the LEACH protocol is the Orphan-LEACH (O-LEACH) protocol [15]. The O-LEACH protocol is developed based on the assumption that the LEACH protocol selects CHs randomly. Thus, some of the CMs no longer remain connected to their CH and become orphans. This protocol suggested the concept of intermediate gateway nodes that collect data from CMs and send data to CHs. However, the gateway nodes are selected based on the first-come, first serve-basis. Any node can be selected as a gateway without considering its residual battery. Compared with LEACH, O-LEACH provides better coverage and energy efficiency. However, a major limitation of the work is finding the information on orphan nodes. Data delivery delay and control overhead are also some issues that need to be resolved. Very few studies address clustering and routing as a single unified problem in WSNs [2]. For example, the JCR protocol [1] uses a back-off timer and gradient routing to develop a network topology for data collection in a large-scale WSN.

All the existing studies have attempted to solve hot-spot problems in different ways. Although different protocols, such as JCR [1], and O-LEACH [15], provided multi-hop routing solutions, the intermediate nodes between CHs are dynamically introduced, resulting in inefficient network energy consumption. Thus, clustering and routing must be addressed simultaneously as a single unified solution.

3 Method/experimental

3.1 Preliminaries

Consider that the sensing nodes and BS are deployed in the same sensing area. The aim of the sensing network is to collect sensing data from sensing nodes and send it to the BS. The network has the following properties:

- The nodes are organized into clusters where each cluster has a CH and all CMs can transmit data directly to the CH. The CH forwards the data to the BS via multiple hops.
- Every node has a minimum transmission range (Tr_{min}) and data transmission range (Tr_{d}). These ranges are defined based on the specification of the sensing device such as Telosb Tmote Sky platform [16].
- Throughput is defined as number of packets reaching to BS and originated from member nodes.

The following assumption are made:

- The nodes are aware of their physical location.
- Nodes are uniformly distributed and randomly located.

The first-order radio model is used for data transmission. In this model [1, 2, 15], energy spent for H-bit data transmission, $E_{\mathrm{tx}}(H, d)$, is given as:

$$E_{\mathrm{tx}}(H,\ d) = (E_{\mathrm{elec}} +\ E_d.\ d^e).H \tag{1}$$

where E_{elec} is the energy dissipated in the transmission circuit for a single bit, E_d is the single bit amplification energy, $E_d.\ d^e$ is the energy dissipated for a single bit transmission over a distance d, and e is the path loss exponent. The value of e is 2 for a free space and 4 for a multi-path space. Thus, the total energy dissipated for transmission of the H-bit packet is $E_{\mathrm{tx}}(H, d)$.

Similarly, if E_{rx} is the energy required to receive a single bit, then the energy required to receive the H-bit packet is:

$$E_{\mathrm{rx}}(H) = E_{\mathrm{elec}}.H \tag{2}$$

where E_{elec} for transmitter and receiver circuit is the same. The number of bits transmitted by CH after data compression and aggregation is [17]:

$$L_{\mathrm{CH}} = \alpha.\sum_{i=1}^{n} H_i + c \tag{3}$$

where L_{CH} is the total number of bits after applying compression and aggregation overhead at CH, n is the number of CMs, α is the compression ratio, and c is the aggregation overhead. In most of the cases [1, 15], the value of $\alpha = 0$ and $c = 1$ are the default settings.

3.2 Problem statement

As discussed in the related work section, previous studies did not consider the routing path selection during the process of CH selection. This leads to inefficient CH selection, routing topology, and frequent disconnections. It is observed that increasing the transmission range of nodes from 20 to 70 m decreases the number of clusters from 58 to 8 [1]. The LEACH protocol [9] does not adopt any routing mechanism. Therefore, a greedy routing approach is used where the current CH finds another CH with a distance lesser than its distance from the CH. The newly selected CH in the routing path acts as the relay node. However, it is found through various examples that the length of edge (LEH) between the CHs is greater than the maximum transmission range of nodes.

From LEACH [9] and JCR [1], some significant results are obtained:

- The average LEH among CHs is greater than the transmission range in the case of LEACH. This is due to variable inter-cluster topologies generated due to random CH selection.
- The LEH and transmission range are strongly related. Maximum LEH should be smaller than the transmission range. However, maximum network lifetime cannot be achieved if transmission range is not large enough.

According to the literature review, few works have considered clustering and CH selection along with routing [3, 9]. In this paper, the goal is to develop a protocol that generates interconnected and reliable network topology that ensures increased network-lifetime.

3.3 Improved Clustering and Routing (ICR) Protocol

Improved Clustering and Routing (ICR) Protocol is based on distributed responsibilities of BS and CHs. The BS is responsible for the initial sub-area distribution and defining the forward, backward and equal nodes of every node based in the network. CHs are responsible for data collection, data transmission to the BS and the distribution of TDMA schedule (time-division multiple access protocol) to the nodes in the clusters. The following sections provide detail on this approach.

The setup phase in ICR is the process of clustering and CH/CM selection. All the nodes in a cluster act as candidate nodes to be selected as the CH. In the data transmission phase, every CH collects data from member nodes, aggregates it, and transmits it to the BS via multi-hop communication.

The goal of ICR is to perform clustering and CH selection in such a way that network lifetime is increased through the efficient backbone of inter-cluster routing. The basic idea of fixed-area-based clustering (FAC) is adopted for simplified clustering, CH selection process, and inter-cluster routing among CHs. For inter-cluster routing, every node identifies its forward and backward nodes with reference to the BS. This information is used during CH selection in a CH selection process and routing discovery.

The division of a whole area based on minimum transmission range, ($Tr_{\min}$), into small sub-areas and sub-regions is shown in Fig. 1. For a given node (x), the forward (u), backward (v), and equal (m) nodes are provided with the data transmission range (Tr_d). Each node has two transmission ranges, which are clustering range ($Tr_{\min}$) and data transmission range (Tr_d), such that $Tr_{\min} < Tr_d$.

3.3.1 Clustering area distribution

To realize the integrated design of clustering and routing, it is important to clarify few definitions:

- **Definition 1** *(Fixed Area (A))* the given minimum transmission range ($Tr_{\min}$) of the node, fixed area can be defined as the area where $Width = Height = (Tr_{\min})$. This is the area that acts as a clustering area for all nodes inside it.

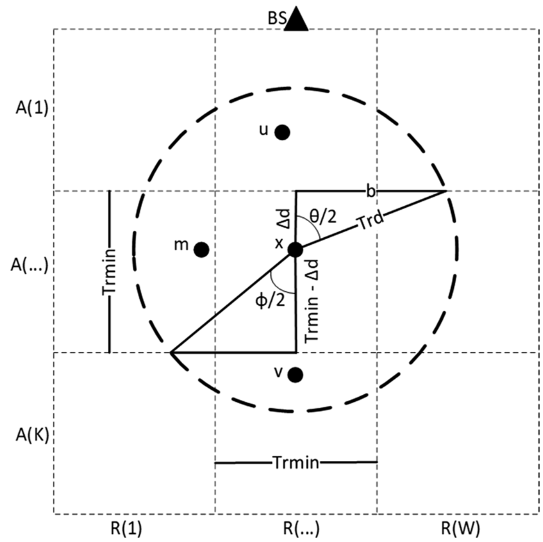

Fig. 1 ICR protocol description

- **Definition 2** *(Cluster)* the set of nodes $C(z)$ having the same cluster (K) can be denoted as follows:

$$C(z) = x : A(x) = K \cap x : R(x) = W; x \in S \tag{4}$$

where $A(x)$ is the fixed Area, $R(x)$ is region W of sensing node x in the area K, and S is all the sensor nodes.

- **Definition 3** *(Forward and Backward Nodes)* Given the Area $A(K)$ of node (x), the forward nodes, $F(x)$, are from the Area ($K-1$) within the data transmission range (Tr_d), and Backward nodes, $B(x)$, are from the Area ($K+1$) with nodes within the data transmission range (Tr_d).

$$F(x) = x : distance(x, u) < Tr_d \cap A(u) = K - 1 \tag{5}$$

$$B(x) = x : distance(x, v) < Tr_d \cap A(v) = K + 1 \tag{6}$$

- **Definition 4** *(Equal Nodes)* The nodes within the same $Area(K)$ such that:

$$E(x) = x : distance(x, m) < Tr_d \cap A(m) = K \cap \forall R \in A(K) \tag{7}$$

In ICR, purpose of fixed-area selection is to minimize the time for clustering, CH selection, and finding forward and backward nodes. This information is useful not only during the CH selection process but also in developing the robust network topology.

Initial values of area A, forward nodes F, and backward nodes B are set to zero (Fig. 1). The BS calculates the first hop neighbors by sending a message to Area (1) nodes within Tr_d. Upon receiving the message, the Area (1) nodes set the BS as the forward nodes and broadcast the message within its Tr_d. On receiving this message from Area (1), Area (2) nodes set Area (1) nodes as the forward nodes and broadcast the message to Area (3). This process stops at the K_{th} area (last area) where no backward nodes are identified.

A node may receive more than one broadcast message from the $K - 1$ area. Upon receiving the first message in the K_{th} area, the node sets its forward node ID, and upon receiving the second forward node and third forward node, it records the forwards node's list.

Once the broadcast messages are complete in sending messages to all (K) areas, the BS initiates a 'TEMP' message asking all the nodes in the network if they have set up their forward or backward nodes. If nodes have received, it will send an 'ACK' message, else it will send an 'ERR' message. The process of network topology is a one time activity where each node establishes its forward and backward nodes. However, in a specific scenario where forward node is died, the equal node (CH) can be selected as forward node in an attempt to find data transmission path.

The ICR protocol is proposed based on small and fixed transmission ranges of IoT devices, as specified in their technical specifications. The whole area is divided into equal-sized sub-areas based on the transmission range. Many existing protocols select CHs irrespective of their positions [15]. ICR restricts selection of CH within a fixed-area.

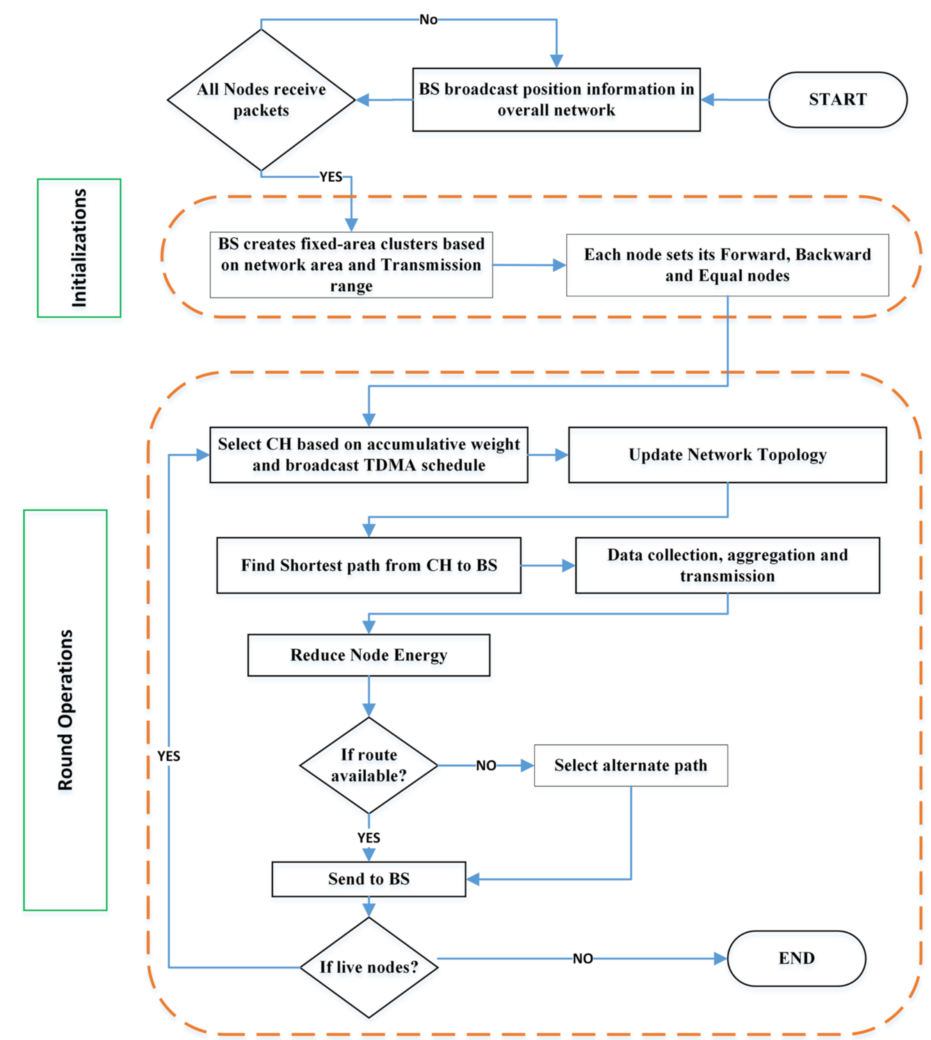

Fig. 2 Flow chart of ICR protocol

Table 1 Variables and definitions for ICR protocol

Variable	Definitions
BSPosition	Position of sink node
BS	Base station responsible for data collection
NetArea	Network area size
EInit	Initial energy of nodes
Nd	Node
NumNd	Total number of nodes
N_F	Number of forward nodes
N_B	Number of backward nodes
NdPosition	Position of nodes
Round	Current round of simulation
MaxRound	Maximum number of rounds
NdLive	Number of alive nodes in network
SubAreasList	List of sub-areas based on transmission range of node
ThrshdDist	Threshold distance
CHList	List of CHs
CH	Cluster head node of a cluster
CM	Member node of a cluster
EDissip	Energy dissipation
SNd	Source node that sends packets
TNd	Target node that receives packets
T	Back-off timer of a node
Nc	Number of clusters in the area based on transmission range
CHSelect	Selection of CH
NdWeights	The accumulative weight of node to be candidate for CH
WeightList	Sorted list of nodes based on accumulative weights inside a cluster
EConsumed	Energy consumed

The generic algorithm for the ICR protocol covering energy dissipation, clustering, and CH selection is provided as Algorithm 1 (graphical representation or flow graph is given in Fig. 2). Details of variables and their definitions are provided in Table 1.

Algorithm 1 ICR Protocol

```
Procedure Main ( )
BSPosition : Set BS position (x,y)
NetArea : Set area size
NumNd : Set number of nodes
EInit : Initial node energy in Joules
NdPosition ← Node position (x,y)
Round : Initialize with zero
MaxRound : Set max rounds
NdLive : Total number of live nodes
SubAreasList ← Clustering(NetArea, TrRange)
N_F, N_B, N_E : BS set forward, backward and equal nodes of each node
ThrshdDist : Set threshold distance
while Round < MaxRound do
    if (NdLive == 0) then
        Exit
    end if
    for SubArea in SubAreasList do
        CH ← CHSelect(SubArea)
        CH set TDMA schedule and send to CMs
        NetTopology ← update network backbone with new CH
    end for
    For each CH, calculate the shortest path from the CH to BS using Dijkstra algorithm
    for SNd and TNd in SubArea do
        Select SNd and TNd from NetTopology
        EInit ← EDissip(SNd, TNd)
    end for
    Aggregate data at the CH and send to the BS using multi-hop
    EInit ← EDissip(CH, N_F)
    Round ← Round + 1
end while
return Round
End Procedure
Procedure Clustering (NetArea, TrRange)
N_c ← calculate number of clusters based on NetArea and TrRange
SubAreasList ← list of sub-areas based on Nc
return SubAreasList
End Procedure
Procedure CHselect (SubArea)
SubNdLive in SubArea
for Nd = 1 to SubNdLive do
    NdWeight ← calculate node weight using Eq. (8)
    Set back-off timer T of Nd based on NdWeight
end for
return CH ← Nd with first expiring the timer and having N_F, N_B and N_E
End procedure
Procedure EDissip (SNd, TNd)
Dist ← calculate distance between SNd and TNd
if (ThrshdDist > Dist) then
    EConsumed ← energy for free space data transmission Eq. (1)
else
    EConsumed ← energy for multi-path data transmission Eq. (1)
end if
EConsumed ← EConsumed + energy for data receiving Eq. (2)
EInit ← EInit - EConsumed from SNd and TNd
return EInit
End Procedure
```

The proposed algorithm performs clustering and routing processes together is such a way to maximize the network performance and lifetime. First of all, In phase 1, BS broadcast position information to overall network. All nodes receiving this information send back response to BS about their position. In phase 2, BS divides the whole network area based on transmission range of nodes. The CH of each area is selected within the area based of network degree and residual battery of nodes. In CH selection process, forward, backward and equal nodes are identified that will later used in topology generation. CH is selected based on accumulative weight and broadcast its schedule to member nodes followed by network topology update. Next, shortest path is identified from source CH to BS for traffic routing. In next phase, data transmission and energy dissipation starts along with alternate routing path selection procedure. Once one data transmission round is completed, the process of phase 2 is started again. This process continues until all network nodes die.

In the toy example provided in Fig. 3, there are 200 nodes uniformly distributed and randomly located in an area of 60 m × 60 m. All the nodes are assumed to have homogeneous hardware. It is also assumed that the BS has no battery power restriction.

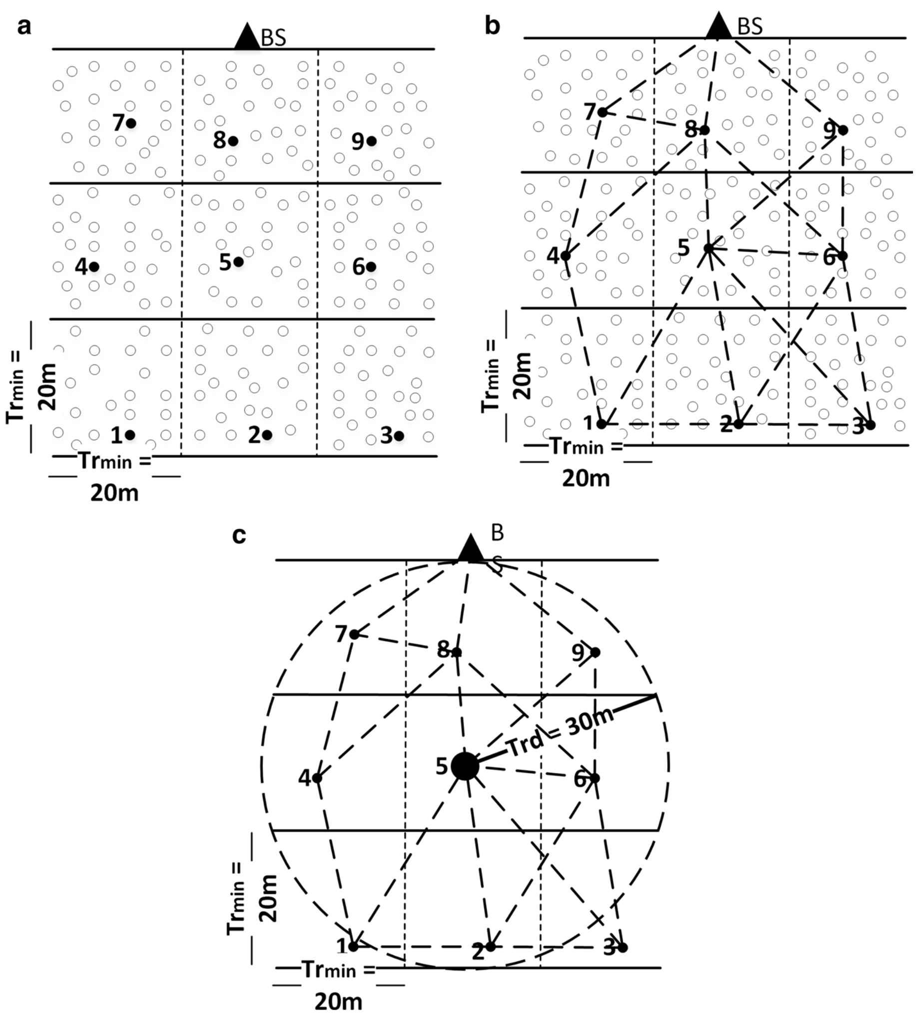

Fig. 3 ICR model phases. **a** Fixed-area clustering based on minimal transmission range. **b** Forward, Backward and Equal node selection. **c** Cluster-head selection based on accumulative weight and forward and backward nodes

In the first step, the BS performs the clustering of nodes by dividing the whole area into sub-areas based on the minimum transmission range (Fig. 3a). All the nodes within a sub-area are considered part of one cluster. This is a one-time process and, thus, saves energy required for repeated clustering and message passing.

3.3.2 Cluster-head selection

Cluster head selection is done based on two parameters: node degree and residual battery of nodes.The equation for accumulative weight, W_c, of a node is as follows:

$$W_c = \beta.\left(\frac{\xi_{\max} - \xi}{\xi_{\max}}\right) + \gamma.\left(\frac{\zeta_{\max} - \zeta}{\zeta_{\max}}\right) \tag{8}$$

where $\beta + \gamma = 1$, both variable represent weighing proportion of residual battery of nodes and connectivity of nodes. Any combination of values can be selected. However we give 0.5 proportion to both weighting parameters. The reason is to give equal importance to network topology generation along with clustering that is major research question of the article. β is the weight of residual battery and γ is the weight of node degree (the number of connected nodes to the node). ξ is the residual battery of node, and ζ is the degree of node, which decreases after every data transmission round due to the die-out of connected nodes. Also, $\xi_{\max}$ is the maximum battery, and $\zeta_{\max}$ is the maximum degree a node.

Let W_m be the maximum accumulative weight any node can have. Then, it can be defined as follows:

$$W_m = (\beta.\xi_{\max} + \gamma.\zeta_{\max}) \tag{9}$$

The value of $\zeta_{\max}$ for uniform node distribution can be computed as follows:

$$\zeta_{\max} = \rho\pi Tr_d^2 \tag{10}$$

Equation (10) represents number of nodes connected to the given node with radius Tr_d and node density ρ.

T is the time set for every node in a cluster with value (W_c). The node with the highest value of residual battery and degree will get its timer, T, zero first and sends an advertisement message (ADV) in the minimum transmission range. This ADV message contains (CH-ID, Area ID, and Region ID). On receiving the message, all candidate nodes will stop their timer and send back an 'ACK' message with (Node ID, CH-ID, Area ID and Region ID). On receiving the message 'ACK' from cluster members, the CH broadcasts a TDMA schedule containing nodes' IDs and their time slot. On receiving this message, the CM sleeps until its time slot.

The second step is the selection of forward and backward nodes of all the nodes. The BS broadcasts a message to all nodes for the selection of forward and backward nodes (Fig. 3b).

In the third step, the CH is selected based on accumulative weight of node and its connection with forward and backward nodes (Fig. 3c). Network topology is generated among CHs based on the data transmission range considering the load balancing and fail over scenarios along with a minimal number of hops to the sink node. Load balancing is

Table 2 Initialization of network topology: step-1

ID	Forward	Backward	Equal	Hop-count
7	BS	4	8	1
8	BS	4, 5, 6	7	1
9	BS	5,6	Nil	1

provided by considering alternate routes for data transmission in the case of network congestion. The same alternate route can serve as the main route when the primary link is down from the CH to the BS. This is how the fail over scenario is addressed.

3.3.3 Routing

The routing process is initiated by the BS, which broadcasts a route update message to all nodes in all regions of Area 1. On receiving the messages, all the nodes in Area (1) set the BS as its forward node and broadcast the message in their data transmission range Tr_d. The nodes in Area(2), on receiving the messages from nodes in Area(1), set these nodes as forward nodes. Area (1) nodes send back the 'ACK' message in Tr_d and set Area(2) nodes as backward nodes. This process continues until all the nodes in different areas set their forward and backward nodes. During the selection for forward node, the preference is given to nodes with the highest Euclidean distance for covering the maximum distance in a single hop [18]. It is important to note that the clustering regions are formed based on the minimum transmission range $Tr_{\min}$, and data is transmitted based on the data transmission range Tr_d of nodes.

A simple topology generated by the BS for the network is shown in Fig. 3b. In the first step, the one-hop neighbor is initiated by the BS. This includes CH nodes 7, 8 and 9. Each CH maintains a routing table indicating its forward and backward nodes. The cumulative routing table is given in Table 2. In the next phase (Fig. 3b), each node (7, 8, 9) sends a route update message to backward CH nodes (4, 5, 6) in Table 3. Similarly, in Table 4, CH nodes (4, 5, 6) send route update messages to backward nodes (1, 2, 3). Each node maintains its routing table for forward and backward nodes. It sends an update message in case of a change in CH in any region. In case of failure of node 4 and 5, node 1 selects node 2 as forwarding node to handle fail-over scenario.

Table 3 Initialization of the network topology: step-2

ID	Forward	Backward	Equal	Hop-count
4	7, 8	1	Nil	1
5	8, 9	1, 2, 3	6	1
6	8, 9	2, 3	5	1

Table 4 Initialization of the network topology: step-3

ID	Forward	Backward	Equal	Hop-count
1	4, 5	Nil	2	1
2	5, 6	Nil	1, 3	1
3	5, 6	Nil	2	1

After the CH selection and network topology generation, route discovery and selection are performed in the last step. The objective of this step is to provide an optimal path for data transmission from source to sink, considering load balancing, fail-over, and energy-efficient scenarios. During selection, the path with the least energy consumption is preferred. From source to sink, intermediate CHs act as relay nodes. We assume relaying energy as zero to achieve accurate and general results that consider all the different types of motes.

3.3.4 Cluster head probabilities

According to conditions defined in the algorithm of ICR, the CH will be selected based on two conditions when its timer expires. These are as follows:

- The $node(x)$ must have at least one forward $node(u)$. This $node(x)$ acts as the backward node of $node(u)$ as well. For $node(u)$ to be selected as the forward node of $node(x)$, the probability P is

$$\prod_{u} \left[1 - P(x,u)\right] \cap u \in F(x) \tag{11}$$

And for $node(x)$ to be the only CH that forwards data to $node(u)$, the probability can be written as:

$$P_{\text{CH}}(x) = \left[1 - \prod_{u}(1 - P(x,u))\right] \cap u \in F(x), A(x) \neq A(u) \tag{12}$$

Where $P(x,u) = P_{\text{CH}}(u).P(F(x))$. This equation explain that if node (x) has to be selected as a CH then it should have a forwarding node (u) that should also be a CH node. This is required condition for network topology generation.

- If there is no backward node of $node(x)$, it means that the area of $node(x)$ is the last area (Kth area), then the equation of $node(x)$ to be CH can be written as follows:

$$P_{\text{CH}}(x) = P(x,m) \cap N_B(x) = 0 \cap N_F(x) \neq 0 \tag{13}$$

where m is the equal node of $node(x)$ within same area but different region.

The Eqs. 12 and 13 represent $node(x)$ to be the CH, where $N_B(x)$ and $N_F(x)$ are the number of backward and forward nodes of $node(x)$.

3.3.5 Number of forward and backward nodes

The number of forward nodes of $node(x)$ can be determined by applying simple trigonometric operations. Let Tr_d be the radius of the transmission range, $\triangle d(x)$ is the distance between node and boundary of the area, and $\theta/2$ is the angle. Then,

$$\theta/2 = \arccos \frac{\triangle d(x)}{Tr_d} \tag{14}$$

The area of the whole sector will be:

$$A(x) = \theta/2 \times {Tr_d}^2 \tag{15}$$

Subtracting the area of two triangles from the whole sector area gives the area of the forward nodes as follows:

$$A(F(x)) = A(x) - b.\bigtriangleup d(x)\sqrt{{Tr_d}^2 - \bigtriangleup d(x)^2} \tag{16}$$

where b is the perpendicular of the right-angle triangle (Fig. 1). The number of forward nodes of *node*(x) can be calculated as:

$$N_F(x) = \rho.A(F(x)) \tag{17}$$

where ρ is the number of nodes per unit area. Similarly, for calculating the number of backward nodes of *node*(x),

$$\phi/2 = \arccos \frac{Tr_{\min} - \bigtriangleup d(x)}{Tr_d} \tag{18}$$

where $\phi/2$ is the angle of the right-angle triangle used to calculate the number of backward nodes of *node*(x) (Fig. 1). Thus, the number of backward nodes of *node*(x) can be computed as follows:

$$N_B(x) = \rho\left[\phi/2.{Tr_d}^2 - (Tr_{\min} - \bigtriangleup d(x)).\sqrt{{Tr_d}^2 - (Tr_{\min} - \bigtriangleup d(x))^2}\right] \tag{19}$$

For a node to be a CH, the necessary and sufficient conditions are:

$$x_{\text{CH}} = \begin{cases} N_F(x) \neq 0, N_B(x) \neq 0, & \forall x \in Area < K \\ N_F(x) \neq 0, N_B(x) = 0, & \forall x \in Area = K \end{cases} \tag{20}$$

Consider the scenario described in Fig. 3c. CH-1 has to transmit data to the BS. There are two paths: the first path is from node 1 to node 4 to node 7 and the BS (P-147B), and the second path is from node 1 to 5 to 8 and the BS (P-158B).

Assuming no energy is consumed while relaying information from the CH to the BS and the distance covered through the path P-158B is less than that through the path P-147B, then the total transmission energy required for P-158B is less than that required for P-147B. The reason for less energy consumption in the former path is less distance compared with the latter path, which has a relatively long distance. The shortest path from the source to the BS is computed using the Dijkstra algorithm with distance as the weight between nodes.

The sample shortest path from node 1 to the BS is P-158B. Algorithm 1 is used for energy dissipation, which implements a first-order radio model. This model is explained in Eqs. 1 and 2. In the fourth step, the cumulative routing table is shown in Table 4.

4 Results

The proposed ICR protocol is compared with existing protocols, LEACH [9], JCR [1] and O-LEACH [15]. The ICR protocol performs better than these protocols, specifically when comparing network lifetimes, and it is suitable for multi-hop IoT-based WSNs.

Table 5 Parameter values

Parameter	Value	Units
Sink position	50, 100	
Number of nodes	100, 200	
Initial energy	0.25, 0.5, 1.0	Joules
Area	(100 m × 100 m) (200 m × 200 m) (400 m × 400 m)	Meters
Packet size	2000	Bits
Control packet (data frame overhead) length	200	Bits
Transmitter energy T_x	50	nJ/bit/m^2
Receiver energy R_x	50	nJ/bit/m^2
Data aggregation energy	5	nJ/bit/m^2
Transmit amplifier (free space)	100	nJ
Transmit amplifier (multipath)	0.0013	nJ

However, before providing the details, it is important to explain the term 'round' used in this context. A round can be defined as a time interval in which the processes of node-setup, cluster organization, CH selection, and data transmission are completed. The first-order radio model is used for energy dissipation, which is explained in Algorithm 1. Simulations are implemented in MATLAB, and details of the parameters used in simulations are provided in Table 5.

In LEACH [9] protocol, cluster formation and selection of the CH are high energy consumption processes. However, ICR performs clustering by predefining sub-clustering areas based on the minimum transmission range. All of the nodes in the sub-area select the next CH based on accumulative weights of parameters such as the residual battery and connectivity. After this step, communication is started by all member nodes based on TDMA schedule. It is important to note that ICR uses major proportion of node energy for data transmission rather than cluster setup, advertisement, and resetting of the network.

4.1 Network throughput/ aggregation/node density/node heterogeneity

The network lifetime is reduced if data aggregation is disabled on CH nodes (Table 6, Figs. 4 and 5). Sensors, such as temperature and humidity sensors, which are co-located and provide similar data, can be aggregated to provide estimated results [3]. These results can be interpreted on a sink node with a predetermined threshold [3]. However, in the IoT context, in the case of sensors related to healthcare, such as blood pressure monitoring units or diabetes-related sensors, the value of each sensor is important because it belongs to individual patients and cannot be aggregated. Thus, the evaluation network related parameters including change in BS position, data aggregation on CH, throughput, node density and heterogeneous initial energy of nodes is essential.

Changes in the BS position increase or decrease the network lifetime. Networks with BS positioned at the border of the network area achieve a larger lifetime compared with BS at the center of the area or outside the network border (Fig. 6). It is also noticed that for a given transmission range and number of clusters, the network lifetime decreases

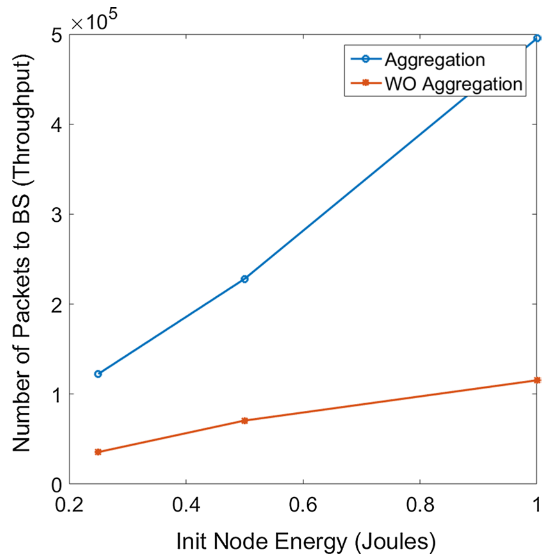

Fig. 4 Packets to the BS with and without aggregation for ICR

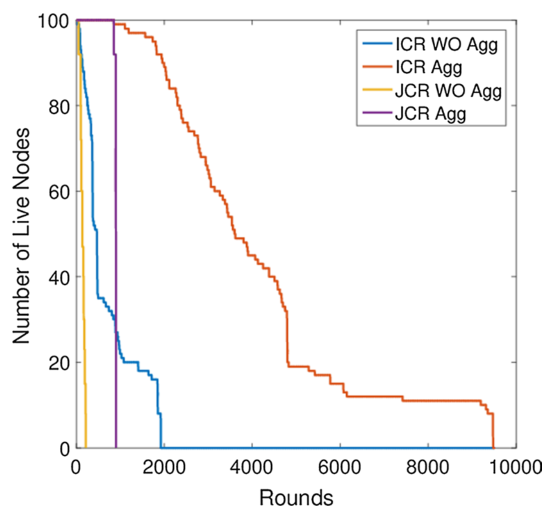

Fig. 5 Network life with and without aggregation on CHs (100 m × 100 m, Init Energy = 1.0 J)

Table 6 Data aggregation and network lifetime (transmission range = 50 m)

Init energy	E with agg (ETX + EDA)	Pks to BS	Rounds	E without agg (ETX)	Pks to BS	Rounds
0.25	4.0364e−04	122,423	2087	6.0550e−05	35,638	504
0.5	7.1335e−04	228,028	3663	1.2300e−04	70,618	998
1.0	0.0016	495,250	8074	1.6275e−04	115,374	1607

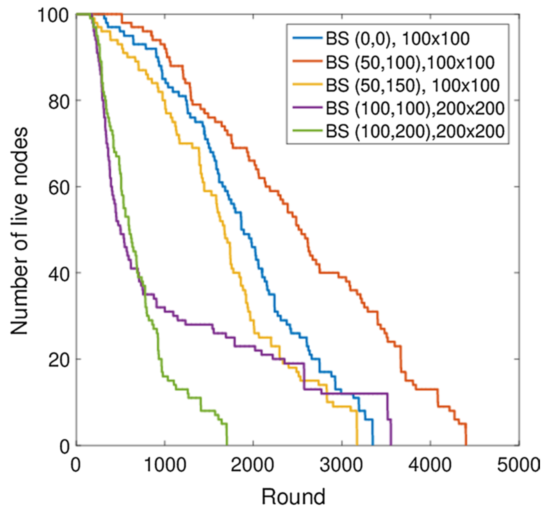

Fig. 6 Impact of the BS position and area size on network lifetime for ICR

with increasing area size. This is true for all protocols, such as LEACH [9], ICR, and JCR [1].

The impact of node density on different area sizes for ICR and JCR is provided in Figs. 17 and 18. The network lifetime is better following ICR compared with the JCR protocol. The network lifetime initially increases with an increase in the number of nodes and becomes stable for a larger number of nodes. This is because higher the node density in an area results in higher inter-packet communication for clustering and CH selection and, therefore, decrease in data transmission energy. Overall, the network lifetime is higher for smaller areas using both protocols (Fig. 17) compared with larger area sizes (Fig. 18). The peak value of the lifetime is observed for 400 nodes in the 100 m × 100 m area and for 500 nodes in the 200 m × 200 m area. The number of rounds in the ICR protocol is twice as much as that in JCR protocol. Additionally, at 1500 rounds, negligible lifetime is found for JCR protocol, while the ICR protocol shows 1000 rounds for 100 m × 100 m area and 1100 rounds for 200 m × 200 m area. It is observed that as soon as the node density increases, a decrease in network lifetime is observed for a given area. Figures 17 and 18 indicate that the network lifetime becomes minimal with a large number of nodes. Practically, it is difficult for the JCR protocol to manage larger node densities, while the ICR protocol still shows a higher value of the lifetime for high node densities. Both figures demonstrate the performance of ICR and JCR protocol for different nodes densities in the area. It is observed that smaller density values (e.g. 50 nodes in 100 m × 100 m area) have less network lifetime. This is because each node gets its CH turn in lesser time and exhausts energy soon. Also, the nodes are relatively sparsely located and consume more energy for data transmission to CH. Contrary to this, for higher node density values (500 and above), the network lifetime reduces again. The reason is the CH that not only have to transmit the data of larger number of CMs but also have to work as relay node for larger number of nodes. Thus, resulting in decrease of overall network lifetime for both protocols.

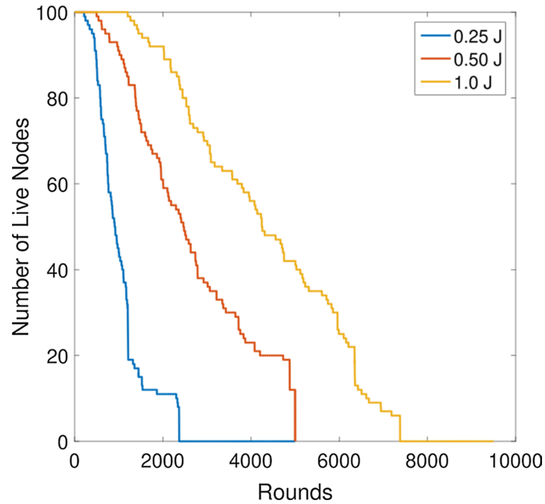

Fig. 7 ICR with different initial energy values

A large decrease in the number of packets reaching the BS (throughput) is observed when the initial energy of the nodes is less than the maximum initial energy. Node energy is selected such that initial energy, $E_i \leq E_{max}$, where E_{max} is the maximum energy a node can have at the start of a simulation (Table 10). Network lifetime and throughput in case of homogeneous and heterogeneous initial energy of nodes is shown in Figs. 20 and 21. With lesser initial energy value of each node, the network life time and throughput, both, have lesser values.

Similarly, higher throughput is observed for ICR protocol as compared with LEACH protocol (Fig. 22). This is due to the partitioning of the area, which results in a small clustering area. Hence, each node has to send its data to a relatively closer CH and save node energy. Second, the CH is only selected from nodes within the assigned area. The selection of CH is based on the node battery and degree of connectivity. In LEACH, the CH is selected randomly from live nodes without consideration of any parameter. Another major reason is that in LEACH, all the nodes are considered within a single transmission range, which is not a practical scenario for large-scale networks.

4.2 Alive nodes vs number of nodes

For the ICR protocol, the results indicate 2300, 5000, and 7400 rounds for 0.25 J, 0.5 J, and 1 J of initial node energy (Fig. 7) that show die out pattern for nodes with different initial energy values of nodes in ICR protocol. The network die-out times for the ICR and LEACH protocols under different initial energy levels are given in Table 7. Figure 8 shows that the network lifetime of ICR is much higher than that of the LEACH protocol.

Although the CH assigns a TDMA-based data transmission schedule to CMs, the CMs in different clusters face the collision of data packets with each other [4]. As a result, the number of packets reaching the BS decreases many orders of times in JCR [1], LEACH [9], and other dynamic-clustering-based protocols. However, ICR is based on fixed-area clustering, and nodes inside one area have considerably less interference with the nodes

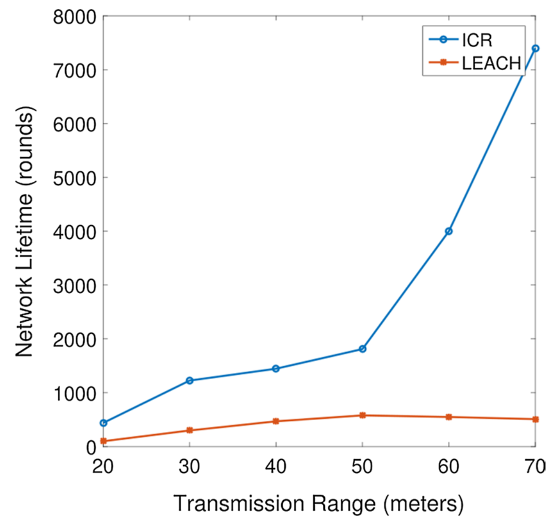

Fig. 8 Network lifetime (100 m × 100 m, 0.5 J)

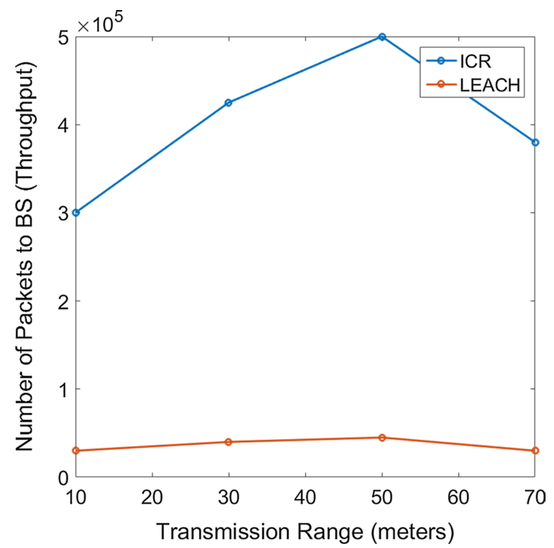

Fig. 9 Transmission range and packets to BS (100 m × 100 m, Init Energy = 1 J)

of other areas. Thus, the packet loss probability reduces many orders of magnitude, and many more packets reach the BS, as presented in Table 9.

Random cluster formation in LEACH protocol results in lesser number of packets reaching to BS. Since the number of disconnected nodes in any given round is lesser in LEACH as compared with ICR, all the nodes in the network are not able to transmit the data. ICR protocol provides complete connectivity and alternate routing path in multi-hop communication. This results in increased throughput in terms of number of packets reaching to BS (Fig. 9).

The LEH between CHs increases with an increase in the transmission range. This clear for both ICR and LEACH protocols (Fig. 10). However, for an area of 200 m × 200 m,

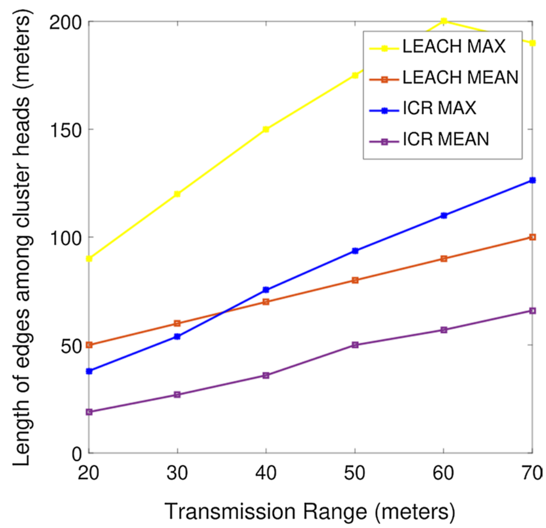

Fig. 10 Length of the edge (200 m × 200 m)

the LEH between two clusters is much higher for LEACH compared with that in the ICR protocol. This is due to FAC where clusters are created based on the transmission range and thus, restricted to a maximum length of the edge rather than probabilistic clustering with an undetermined position of CHs.

For the 200 m × 200 m area and transmission range of 70 m, the network lifetime using JCR and ICR protocols is shown in Fig. 11, where different initial energies of nodes, JCR and ICR are compared. Using the ICR protocol, a change in the transmission range also changes the number of clusters in the network, which ultimately affects the network lifetime. In Fig. 11, the network lifetime for different transmission ranges is compared for 1, 0.5, 0.25 J of initial node energy. For all cases, it is found that there is an optimal value of the number of clusters for given area size where the best network throughput is achieved. This is because a higher number of nodes are formed for a smaller number of clusters, and thus, a node finds many alternative routing paths to the BS. If the number of clusters increases, the number of nodes per cluster decreases, which results in a lower chance for a node to find the next-hop neighbor alive during packet propagation towards the BS. This phenomenon increases the number of hops and network management time and reduces network throughput.

All of these variations in the network lifetime running under different protocols are due to different phases that each protocol performs. ICR develops clusters initially and does not perform dynamic clustering. This increases its network lifetime compared with other protocols. A comparison of different phases of these protocols is provided in Table 8.

4.3 Impact of transmission range

The results in Fig. 12 indicate that an increase in the transmission range results in a lesser number of clusters and hop-counts for both multi-hop protocols, ICR and JCR [1]. However, there is an optimized value of clusters for every area. Table 9 shows that the maximum number of packets reaching the BS was achieved when the transmission

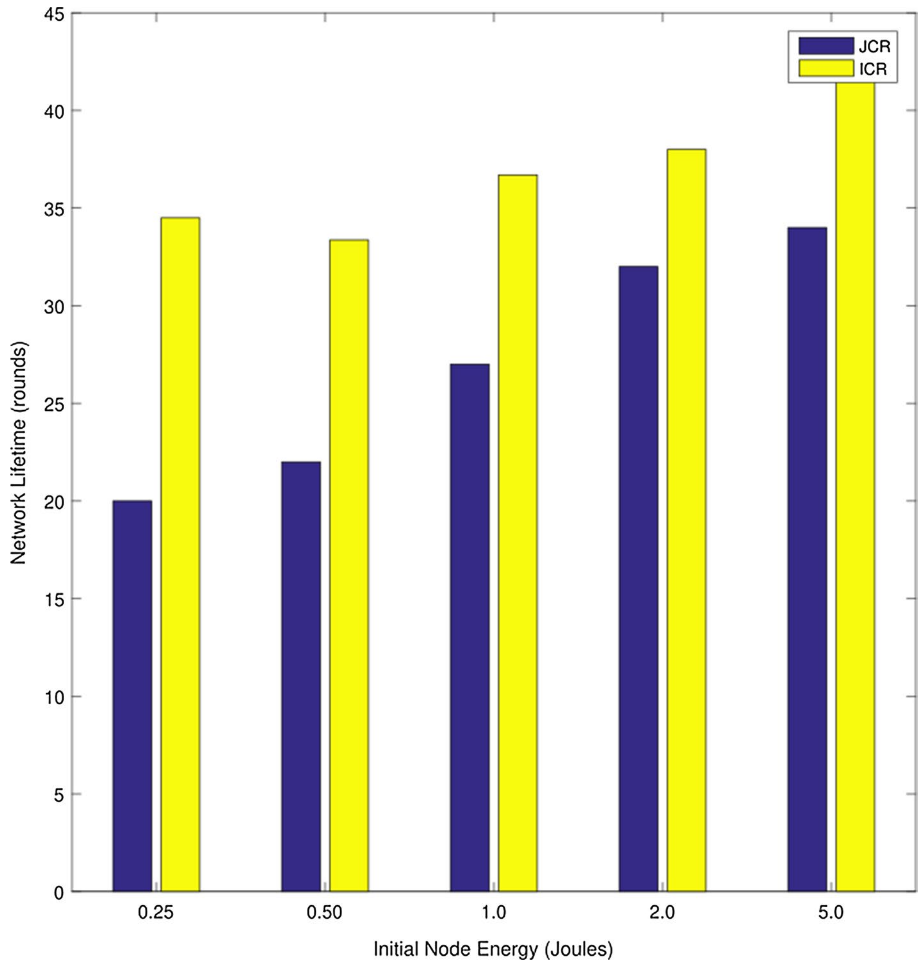

Fig. 11 Network lifetime of the ICR and JCR protocols (200 m × 200 m, Transmission range = 70 m)

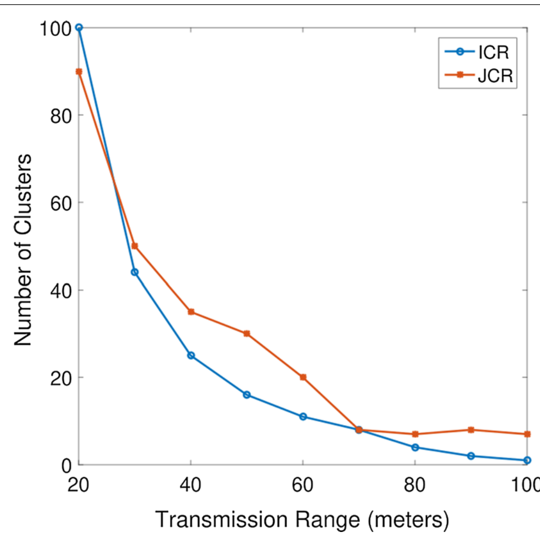

Fig. 12 Transmission range and number of clusters (200 m × 200 m)

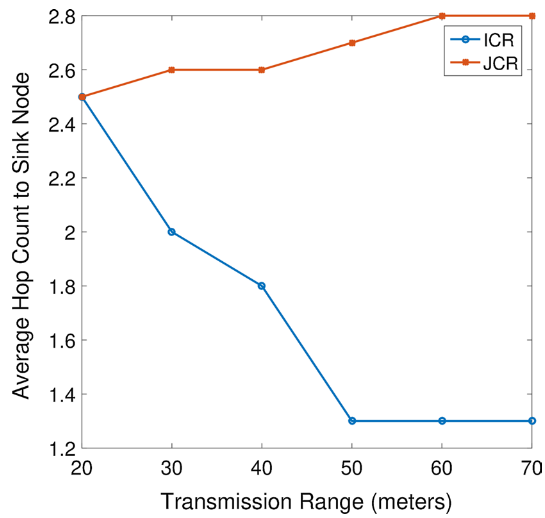

Fig. 13 Average hop-count to the sink node (100 m × 100 m)

range was 50 m (4 clusters) for an area of 100 m × 200 m for all initial energy levels of nodes. For larger numbers of rounds with 25 m and 33 m transmission ranges, fewer packets were observed to reach the BS, even with a higher number of rounds compared with the 50 m transmission range. The reason for the lesser number of packets to the BS for a large number of clusters or small transmission is that links in the multi-hop communication die-out and the transmission needs to be completed in the next round, which results in a higher number of rounds and less throughput. For a large transmission range, such as 100 m, which results in only one cluster based on the ICR protocol, the network dies out earlier because of the larger distance that a packet has to travel from the originating node to the CH and then from the CH to the BS.

The number of clusters formed is based on the transmission range of the nodes. For smaller transmission ranges, although JCR performs better than ICR, the results are comparable. However, for a given area of 200 m × 200 m, ICR is flexible in creating clusters according to the transmission range, while JCR fails to do that and creates more clusters (Fig. 12). JCR does not adjust the number of clusters when T_r is more than 70 m, while ICR succeeds in efficiently adjusting to larger transmission ranges.

4.4 Routing performance evaluation

The ICR protocol addresses clustering and routing as interlinked issues. The performance of routing in the ICR protocol is compared with the O-LEACH [15] protocol. The number of connected nodes, network lifetime, and energy consumption performance are used as evaluation parameters. The area of simulation is 100 m × 100 m, and 0.5 J is the initial energy of all nodes.

The average hop-count to the sink node is another measure that is defined as the total number of hops from CH to BS divided by total number of CHs in the area. We compare the average hop-count to the sink for JCR [1] and ICR protocols. As the transmission range of nodes increases from 20 to 70 m for an area of 100 m × 100 m, the average hop-count of both protocols demonstrates comparable variation (Fig. 13).

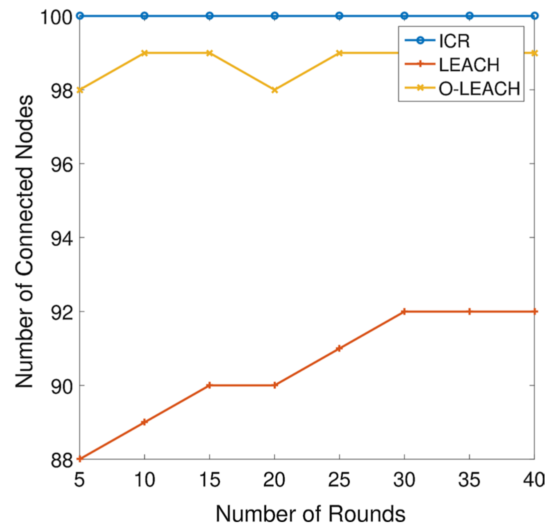

Fig. 14 Number of connected nodes (100 m × 100 m)

When the transmission range is increased from 20 to 70 m, the average hop-count to the BS increased from 2.64 to 2.76 for the JCR protocol, whereas it decreased from 2.5 to 1.3 hops for the ICR protocol. This is because the ICR protocol is adaptive to variable transmission ranges and produces an optimized clustering area. However, hop-count remains constant after a specific transmission range. ICR provides better clustering and routing topology generation by defining fixed area clusters and cluster heads whereas the area-size is variable in JCR that results in increased average hop-count even for increased transmission range. This proves that the transmission path generated by ICR is more efficient than JCR, and the advantage is greater when the transmission range is larger.

The number of connected nodes is compared in Fig. 14. O-LEACH is included because it is multi-hop routing protocol and extended version of benchmark LEACH protocol [9]. The number of connected nodes in every round is 100% for ICR, close to 100% for O-LEACH and 88–92% for LEACH. The reason is that the transmission range of every cluster is fixed, and all nodes within the cluster are covered within the range. However, in O-LEACH, due to gateway nodes, not only the network lifetime decrease, but some nodes also remain disconnected. Since O-LEACH is specifically designed to minimize disconnected nodes, it achieves good connectivity of nodes.

5 Discussion

The JCR [1] and LEACH [9] protocols are considered as benchmark protocols and are extensively compared with the ICR protocol. A comparative review on the performance of all these protocols is provided in this section.

The impact of the transmission range on the network lifetime for both protocols is shown in Fig. 15 for an area of 200 m × 200 m. It is observed that both protocols provide comparable results for shorter ranges. However, ICR performance is better for longer transmission ranges compared with the JCR protocol. This is because the

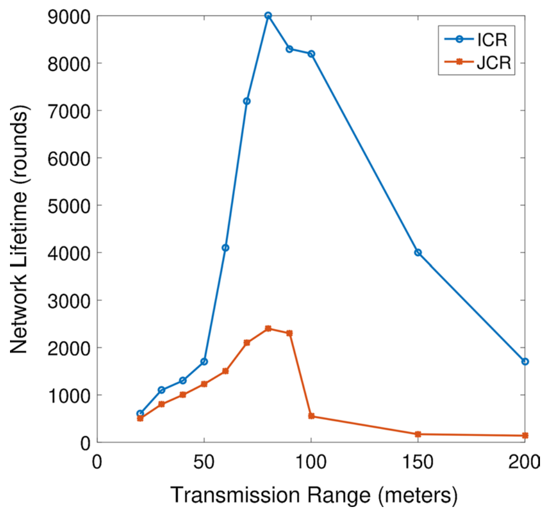

Fig. 15 Impact of the transmission range on network lifetime (200 m × 200 m)

number of clusters in ICR is optimized with a defined coverage area of the transmission range. This causes an increase in network lifetime that increases with the decreasing number of hops. However, it achieves the best throughput values for 2–3 hops communication; after this, it starts decreasing. In the case of the JCR protocol, for a large transmission range (80 m), the clusters are formed in such a way that they overlap transmission ranges. However, in the case of ICR, the cluster does not have overlapping transmission ranges due to FAC. For this reason, the network throughput increases dramatically and then starts decreasing because fewer hops result in overlapping radio ranges.

Dynamic clustering is the process in which clusters are regenerated after every round. During this process, new members are added or removed from the existing cluster, CHs are selected randomly or based on weights of different parameters, and most importantly, the location of the CH is not considered. The clusters could overlap, and the transmission suffers from the problem of radio interference. We propose a different idea for defining sub-areas in the network area and consider each of the sub-areas as clusters. All the nodes inside the sub-area are CMs, and the CH is selected from the sorted list of accumulative weights satisfying conditions. Thus, two performance gains are achieved. (1) The CH selection process is simplified, and the energy is reduced by fewer exchanged messages, and (2) less inter-cluster interference. A comparison of the percentage of energy used for clustering in different fixed-area and dynamic clustering protocols is provided in Fig. 19.

Both JCR and ICR mainly differ in clustering phases. The JCR uses a dynamic cluster where clusters are regenerated after every round and the CH is selected. The cluster generation process requires the exchange of information among nodes and thus requires more energy. Whereas, ICR is based on fixed-area clustering using the transmission range of nodes. Once the clusters are created, they remain the same until the end of network life. This FAC saves a valuable percentage of node energy, as shown by the results

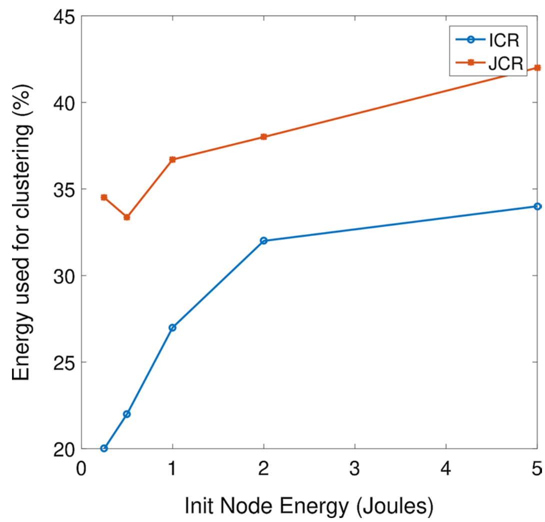

Fig. 16 Percentage of energy for clustering (100 m × 100 m)

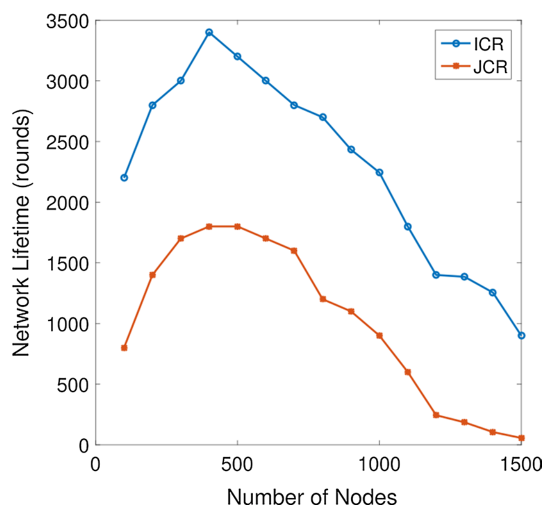

Fig. 17 Node density and area size (100 m × 100 m)

in Fig. 16. The CH selection process is simple and based on a sorted list of accumulative weights of nodes for being candidate CHs.

The performance of the ICR protocol is much better than that of the LEACH [9] and JCR [1] protocols in terms of energy consumption. Both JCR and ICR have common phases of clustering, CH selection, and routing. However, the energy consumption of ICR in clustering and CH selection phases is many orders of magnitude lower than that consumed in JCR. For an area of 100 m × 100 m and transmission range of 50 m, the percentage of energy used for clustering is shown in Fig. 16. The second point of comparison is the number of packets originating from different CMs and reaching the BS. Although the network lifetime of JCR is comparable with that of ICR, the number of packets reaching the BS is much higher in the case of ICR than in JCR. Even with the

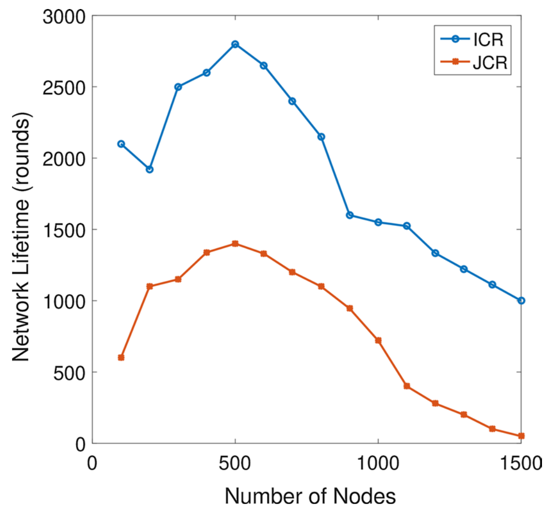

Fig. 18 Node density and area size (200 m × 200 m)

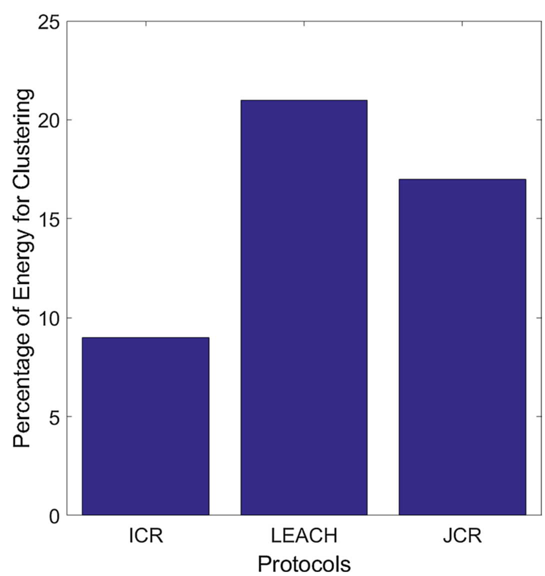

Fig. 19 Comparison of energy used in different protocols

same number of hops, the clustering of ICR is performed in such a way that reduces packet collisions and, therefore, increases the number of packets reaching to BS.

Another major comparison between ICR and LEACH [9] is the number of packets reaching the BS using different initial energies of nodes (Table 7). The number of packets reaching the BS using the ICR protocol is much higher than that using the LEACH protocol. The LEACH protocol performance is not better in terms of energy consumption because most of the energy of devices is utilized during the management phases rather

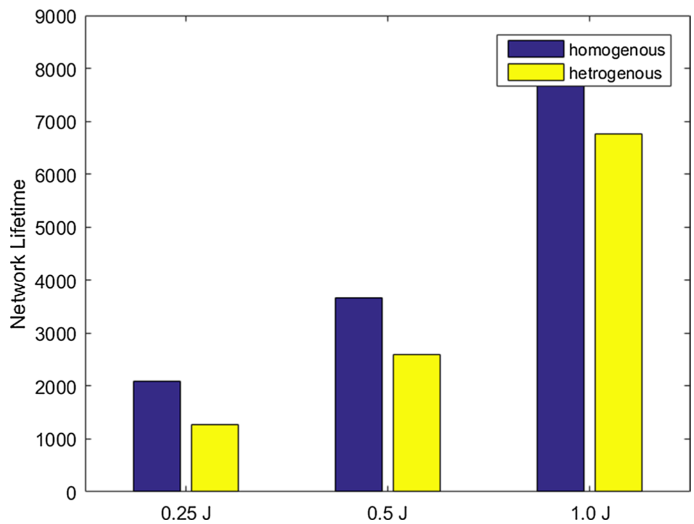

Fig. 20 Network life time of ICR (100 nodes, 100 m × 100 m)

than the data transmission phase. Longer transmission distances from the CM to the CH and from the CH to the BS are another reason for the waste of node energies.

An important difference between ICR and the other protocols is network die-out behavior. All protocols decrease the number of live nodes sharply, whereas ICR decreases rather linearly. This is due to FAC, where, in each cluster, nodes die one by one. Other protocols use dynamic clustering, which computes the CH dynamically based on values of one or more parameters. The preference of these protocols is to utilize each node energy equally. Although the overall network energy drops more rapidly compared with the ICR protocol, more nodes appeared alive. All the nodes reach the battery end-point at the same time, which is the reason for the sharp decrease in the network lifetime curve for these protocols.

6 Conclusion

Clustering and routing are treated as unified problem in this work. In previous studies, both of these issues were treated separately and thus failed to find an energy-efficient solution. Considering the geographical position of nodes, an improved clustering-based mechanism was proposed along with efficient routing of traffic in a multi-hop scenario. The fail-over scenario was also considered when designing the model.The results indicate that the protocol provides a comparatively better network lifetime (maximum 4.5 times for 80-m transmission range in the area of 100 m × 100 m compared with that of the JCR [1] and LEACH protocols. This algorithm can be extended for scenarios considering mobile nodes and BSs. Real-life random node deployment with random node distribution scenarios can also be considered from healthcare, shopping areas and stadiums. This will produce more optimized results in terms of network lifetime and overall network capacity (Figs. 17, 18, 19, 20, 21, 22).

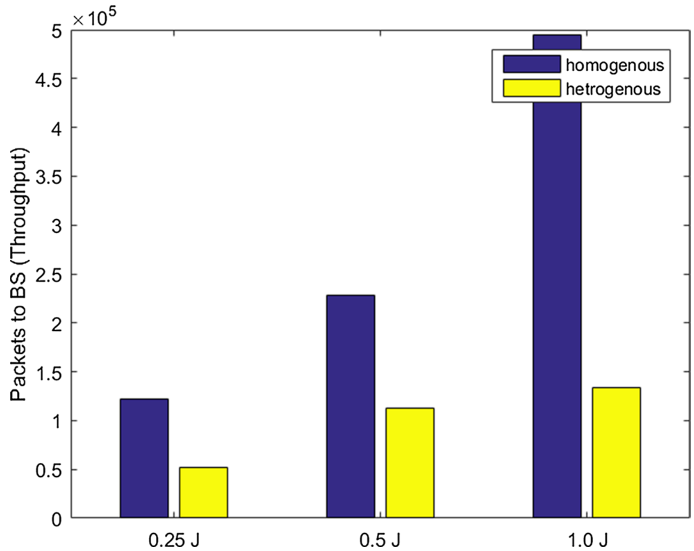

Fig. 21 Packets to BS (throughput) in ICR (100 nodes, 100 m × 100 m)

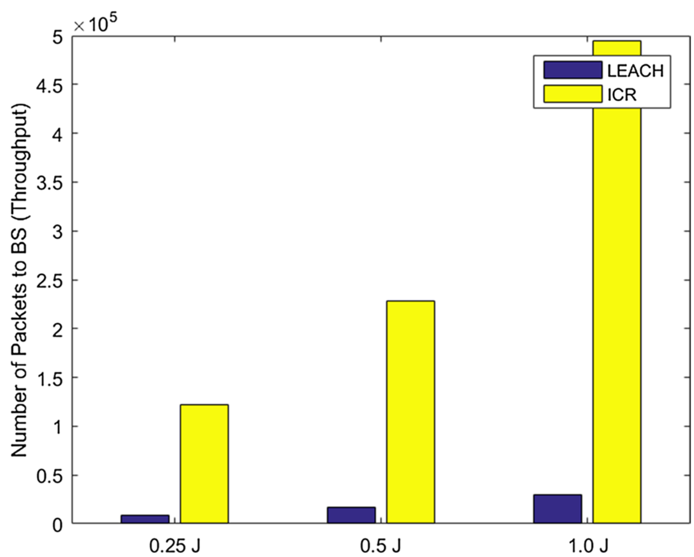

Fig. 22 Comparison of throughput (100 nodes, 100 m × 100 m)

Table 7 LEACH vs ICR

Energy (J/node)	Protocol	Rounds 1st node dies	Rounds last node dies	Packet to BS	E-consumed (Tx + EDA)
0.25	LEACH	394	665	8500	41.6349 e−09
	ICR	350	1850	122,423	4.0364 e−04
0.5	LEACH	932	1312	16,924	83.2754 e−09
	ICR	400	4300	228,028	7.1335e−04
1	LEACH	1848	2608	29,731	148.0973 e−09
	ICR	750	9200	495,250	0.0016

Table 8 Comparison of rounds

Step	LEACH	JCR	ICR
Advertisement	All nodes sharing residual battery information	Gradient field setup	Cluster-sorting
CH selection	Random selection of CH among nodes	CH selection using back-off timer	CH selection based on cumulative weights of parameters
Cluster setup	The CH sends an advertisement message to the CM. The CM joins the desired CH	Similar to LEACH	The CH only sends its info to the CMs in the sub-area
Scheduling	The CH generates the TDMA schedule and broadcasts it to all nodes in the network. All the nodes receive this message and accept or discard based on relevance	The CH sends the TDMA-based schedule information to all nodes in its transmission range	The CH sends the TDMA-based schedule information of only nodes in the sub-area
Route selection	No routing is performed in LEACH	The BS initiates the path discovery and moves from one gradient to the next and selects the optimal the next gradient CHs	The ICR protocol develops a multihop route of the CHs to the BS. This routing is initiated by the BS
Data transmission	The data transmission of LEACH is started from the CM to the CH. Since all the nodes are in the transmission range of each other, the probability of packet loss is very high	Data transmissions from the CM to the CH have a lower probability of collision compared with LEACH but more than that of ICR because of variable clustering ranges	Each clustering area has its own transmission range, the probabilities of packet loss in ICR are much less than LEACH

Table 9 Impact of the transmission range on packets

Energy	Pkts-to-BS	Rounds 25 m	Pkts-to-BS	Rounds 50 m	Pkts-to-BS	Rounds 100 m
0.25	82,082	2552	122,423	2087	81,256	1535
0.5	187,403	4761	228,028	3663	168,638	3224
1.0	298,188	9135	495,250	8074	385,579	7291

Table 10 Homogeneous and heterogeneous node energies ($T_r = 50$ m)

No.	Init energy	Pkts to BS Homogeneous	Rounds	Pkts to BS Heterogeneous	Rounds
1	0.25	104,726	1622	78,823	1208
2	0.5	207,463	3269	131,270	2671
3	1.0	491,399	8000	272,134	4983

Abbreviations
ICR: Improved clustering and routing; JCR: Joint clustering and routing; LEACH: Low energy edaptive hierarchical clustering; WSN: Wireless sensor network; IoT: Internet of things; CH: Cluster-head; CSMA/CA: Carrier sense multiple access/collision avoidance; FAC: Fixed-area-based clustering; LEACH-C: LEACH-centralized; BS: Base station; GEEC: Game theory based energy efficient clustering routing protocol; DEEC: Distributed energy-efficient clustering; DDEEC: Developed distributed energy-efficient clustering; EDDEEC: Enhanced developed distributed energy-efficient clustering; BEEN-ISH: Balanced energy efficient network integrated super heterogenous protocol; CM: Cluster member; LEACH-XMP: LEACH-eXtended Message-Passing; O-LEACH: Orphan-LEACH; LEH: Length of edge; TDMA: Time-division multiple access protocol.

Acknowledgements
Not applicable.

Authors' contributions
Conceptualization, MAA; methodology, MAA; software, MAA; validation, MAA, MK and MZ; formal analysis, MAA; investigation, MAA; writing (original draft preparation), MAA; writing (review and editing), MAA, MK and MZ; supervision, MK; project administration, MAA. All authors read and approved the final manuscript.

Funding Information
Not applicable.

Availability of data and materials
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK. [2] National University of Science and Technology (NUST), H-12, Islamabad 44000, Pakistan.

Received: 10 October 2020 Accepted: 26 January 2021
Published online: 06 March 2021

References

1. Z. Xu, L. Chen, C. Chen, X. Guan, Joint clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks. IEEE Internet Things J. **3**(4), 520–532 (2016)
2. H. Singh, D. Singh, Multi-level clustering protocol for load-balanced and scalable clustering in large-scale wireless sensor networks. J. Supercomput. **75**(7), 3712–3739 (2019)
3. W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**(4), 660–670 (2002)
4. P. Feng, D. Qin, P. Ji, M. Zhao, R. Guo, T.M. Berhane, Improved energy-balanced algorithm for underwater wireless sensor network based on depth threshold and energy level partition. EURASIP J. Wirel. Commun. Netw. **1**(1), 1–15 (2019)

5. N. Ayoub, M. Asad, M. Aslam, Z. Gao, E. U. Munir, R. Tobji, MAHEE: multi-hop advance heterogeneity-aware energy efficient path planning algorithm for wireless sensor networks, in *2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, Aug. 2017
6. G.S. Arumugam, T. Ponnuchamy, EE-LEACH: development of energy-efficient LEACH protocol for data gathering in WSN. EURASIP J. Wirel. Commun. Netw. **1**, 76 (2015). https://doi.org/10.1186/s13638-015-0306-5
7. J.A. Khan, H.K. Qureshi, A. Iqbal, Energy management in wireless sensor networks: a survey. Comput. Electr. Eng. (2014). https://doi.org/10.1016/j.compeleceng.2014.06.009
8. A. Dhumane, R. Prasad, J. Prasad, Routing issues in Internet of Things: A survey, in *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, vol. 1, pp. 16–18 (2016)
9. W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. Piscataway: IEEE, Jan. 2000, p. 10
10. T.N. Qureshi, N. Javaid, A.H. Khan, A. Iqbal, E. Akhtar, M. Ishfaq, BEENISH: balanced energy efficient network integrated super heterogeneous protocol for wireless sensor networks. Proc. Comput. Sci. **19**, 920–925 (2013)
11. L. Qing, Q. Zhu, M. Wang, Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. Comput. Commun. **29**(12), 2230–2237 (2006)
12. B. Elbhiri, R. Saadane, D. Aboutajdine, Developed distributed energy-efficient clustering (DDEEC) for heterogeneous wireless sensor networks, in *2010 5th International Symposium on I/V Communications and Mobile Network*. Morocco: IEEE, (2010), pp. 1–4
13. N. Javaid, T.N. Qureshi, A.H. Khan, A. Iqbal, E. Akhtar, M. Ishfaq, EDDEEC: enhanced developed distributed energy-efficient clustering for heterogeneous wireless sensor networks. Proc. Comput. Sci. **19**, 914–919 (2013)
14. J. Kang, I. Sohn, S. Lee, Enhanced message-passing based LEACH protocol for wireless sensor networks. Sensors **19**(1), 75 (2019)
15. W. Jerbi, A. Guermazi, and H. Trabelsi, O-LEACH of routing protocol for wireless sensor networks, in *Proceedings of 13th International Conference Computer Vision Graphics and Image Processing (CGiV)*, pp. 399–404 (Mar. 2016)
16. Advanticsys. https://www.advanticsys.com/shop/mtmcm5000msp-p-14.html
17. V. Mhatre, C. Rosenberg, Design guidelines for wireless sensor networks: communication, clustering and aggregation. Ad Hoc Netw. **2**(1), 45–63 (2004)
18. M. Penrose, The longest edge of the random minimal spanning tree. Ann. Appl. Probab. **7**(2), 340–361 (1997)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.