IET Communications

*Research Article*

IET Journals
The Institution of Engineering and Technology

# CSDGP: cluster switched data gathering protocol for mobile wireless sensor networks

ISSN 1751-8628
Received on 21st November 2018
Revised 17th June 2019
Accepted on 20th August 2019
E-First on 16th October 2019
doi: 10.1049/iet-com.2018.6152
www.ietdl.org

*Kalaivanan Karunanithy*[1] ✉, *Bhanumathi Velusamy*[1]

[1]*Department of ECE, Anna University Regional Campus, Coimbatore, Tamilnadu 641046, India*

✉ *E-mail: kalaivaanankk@yahoo.com*

**Abstract:** Nowadays, most of the research focuses on the distributed algorithm to form the clusters in wireless sensor networks, in which the sensor nodes act autonomously to self-configure themselves with the help of the local information. However, the issues in such distributed mechanisms are (i) how to assign the time division multiple access (TDMA) schedule to collect the data from the cluster member (CM), (ii) how to distribute and balance the overall energy consumption of the cluster heads (CHs) under the circumstance of node's mobility. To overcome these problems, a cluster switched data gathering protocol (CSDGP) is proposed which ensures a uniform distribution of CHs by utilising a waiting time-based CH selection and an effective cluster switch based TDMA scheduling mechanism. The salient feature observed in the proposed CSDGP is that the migrated node has a chance to transmit the sensed data to the CH, whenever the CH has a free time slot. From the simulation results, it is proved that CSDGP provides a good performance under three different scenarios in comparison with existing protocols in terms of average energy consumption, average end-to-end delay, packet delivery ratio, and control overhead.

## Nomenclature

| | |
|---|---|
| $k$ | size of data packets |
| $B$ | transmission bit rate |
| $d_{i,j}$ | distance between the sensor nodes |
| $\gamma$ | propagation speed ($3 \times 10^8$ m/s) |
| $r$ | average packet arrival rate |
| $C_{ld}(t)$ | total link delay of the cluster |
| $N_{\mathrm{m}}$ | minimum number of sensor node is needed to cover the entire sensing area |
| $A_{\mathrm{h}}$ | area of the regular hexagon, i.e. coverage of sensor node |
| $A_n$ | area of the targeted sensing region |
| $N_a$ | number of deployed sensor node in the sensing area |
| $\varpi$ | ratio of the minimum number of sensors required to the total number of sensor nodes deployed in the sensing area |
| $\zeta(n)$ | threshold value |
| $r(n)$ | random number |
| $\psi$ | index value of the current round |
| $\varrho$ | number of sensor nodes waiting to act as TCH |
| $\chi$ | multiplication factor |
| $\xi_{\mathrm{ACT}}$ | normalised value of average connection time |
| $\xi_{\mathrm{NN}}$ | normalised value of number of neighbour nodes |
| $\xi_{\mathrm{res}}$ | normalised value of residual energy |
| $L_{i,j}(t)$ | link delay between the two nodes |
| $\delta_Q(t)$ | queuing delay |
| $\delta_T(t)$ | transmission delay |
| $\delta_{Pr}(t)$ | processing delay |
| $\delta_P(t)$ | propagation delay |
| $M_{RA}$ | maximum number of attempts assigned for data retransmission |
| $M_{\mathrm{CM}}$ | maximum number of member nodes in each cluster |
| $\nu_1$, $\nu_2$ | speed of sensor nodes |
| $\theta_1$, and $\theta_2$ | moving direction of the sensor nodes |
| $R_m$ | transmission range |
| $\Delta t_{i,j}$ | connection time between the sensor nodes |
| $T_{LT}(t)$ | average life time of the cluster |
| $\hat{N}_{\mathrm{TCH}}$ | number of TCH's neighbours |
| $\eta_i$ | maximum connection time with its members |
| $E_{\mathrm{c}}$ | current residual energy of TCH |
| $E_{\mathrm{m}}$ | maximum energy of the sensor node at the time of the deployment |
| $B_{\mathrm{wait}}(t)$ | delay time of the final CH broadcasting message |
| $t_{\mathrm{cf}}$ | total duration for cluster formation |
| $\delta_r$ | very small time frame ($\delta_r \ll t_{\mathrm{cf}}$) |
| $\theta_{aoa}$ | angle of arrival of the transmitting signal |
| $T_{ci}(n)$ | cost value |
| $d_{CS_i}$ | distance between CH node $i$ and SN |
| $d_{CS_j}$ | distance between CH node $j$ and SN |
| $d_{SB}$ | distance between SN and BS |
| $d_{C_{i,j}}$ | distance between CH node $i$ and CH node $j$ |
| $d_{CB}$ | distance between CH and BS |
| $d_{CB_i}$ | distance between CH node $i$ and BS |
| $d_{GB_i}$ | distance between GN node $i$ and BS |
| $d_{CB_j}$ | distance between CH node $j$ and BS |
| $d_{GB_j}$ | distance between GN node $j$ and BS |
| $V_c$ | current velocity of the sensor node |
| $V_m$ | maximum velocity of the sensor node |
| $T_{cp}(n)$ | cost value |
| $d_l$ | depth of the layer |
| $\epsilon$ | energy consumption of amplifier |
| $\alpha$ | amplification factor for multi-path or free space |
| $E_e$ | energy consumption of the radio electronics |
| $k$ | size of data packets |
| $d$ | distance between the nodes |
| $E_D^{\mathrm{CM}}$ | energy consumed by CM for transmitting and sensing the data |
| $N_{CH}$ | number of CH in each round |
| $E_{\mathrm{c}}^{\mathrm{CM}}$ | energy consumed by CM for transmitting and receiving the control packets |
| $L_c$ | length of control packets |
| $E_T^{\mathrm{CM}}$ | total energy consumed by CM |
| $E_{\mathrm{c}}^{\mathrm{SN}}$ | energy consumed by the sensor nodes for exchanging the node discovery message |
| $N_{nn}$ | number of neighbours of sensor node |
| $N_{GN}$ | number of gateway nodes in each round |
| $E_D^{\mathrm{GN}}$ | energy consumed by GN for transmitting the data |

| | |
|---|---|
| $N_{dp}$ | number of processed data packets |
| $E_{\mathrm{c}}^{\mathrm{GN}}$ | energy consumed by GN for transmitting and receiving the control packets |
| $E_T^{\mathrm{GN}}$ | total energy consumed by the GN |
| $E_D^{\mathrm{CH}}$ | energy consumed by CH for data communication and sensing the data |
| $N_{\mathrm{CM}}$ | number of CM in each cluster |
| $E_{\mathrm{c}}^{\mathrm{CH}}$ | energy consumed by CH for transmitting and receiving the control packets |
| $E_T^{\mathrm{CH}}$ | total energy consumed by the CH |
| $\max_{\mathrm{gen}}$ | maximum number of iterations |
| $n_{\mathrm{gen}}$ | number of candidates |

## 1 Introduction

In recent years, wireless sensor networks (WSNs) have attracted the attention of the research community and the industrialists towards its recent advancement in the fields such as micro-electro-mechanical systems, embedded system, and wireless communication. These technologies have facilitated the betterment in the design of inexpensive sensor nodes with low power [1, 2]. However, there are some constraints on the sensor node such as memory, battery energy, processing capability, communication bandwidth, and so on [3–5].

In WSNs, numerous protocols have been evolved for many applications like tracking of wildlife animal, rescue in a landslide, earthquake, flooding, and so on. However, the replacement or recharging of the battery from the man inaccessible area is not possible. Therefore, the conservation of energy is an important task in the battery operated WSNs. In mobility-based WSNs, the sensor node often changes its position leading to link failure and makes the network unstable [6]. In order to improve the scalability of the networks, the clustering mechanism has been introduced in [7] which divides the sensing area into smaller groups and each group has one head node, called as cluster head (CH). In clustering mechanism, there are two possible ways to select the CH including centralised and distributed manner.

In the centralised CH selection scheme, the base station (BS) acts as a main control centre of the networks to select the required number of CHs and coordinate the data transmissions based on the information such as location, residual energy, number of neighbours, memory, and so on, gathered from the sensor nodes (i.e. the BS requires the global knowledge of the networks). However, the centralised scheme has a constraint in large-scale sensor networks because of its inability to send the sensor node's information to the BS due to its limitation in the transmission range. In LEACH-C [7], the CH selection was performed by the BS in a centralised manner, therefore, the sensor nodes sent its information to the remote BS over a long distance in each and every round causing high energy consumption and leads to the early death of the sensor nodes. Muruganathan *et al.* [8] proposed a BS controlled dynamic clustering protocol (BCDCP) to alleviate the issue in LEACH-C by allowing only one CH forward the data to the BS instead of all the CHs attempt to communicate with the BS. And also, it reduces the probability of multiple CHs to come in the same cluster. Thus, it ensures the uniform distribution of CHs. Sabor *et al.* [9] proposed an adjustable range based immune hierarchy clustering (ARBIC) protocol which determined a CH's position based on the mobility, link connection time, and energy. The BS reorganises clustering process whenever the residual energy of the CH goes below the threshold value resulting in reduced computation time. However, the clustering process takes more computational time to find the best position of the CHs. And also, it meets increased control overhead due to the CH requisitions for sending the data from each cluster member (CM). And also, the scalability of the network cannot be extendable due to the clustering process which was controlled by the BS.

Hence, the researchers are focusing on the distributed scheme, in which the sensor nodes autonomously organise the cluster and select the CH based on the local information. The CHs in each cluster act as a local controller to coordinate the CM and data communication. Thus, it avoids direct communication with the BS over a long distance. By this way, it conserves the battery energy significantly. And also, each CH concurrently performs the network operation and enhances the scalability. The benefits of unequal clustering algorithm such as alleviating the hot spot problem, reducing the traffic burden of the nodes nearer to the BS were discussed in [10]. However, the presence of a large number of nodes in the cluster shrinks the longevity of the WSNs due to the overburdening of the CH and also causes high collision and contentions in accessing the channel [11]. The issues of control overhead and load balancing among the CHs were discussed in [12]. In WSNs, most of the battery energy is wasted in the collision and overhearing as mentioned in [13, 14]. Han *et al.* [15] addressed the issue of high energy expenditure due to the redundant coverage in the overlapped cluster formation. And also, the quality of service was suffered due to the simultaneous data transmission [16, 17].

The most popular distributed algorithm in WSNs, called as low energy adaptive clustering hierarchy (LEACH) was proposed in [7], in which the probability value was considered for finding the chance to become a CH. However, the low energy node is often elected as a CH which leads to the non-uniform distribution of CH and unbalanced workload than other nodes causing isolated nodes and poor network lifetime. LEACH-Mobile (LEACH-M) presented in [18] achieved successful packet delivery ratio by sending a consecutive time division multiple access (TDMA) schedule. The CM receives a data request message from the CH during its TDMA schedule and it sends the data to its CH, otherwise, the CM will know itself that it is leaving from its current cluster and it has to search for a new CH by broadcasting a join request message. Hybrid energy-efficient distributed clustering (HEED) was proposed in [19] which utilised a joint strategy including the communication cost and residual energy to select the CHs. It avoids the multiple CHs to lie in the same CH transmission range and ensures the uniform distribution of CHs throughout the networks. But, a random setup in the initial consideration of the number of CHs leads to the degradation in its performance. Malathi *et al.* [20] proposed hybrid unequal clustering with layering (HUCL), in which the CH candidates wait a random time before broadcasting CH awareness message based on the residual energy, distance to the BS, and the number of neighbour nodes. Thus, it alleviates the hot spot problem nearer to the BS and balances the traffic load among the sensor nodes. However, the utilisation of compression technique increases the computational and time complexity.

Energy-aware distributed dynamic Clustering Protocol using Fuzzy logic (ECPF) was proposed in [21], in which the tentative CHs (TCHs) were selected based on the delay time, i.e. inversely proportional to the residual energy. Thus, it ensures the uniform distribution of CHs and minimises the isolated nodes. However, the cluster formation initiated by the BS based on the residual energy of the CHs increases the complexity and limits the scalability of the networks. Lee *et al.* [22] proposed a location-based unequal clustering algorithm (LUCA), in which each node waited for a random back-off time before broadcasting CH awareness message to the neighbours. However, the low energy sensor nodes were often elected as CH due to the random backup time resulting in reduced network lifetime. Chao *et al.* proposed [23] energy efficient clustering algorithm (EECA), in which the set of TCHs was independently selected by using the residual energy and the distance to its neighbour nodes. Then, it broadcasts CH candidate awareness message to the neighbours and waits for some time to select one with the highest residual energy among the CH candidates to act as a final CH. And also, it utilised a data aggregation technique to remove the redundant data packets, thus it lessens the energy consumption. Energy-balancing unequal clustering approach for gradient-based routing (EBCAG) was presented in [24] which used a competition based CH selection with respect to the residual energy and determined the cluster size according to the hop count from the BS. Thus, it balances the energy consumption. However, the BS utilised a flooding method to divide the sensing area into the unequal size of clusters causing extra communication overheads. Therefore, it is not suitable for large-scale WSNs. Deng *et al.* [25] proposed mobility-based clustering (MBC) protocol which primarily considered the node's

residual energy and velocity to select the CH. In intra-cluster communication, the CHs assign TDMA schedule to the member nodes based on the connection time that enhances the network stability. In order to increase the successful packet delivery rate, it has a chance to receive the data from the migrated sensor node whenever it has a free time slot. However, it severely affected due to the isolated nodes causing non-uniform energy consumption and diminishing lifespan of the networks. In [26], velocity energy-efficient and link-aware cluster-tree (VELCT) considered the node's residual energy, number of neighbour nodes, coverage distance, and velocity for electing the CH. It separately organises the data collection tree (DCT) in order to collect the data from CH and enhance network stability. Due to the flooding-based DCT construction leads to high control overhead, when the scale of the network increases. And also, the received signal strength was utilised to calculate the connection and coverage time which evinces the unreliable link stability. A new routing strategy in WSNs namely tap root based inter-cluster data collection mechanism was introduced in [27]. Additionally, it provided a reliable measurement of the connection time by using a magnetic compass.

The main contributions of this paper are summarised as follows:

(i) A delay time based distributed CH selection method is introduced to ensure the even distribution of workload and to avoid the isolated node problem.
(ii) A cluster switched data gathering is presented to reduce the contentions in MAC level and data retransmission.
(iii) Tree topology based inter-cluster data communication is implemented to maintain the connectivity and stability of the network topology by considering the node's energy, velocity, and distance.
(iv) Representing a mathematical analysis to examine the energy consumption, control overhead, and time complexity of the proposed protocol.
(v) Conducting the simulation using NS-2 to analyse the effectiveness of the proposed cluster switched data gathering protocol (CSDGP) in comparison with ARBIC [9], VELCT [26], MBC [25], and LEACH-M [18].

The rest of the paper is organised as follows. Section 2 presents the problem statement. The design of CSDGP is described in Section 3. In Section 4, the mathematical analysis of CSDGP is performed. Analysis of simulation results is discussed in Section 5. Finally, the conclusion is presented in Section 6.

## 2 Problem statement

Most of the clustering protocols focused to avoid the interference within the cluster by using the TDMA schedule to the member nodes and also allows simultaneous data transmission in the neighbour clusters. But, these protocols do not bother about the interference due to the signal coming from the neighbour clusters. For example, the distributed clustering schemes in LEACH, LEACH-M, MBC, VELCT, HEED, EECA, and HUCL cannot attain the non-overlapped clusters. In this case, it involves high data collision and energy dissipation due to the CM of the neighbour clusters which may also utilise the same TDMA time slot for transmitting the data to its CH. As shown in Fig. 1, due to the overlapped cluster formation, the transmission power of the interfering node (IN) is shared among the multiple CHs. If the IN sends its collected data to the concerned CHs during its TDMA schedule, then it may cause the data reception of neighbour CHs, resulting in packet collision and loss. Another critical issue in the cluster is an isolated node (i.e. the sensor nodes are uncovered by the CH and live alone) which occurs due to the improper selection of CHs. These uncovered nodes are continuously searching a CH or directly communicating with the BS for transmitting the sensed data, there by depleting its energy very quickly. This is because of the LEACH, MBC, and VELCT protocols utilised only some percentage of sensor nodes to act as a CH. But, these protocols used a probabilistic method for selecting the CHs, if all the selected CHs lie near the border of the sensing area causing non-uniform distribution of CH. Therefore, the probabilistic selection with a fixed number of CHs cannot ensure the link available to all the regular sensor nodes, resulting in isolated nodes. Moreover, the waiting time based CH selection performed in EBCAG, ECPF, EECA, and HUCL meet the convergence issues. If all the deployed sensor nodes participated in the CH selection competition makes more complexity in achieving the optimum number of CHs. And also, it increases the convergence delay and control overhead. Table 1 shows the methodology applied and drawbacks for various clustering protocols with the proposed CSDGP. The symbols used in the proposed CSDGP are listed in Nomenclature section.

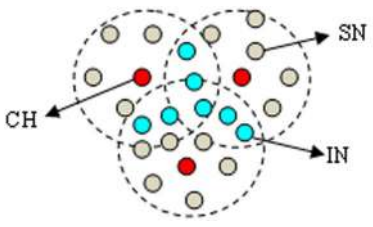


**Fig. 1** *Overlapped cluster formation*

## 3 CSDGP design

The operation of the CSDGP is divided into the rounds, as shown in Fig. 2. Each round consists of the setup phase and the steady-state phase. In the setup phase, it performs the cluster formation through the TCH selection and the waiting time based final CH awareness message broadcasting and also organises the tree topology for inter-cluster communication. Scheduling the TDMA slot and the duration of the data collection for each cluster is performed based on its cluster switch state (CSS). Finally, the CH collects the data from the CM which in turn transmits to the BS through the established tree topology in the steady-state phase.

### *3.1 Preliminary considerations in node deployment*

The network considered here comprises of ($N_a$) number of sensor nodes deployed over the targeted sensing area of $M \times M\text{m}^2$. To develop CSDGP, a few assumptions are made as follows:

(i) All the sensor nodes know their location and moving direction in the sensing area.
(ii) All the CHs have a capability to transmit its signal upto two hop distance.
(iii) All the sensor nodes know their $x$, $y$ coordinate of the BS.
(iv) Sensor nodes can adjust their transmission power based on the distance between them.
(v) The links are assumed to be symmetric.

The minimum number of sensor nodes ($N_\text{m}$) needed to cover the entire sensing area is found by using the following equation:

$$N_\text{m} = \frac{A_n}{A_\text{h}} \tag{1}$$

where $A_\text{h}$ is the area of the regular hexagon, i.e. coverage of the sensor node, $A_n$ is the area of the targeted sensing region.

In (2), $\varpi$ is denoted as the ratio of the minimum number of sensors required to the total number of sensor nodes deployed in the sensing area

$$\varpi = \frac{N_\text{m}}{N_a} \tag{2}$$

### *3.2 Setup phase*

*3.2.1 CH selection phase.:* To overcome the issues found in the literature, CSDGP is introduced which elects the CH in two steps: (i) TCH election and (ii) a delay time based final CH selection.

**Table 1** Comparison of the existing clustering protocols with proposed CSDGP

| Protocols | Clustering mechanism | Number of CHs (fixed-F, not fixed-NF) | Sufficient number of CHs to cover the whole network | Convergence time for getting optimum number of CH | Load balancing | Isolated node | Scalability | TDMA schedule assigned by either CH or BS | Concurrent data transmission allowed in the neighbour cluster | Data collision due to signal from neighbour cluster |
|---|---|---|---|---|---|---|---|---|---|---|
| LEACH [7] | distributed | F | no | low | no | yes | yes | CH | yes | high |
| LEACH-C [7] | centralised | F | yes | low | yes | no | no | CH | yes | high |
| LEACH-M [18] | distributed | F | no | low | no | no | yes | CH | yes | high |
| MBC [25] | distributed | F | no | low | no | yes | yes | CH | yes | high |
| VELCT [26] | distributed | F | no | low | no | yes | yes | CH | yes | high |
| ARBIC [9] | centralised | NF | yes | high | yes | no | no | CH | no | low |
| EBCAG [24] | distributed | NF | yes | high | yes | no | yes | periodical | no | low |
| ECPF [21] | distributed | NF | yes | high | yes | no | yes | CH | yes | high |
| BCDCP [8] | centralised | NF | yes | low | yes | no | no | CH | yes | high |
| HEED [19] | distributed | NF | no | low | no | yes | no | CH | yes | high |
| EECA [23] | distributed | NF | yes | high | yes | yes | yes | CH | yes | high |
| HUCL [20] | distributed | NF | yes | high | yes | no | yes | CH | yes | high |
| proposed CSDGP | distributed | NF | yes | low | yes | no | yes | CH | no | low |

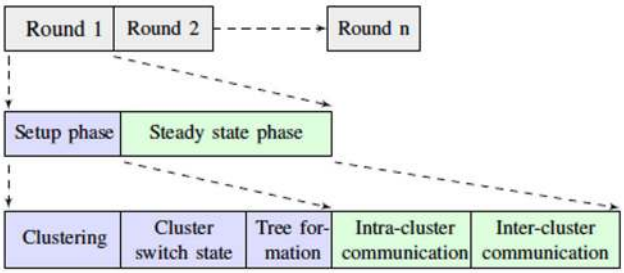


**Fig. 2** *Operation of CSDGP*

Initially, each sensor node broadcasts the HELLO message with its location, moving direction and speed to the neighbours using carrier sense multiple access and collision avoidance (CSMA/CA). It is used to have an idea about the neighbour sensor nodes within its transmission range. Therefore, the sensor nodes can determine the connection time and number of neighbours. The CH selection and cluster formation is described in Fig. 3.

In each round, a set of normal nodes is selected as TCH in order to ensure the uniform distribution of CHs to the entire networks. In TCH selection phase, each sensor node calculates its threshold value in order to get a chance to become a TCH based on (3) and also generates a random number between [0, 1]. To avoid the low energy sensor nodes selected as TCHs and to ensure the uniform energy consumption of the sensor nodes, the residual energy of the nodes is considered in (3). If the threshold value $\zeta(n)$ is greater than the random number, then the sensor node identifies itself to become a TCH.

The threshold value of TCH is expressed as

$$\zeta(n) = \begin{cases} \dfrac{\chi \times \varpi}{1 - \chi \times \varpi[\psi \bmod \frac{1}{\chi \times \varpi}]} \times \dfrac{E_c}{E_m} & \forall N_a \in \varrho \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\psi$ is the index value of the current round, $\varrho$ denotes the number of sensor nodes waiting to act as TCH, $E_c$ is the current residual energy, $E_m$ is the maximum energy of the sensor node at the time of the deployment, $\chi$ is a multiplication factor which depends on the requirement of applications, node coverage, sensing area, and node density, that is, its range is $1 \le \chi \le (\text{Node density}/N_m)$. The $\chi$ is used to define the number of final CHs which lie between $N_m$ and $\chi N_m$. The multiplication factor considered in the proposed CSDGP is 2, i.e. $\chi = 2$.

The average connection time and the number of neighbour nodes are considered in order to select the required CH from the TCH set.

(a) *Average connection time*

To enhance the stability of the networks and to find the average lifetime of the cluster, the link delay between the nodes is considered in CSDGP. Based on [27], the link delay $L_{i,j}(t)$ between the two nodes is the summation of queuing delay $\delta_Q(t)$, transmission delay $\delta_T(t)$, processing delay $\delta_{Pr}(t)$ and propagation delay $\delta_P(t)$. It is calculated by using (4)–(7)

$$\delta_T(t) = \frac{k}{B} \quad (4)$$

where $k$ is the size of data packets, $B$ is the transmission bit rate

$$\delta_P(t) = \frac{d_{i,j}}{\gamma} \quad (5)$$

where $d_{i,j}$ is the distance between the sensor nodes, $\gamma$ is the propagation speed ($3 \times 10^8$ m/s)

The queuing model M/M/1 is utilised in the proposed design and its queuing delay is computed as

$$\delta_Q(t) = \frac{\frac{r \times k}{B}}{1 - \frac{r \times k}{B}} \times \frac{k}{B} \quad (6)$$

where $r$ is the average packet arrival rate

$$L_{i,j}(t) = M_{RA} \times (\delta_Q(t) + \delta_T(t) + \delta_P(t) + \delta_{Pr}(t)) \quad (7)$$

where $M_{RA}$ is the maximum number of attempts assigned for data retransmissions.

The total link delay of the cluster $C_{ld}(t)$ is expressed as

$$C_{ld}(t) = M_{CM} \times L_{i,j}(t) \quad (8)$$

where $M_{CM}$ is the maximum number of member nodes in each cluster (i.e. $M_{CM} = N_a/N_m$).

The estimation of connection time between the nodes is a key factor in mobility-based WSNs because it will measure how long the sensor nodes remain in contact with another node. Let us

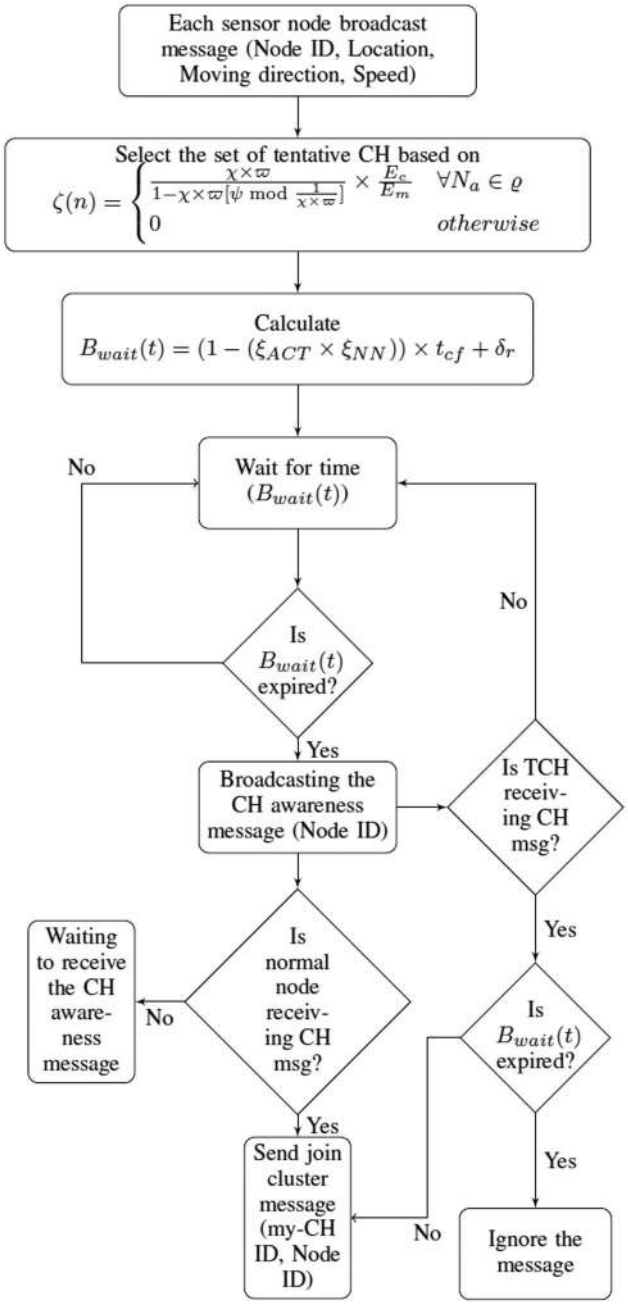


**Fig. 3** *Flowchart for cluster formation*

consider the sensor nodes are with their ($x$, $y$) co-ordinates as $P_1(x_i, y_i)$ and $P_2(x_j, y_j)$, and their Cartesian position, $x_i = x_1 + \nu_1 t\cos\theta_1$, $y_i = y_1 + \nu_1 t\sin\theta_1$, $x_j = x_2 + \nu_2 t\cos\theta_2$, $y_j = y_2 + \nu_2 t\sin\theta_2$. Here, $\nu_1$, $\nu_2$, $\theta_1$, and $\theta_2$ are the speed and moving direction of the sensor nodes in the sensing area at any instant of time $t$ [25]. The distance between $P_1$, and $P_2$ within the transmission range $R_m$ is given in (9)

$$(x_j - x_i)^2 + (y_j - y_i)^2 = R_m^2 \tag{9}$$

$$\begin{aligned}((x_1 - x_2) + (\nu_1 t\cos\theta_1 - \nu_2 t\cos\theta_2))^2 + ((y_1 - y_2) \\ +(\nu_1 t\sin\theta_1 - \nu_2 t\sin\theta_2))^2 = R_m^2\end{aligned} \tag{10}$$

Expanding (10) and considering that $\varkappa = x_1 - x_2, \varsigma = y_1 - y_2, \varrho = \nu_1\cos\phi_1 - \nu_2\cos\phi_2, \varphi = \nu_1\sin\phi_1 - \nu_2\sin\phi_2$
and we get

$$t^2(\varrho^2 + \varphi^2) + 2t(\varkappa\varrho + \varsigma\varphi) + (\varkappa^2 + \varsigma^2 - R_m^2) = 0 \tag{11}$$

The connection time ($\Delta t_{i,j}$) between the sensor nodes is expressed in (12) by solving (11) with respect to $t$. The negative value of connection time shows that there is no link between them

$$\Delta t_{i,j} = \frac{-(\varkappa\varrho + \varsigma\varphi) \pm \sqrt{R_m^2(\varphi^2 + \varsigma^2) - (\varkappa\varphi - \varsigma\varrho)^2}}{\varrho^2 + \varphi^2} \tag{12}$$

The average life time of the cluster $T_{LT}(t)$ (i.e. the average connection time between the TCH and member nodes in each cluster) is expressed as

$$T_{LT}(t) = \frac{1}{N_{\text{CM}}} \times \sum_{n=1}^{N_{\text{CM}}} \Delta t_{i,j}(n) \tag{13}$$

where $N_{\text{CM}}$ is the number of CM.

The designed algorithm should satisfy the conditions (i) $T_{LT}(t) > C_{ld}(t)$, (ii) $\Delta t_{i,j} > L_{i,j}(t)$ to achieve high stable cluster formation, thereby increasing the network packet delivery ratio, and network lifetime. The normalised value of the average connection time is computed based on the following equation:

$$\xi_{\text{ACT}} = \frac{\sum_{n=1}^{\hat{N}_{\text{TCH}}} \Delta t_{i,j}(n)}{\hat{N}_{\text{TCH}} \times \eta_i} \tag{14}$$

where $\Delta t_{i,j}(\cdot)$ is the connection time between the nodes, $\hat{N}_{\text{TCH}}$ is the number of TCH's neighbours. Here, $\eta_i = \max(\Delta t_{i,j}(n))_{\hat{N}_{\text{TCH}}}$ which is used to select a maximum connection time [other than that of the nodes (TCH and CM] having the same speed and moving direction) from its members. If $v_1 = v_2$, $\theta_1 = \theta_2$, then $\Delta t_{i,j}(\cdot)$ is infinite, i.e. the connection time $\Delta t_{i,j}(\cdot)$ between two nodes continuously stays constant until the end. In such a case, the connection time of infinite duration is assumed as maximum of one ($\Delta t_{i,j}(\cdot) = 1$).

(b) *Number of neighbour nodes*

The number of CMs in each cluster play a vital role to stipulate the lifespan and load balance of the CHs. The normalised value of the number of neighbour nodes is calculated as

$$\xi_{\text{NN}} = \frac{\hat{N}_{\text{TCH}} \bmod (M_{\text{CM}} + 1)}{M_{\text{CM}}} \tag{15}$$

The consideration of $\xi_{\text{NN}}$ reduces the chance value (i.e. increases the delay time of broadcasting of the final CH awareness message) of TCH to become a final CH when the number of CM nodes in each cluster goes beyond or goes very low within the defined level, i.e. $M_{\text{CM}}$.

The TCH candidate calculates the delay time ($B_{\text{wait}}(t)$) for broadcasting the final CH awareness message by using the following equation:

$$B_{\text{wait}}(t) = (1 - (\xi_{\text{ACT}} \times \xi_{\text{NN}})) \times t_{\text{cf}} + \delta_r \tag{16}$$

where $t_{\text{cf}}$ is the total duration for cluster formation, $\delta_r$ is the random time frame which is very small ($\delta_r \ll t_{\text{cf}}$) utilised to minimise the contention in the channel accessing.

The TCH becomes either CH or CM based on (16). The TCH waits a time $B_{\text{wait}}(t)$ before it broadcasts the final CH awareness message. Once the waiting time is expired, then it broadcasts the CH awareness message to neighbours and also establishes the cluster. It means that the high chance value has a low waiting time, i.e. it has high connection time and optimum number of neighbours. If any TCH receives the CH awareness message from other candidate nodes before the expiry of its $B_{\text{wait}}(t)$, then the TCH gives up its competition and joins as a CM. Upon receiving this message, the normal nodes also send the registration message to the concerned CH to join as a CM. The CH assigns the TDMA schedule to the CM depending on the connection time in ascending

order. Then, CHs collect the data from its member nodes with respect to the assigned TDMA schedule.

*3.2.2 Construction of cluster switching.:* We assume that the targeted sensing area is equally clustered by the area of regular hexagon (i.e. the coverage area of the CH). The numbers (1, 2, 3) in Fig. 4 indicate the CSS.

In CSDGP, the CSS-MESSAGE (CSS-MSG) plays a vital role in assigning a CSS and tree topology organisation. The BS initiates the broadcasting of the CSS-MSG by selecting the nearest CH and assigns the CSS to it. After that, the CHs alone broadcast CSS-MSG up to its two-hop distance (i.e. twice its transmission range) by using CSMA/CA. The CSS-MSG consists of its ID, location, CSS, the moving direction, and speed. Fig. 5 depends on (18) and

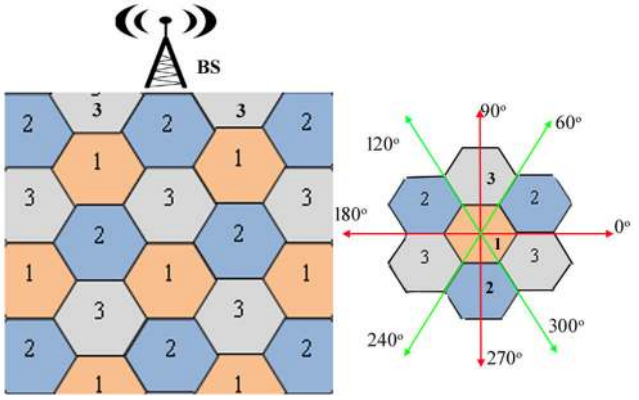


**Fig. 4** *Organisation of cluster switched mechanism*

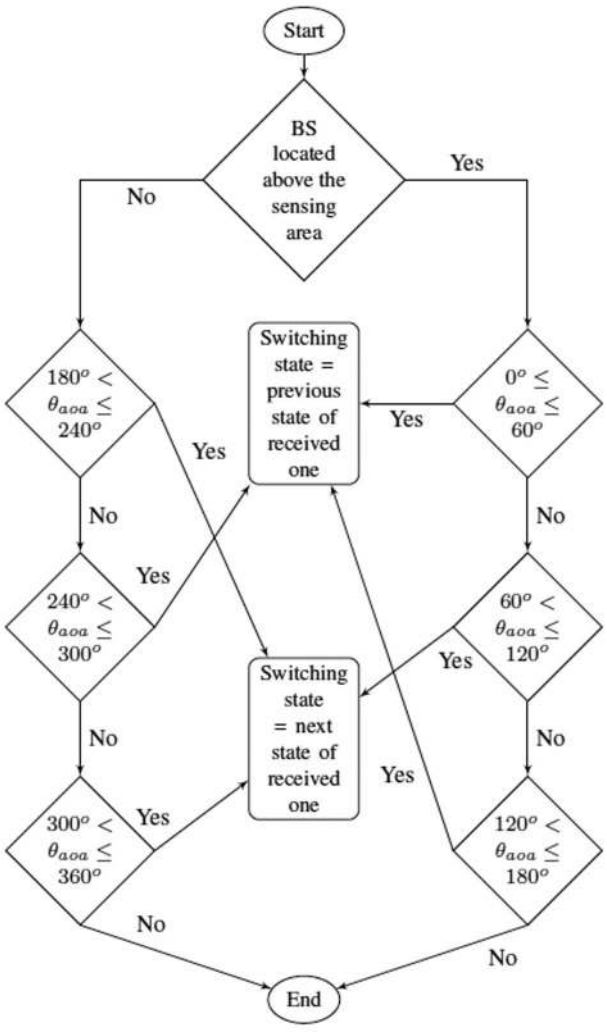


**Fig. 5** *Selection of CSS*

Fig. 4. By using the location information received in the CSS-MSG packets, each CH finds the distance from its neighbour CHs and also calculates the angle of signal arrival, as given in the following equations:

$$d_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{17}$$

$$\theta_{aoa} = \tan^{-1} \frac{y_j - y_i}{x_j - x_i} \tag{18}$$

where $(x_i, y_i)$ and $(x_j, y_j)$ are the positions of the transmitting node and receiving node, respectively, $d_{i,j}$ is the distance between the transmitting and receiving node, $\theta_{aoa}$ is the angle of arrival of the transmitted signal.

Based on the angle of arrival of the signal, the CH knows itself the switch state of the cluster based on Fig. 5, when the BS is located at the top of the sensing area. If any CH receives multiple CSS-MSG from the various CHs, then it selects the closest CH and its CSS-MSG among them for finding the CSS. For example, the $CH_4$ receives the CSS-MSG from $CH_2$ and $CH_3$ respectively. But, the $CH_4$ selects the CSS-MSG of $CH_2$ based on the distance between them, and also finds its CSS with respect to the $CH_2$ as depicted in Fig. 6, thus it attains the converged value in the CSS quickly.

*3.2.3 Adaptive TDMA scheduling.:* The proposed time slot structure is shown in Fig. 7, in which the total time frame is divided into the intra-cluster frame and inter-cluster frame. The total time for intra-cluster communication is considered as $t_{iAc}$ and inter-cluster communication is $t_{iEc}$. The total time frame of intra-cluster is further divided into three CSS frames (CSS 1, CSS 2, CSS 3) and the duration for each switch state is $t_{cs}$. Based on the switch state, the CH assigns the TDMA schedule to the CM and one more slot is reserved for collecting the data from the migrated node (MN). In CSDGP, if any sensor node leaves the current cluster and enters into the new one before sending the data to the concerned CH, then it has the chance to send the collected data to the newly entered region's CH. In this dynamic slot allocation, the CH and CM can estimate and maintain the information about the connection time between them. The CH allocates the TDMA slot based on the connection time in ascending order to its member nodes which remain in the same cluster till the end of $t_{iAc}$. If any sensor node does not remain in the same cluster, then the CH removes the sensor node from the TDMA schedule. These eliminated nodes also know about the connection duration with its current CH, once its connection is expired (moved out of its current cluster region), it starts broadcasting a cluster join request message. Upon receiving this message, the CH of newly entered cluster accepts the request and adjusts its TDMA schedule and then broadcasts the new time slot to its member.

During the intra-cluster communication, the switch state of the CHs is decided based on the CH location in that sensing area. For example, if one of the CH is in an active state, then its neighbours

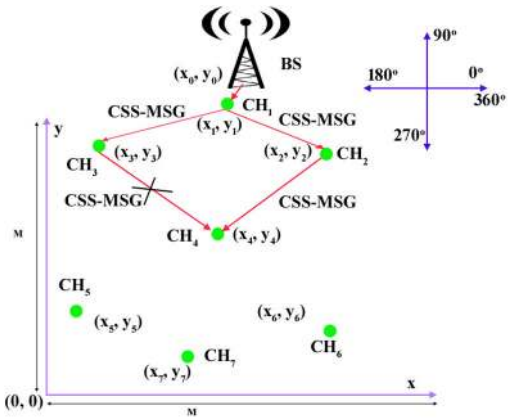


**Fig. 6** *Processing and selection of CSS-MSG*

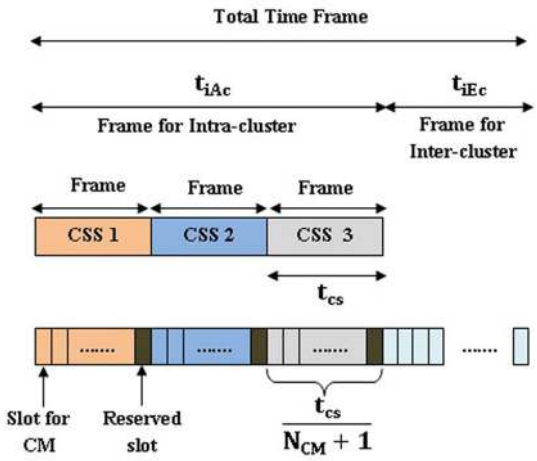


**Fig. 7** *CSDGP time slot structure*

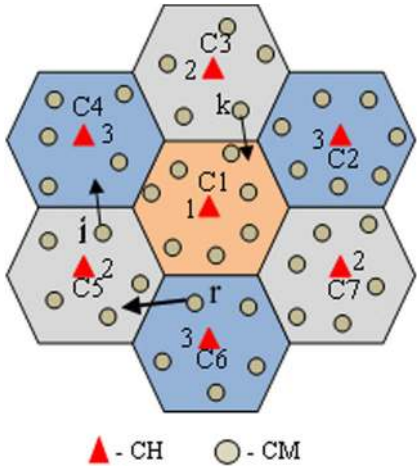


**Fig. 8** *Dynamic slot allocation for mobile nodes*

| | | | | | | |
|---|---|---|---|---|---|---|
| Old time slot | Node m | Node 0 | Node l | Node w | Node s | Reserved |
| New time slot | Node m | Node 0 | Node l | Node w | Node s | Node i |

**Fig. 9** *Utilisation of the reserved slot*

| | | | | | | |
|---|---|---|---|---|---|---|
| Old time slot | Node n | Node b | Node d | Node x | Node i | Reserved |
| New time slot | Node n | Node b | Node d | Node x | Node r | Reserved |

**Fig. 10** *Utilisation of the emigrated node slot*

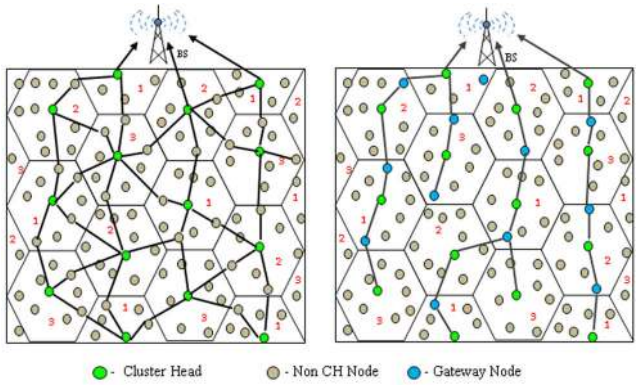


**Fig. 11** *Tree topology formation and final network structure of CSDGP*

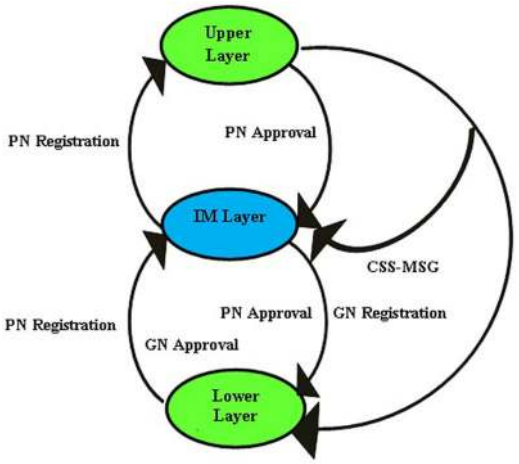


**Fig. 12** *Flow diagram of the tree topology organisation*

go to the idle state. As shown in Fig. 8, the CH of the cluster 1 (C1) is in an active state and its neighbour CHs (i.e. the CH of C2, C3, C4, C5, C6, C7) go to the idle state during the CSS of 1. Similarly, the CHs of C3, C5, and C7 are in an active state and the remaining neighbour CHs (i.e. the CH of C1, C2, C4, C6) go to the idle state during the CSS of 2. And also, the CHs of C2, C4, and C6 are in an active state and the remaining neighbour CHs (i.e. the CH of C1, C3, C5, C7) go to the idle state during the CSS of 3. Thus, only one CH is allowed with respect to the neighbour CHs to receive the data from the CM according to the allotted TDMA time schedule, thereby evading the unwanted signal from the neighbour clusters.

Figs. 9 and 10 depict the process of allocating the TDMA schedule to the newly entered node, in which there is a possible way in assigning the time slot to the MN with respect to Fig. 8 as mentioned below:

*Case (i)*

In this case, the sensor node *i* leaves from the cluster 5 (*C*5) and joins into the new cluster 4 (*C*4). The CH of cluster 4 adds the MN *i* into the reserved slot of the TDMA schedule as shown in Fig. 9.

*Case (ii)*

The sensor node *i* leaves from the cluster 5 (*C*5) and enters into the new cluster 4 (*C*4). Thus, the CH of cluster 5 removes the node *i* and utilises the emigrated node (node *i*) slot to the MN *r* as shown in Fig. 10.

*Case (iii)*

If the CH of any cluster area has completed the data collection from the CM before the expiry of the intra-cluster communication duration then it accepts the request of migration and assigns a time slot.

For example, the CH of cluster area 1 (*C*1) has completed the data collection from the CM in the CSS of 1, thus it immediately accepts the request of the node *k* and assigns a new slot to collect the data.

### *3.3 Routing organisation*

In CSDGP, the routing organisation process is initiated by the BS. The targeted sensing area is divided into the layers based on the position of the CHs and its transmission range. Fig. 11 depicts the tree topology formation and the final structure of CSDGP. Based on the CSS-MSG message, the nodes between CHs of the lower layer (layer 1) and the BS of the upper layer (layer 0) acts as a gateway node (GN) and forms the intermediate (IM) layer. By this way, the CHs of layer 1 now acts as the upper layer that broadcasts CSS-MSG to its two-hop distance neighbours to divide the network area further. The same process is continued until all the CHs including orphan CHs are covered in the routing topology. The flow diagram of the tree topology organisation is described in Fig. 12.

Once, the network partitioning is completed, the sensor nodes in the intermediate layers use a joint strategy including the distance, residual energy, and velocity in order to find their cost value $T_{ci}(n)$,

as expressed in (23). The parameters $\vartheta_{w1}$ for selecting the highest residual energy node, and $\vartheta_{w2}$ for choosing the low-velocity node, $\vartheta_{w3}$ and $\vartheta_{w4}$ are used for finding the shortest path between the CHs and BS. Then, these sensor nodes send the GN registration message which contains Node_ID, upper layer_ID, and lower layer _ID, different combinations of the cost values, location, residual energy level, and velocity. The CH of the lower layer selects the highest cost value from the registrations and gives an approval to act as a gateway node between the concern CHs or between the CH and BS.

The parameters $\vartheta_{w1}$ and $\vartheta_{w2}$ are calculated by using the following equations:

$$\vartheta_{w1} = \frac{E_c}{E_m} \tag{19}$$

$$\vartheta_{w2} = \frac{V_m - V_c}{V_m} \tag{20}$$

where $V_m$ is the maximum velocity of the sensor node, $V_c$ is the current velocity of the sensor node.

Equation (21) for finding the parameter $\vartheta_{w3}$ is based on Fig. 13. Each node finds the value to act as a GN and also ensures that the selected GN has shortest path to reach the BS

$$\vartheta_{w3} = \frac{d_{CB_i} - d_{CB_j}}{2R_m} \tag{21}$$

$d_{CB_i}$ is the distance between CH node $i$ and BS, $d_{CB_j}$ is the distance between CH node $j$ and BS.

Equation (22) is derived based on Fig. 14.

The parameter $\vartheta_{w4}$ is considered by the sensor node to act as a GN and also ensures that the selected GN has straight path to reach the BS

$$\vartheta_{w4} = \frac{d_{C_{i,j}}}{(d_{CS_i} + d_{CS_j})} \text{ (or) } \frac{d_{CB_i}}{(d_{CS_i} + d_{SB})} \tag{22}$$

where $d_{CS_i}$ is the distance between CH node $i$ and SN, $d_{CS_j}$ is the distance between CH node $j$ and SN, $d_{SB}$ is the distance between BS and SN, $d_{C_{i,j}}$ is the distance between CH node $i$ and CH node $j$, $d_{CB_i}$ is the distance between CH node $i$ and BS

$$T_{ci}(n) = \vartheta_{w1} \times \vartheta_{w2} \times \vartheta_{w3} \times \vartheta_{w4} \tag{23}$$

The farthest child node of the tree initiates the selection of the parent node which depends on the residual energy, velocity, and distance to the BS. $\vartheta_{w5}$ is used to find the parent node which has the shortest path towards the BS and it is given in (24) which is derived based on Fig. 15

$$\vartheta_{w5} = \frac{|d_{CB_i} - d_{GB_i}|}{R_m} \text{ (or) } \frac{|d_{GB_j} - d_{CB_j}|}{R_m} \tag{24}$$

where $d_{CB_i}$ is the distance between CH node $i$ and BS, $d_{GB_i}$ is the distance between GN node $i$ and BS, $d_{CB_j}$ is the distance between CH node $j$ and BS, $d_{GB_j}$ is the distance between GN node $j$ and BS.

The GN or CH finds the parent node with high-cost values based on (25). This ensures that the selected node has a low work burden than other neighbour CHs or GNs. Then, these nodes send the PN registration message (Node_ID, Parent Node_ID) to the selected parent node in the upper layer in order to get the confirmation

$$T_{cp}(n) = \vartheta_{w1} \times \vartheta_{w2} \times \vartheta_{w5} \tag{25}$$

After establishing the tree topology between the CHs through the gateway node, the CH gathers the sensed data from its member with respect to the TDMA schedule during its CSS of the $t_{iAc}$. At the end of the intra-cluster communication, the CHs start to send their collected data packets to the BS through the parent nodes. The child node uses a CSMA-CA technique to access the channel and forward the data to the parent without any collision.

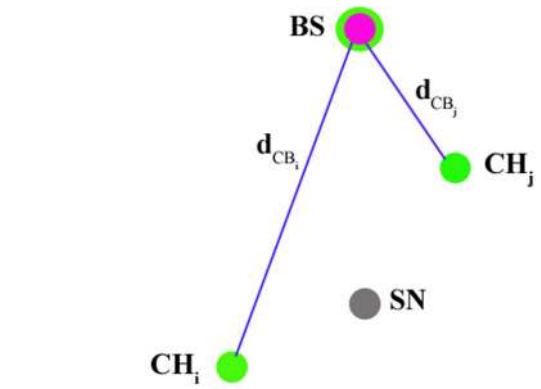


**Fig. 13** *Calculation of* $\vartheta_{w3}$

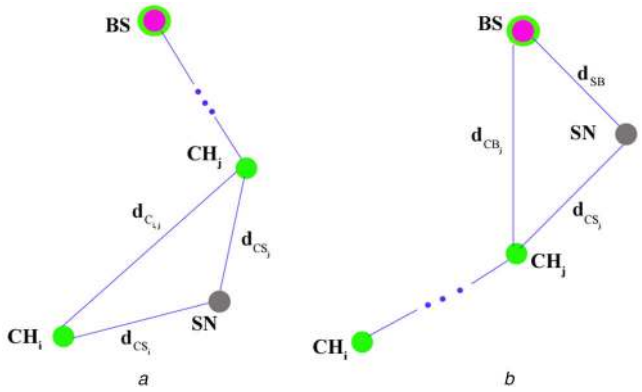


**Fig. 14** *Calculation of* $\vartheta_{w4}$

***(a)*** $\vartheta_{w4}$ considered between the CHs, ***(b)*** $\vartheta_{w4}$ considered between the CH and BS

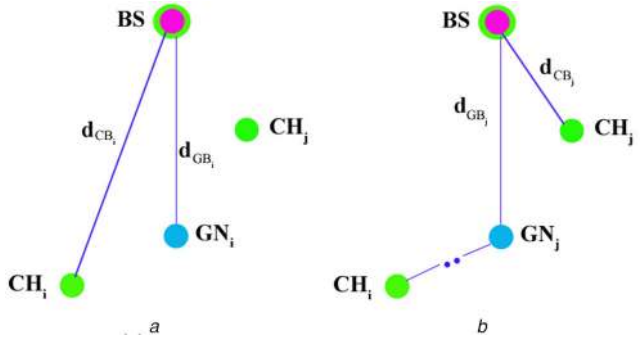


**Fig. 15** *Calculation of* $\vartheta_{w5}$

***(a)*** $\vartheta_{w5}$ considered from CH to GN, ***(b)*** $\vartheta_{w5}$ considered from GN to CH

### *3.4 Fault tolerance mechanism*

During the inter-cluster communication, CH or GN in the tree topology will not receive the acknowledgment from the upper-level node (i.e. PN) within the timeout. The lower-level node (CH or GN) selects the new upper-level node by: (i) if connectivity failure occurs from CH to GN, then CH reestablishes the link connectivity by sending CSS-MSG towards the BS; (ii) if connectivity failure from GN to CH, then GN finds new PN by sending the PN registration message. By this way, it reduces the packet loss and achieves consistency in successful packet delivery ratio.

## 4 Mathematical analysis of CSDGP

### 4.1 Energy consumption

In this paper, a simple radio transceiver model [28] is considered, where, $\epsilon$ is the energy consumption of amplifier, $\alpha$ is the amplification factor for multi-path or free space, $E_e$ is the energy consumption of the radio electronics, $k$ is the size of data packets, $d$ is the distance between the nodes. Therefore, the energy spent for transmitting and receiving the data is given in the following equations:

$$E_t(k,d) = k(E_e + \epsilon \times d^{\alpha}) \tag{26}$$

$$E_r(k) = k \times E_e \tag{27}$$

The average energy consumption of network is calculated based on the following equations:

$$E_D^{\mathrm{CM}} = (N_a - N_{CH}) \times (E_s + E_t(k,d)) \tag{28}$$

where $E_D^{\mathrm{CM}}$ is the energy consumed by CM for transmitting the data, $E_s$ is the energy consumed by the sensor node for sensing, $N_{CH}$ is the number of CHs in each round

$$E_{\mathrm{c}}^{\mathrm{CM}} = (N_a - N_{CH}) \times (E_t(L_c, R_m) + 2E_r(L_c)) \tag{29}$$

where $E_{\mathrm{c}}^{\mathrm{CM}}$ is the energy consumed by CM for transmitting and receiving the control packets, $L_c$ is the size of the control packet

$$E_T^{\mathrm{CM}} = E_D^{\mathrm{CM}} + E_{\mathrm{c}}^{\mathrm{CM}} \tag{30}$$

$E_{\mathrm{Total}}^{\mathrm{CM}}$ is the total energy consumed by the CM

$$E_{\mathrm{c}}^{\mathrm{SN}} = N_a \times (E_t(L_c, R_m) + N_{nn} \times E_r(L_c)) \tag{31}$$

$E_{\mathrm{c}}^{\mathrm{SN}}$ is the energy consumed by the sensor nodes for exchanging the node discovery message, $N_{nn}$ is the neighbours of sensor node

$$E_D^{\mathrm{GN}} = N_{GN} \times (N_{dp} \times ((E_t(k,d) + E_r(k)))) \tag{32}$$

where $E_D^{\mathrm{GN}}$ is the energy consumed by GN for transmitting and sensing the data, $N_{dp}$ denotes the number of processed data packets, $N_{GN}$ is the number of gateway node in each round

$$E_{\mathrm{c}}^{\mathrm{GN}} = N_{GN} \times (2E_t(L_c, R_m) + 2E_r(L_c)) \tag{33}$$

where $E_{\mathrm{c}}^{\mathrm{GN}}$ is the energy consumed by GN for transmitting and receiving the control packets

$$E_T^{\mathrm{GN}} = E_D^{\mathrm{GN}} + E_{\mathrm{c}}^{\mathrm{GN}} \tag{34}$$

where $E_T^{\mathrm{GN}}$ is the total energy consumed by the gateway node

$$\begin{aligned} E_D^{\mathrm{CH}} = N_{CH} \times (N_{\mathrm{CM}} \times E_r(k) + E_s) + E_t(k,d) \\ + N_{dp} \times (E_t(k,d) + E_r(k)) \end{aligned} \tag{35}$$

$E_D^{\mathrm{CH}}$ is the energy consumed by CH for data communication and sensing the data, $N_{\mathrm{CM}}$ is the number of CM nodes

$$\begin{aligned} E_{\mathrm{c}}^{\mathrm{CH}} = N_{CH} \times (4E_t(L_c, R_m) + N_{\mathrm{CM}} + 3) \\ \times E_r(L_c) + E_t(L_c, 2R_m) \end{aligned} \tag{36}$$

$E_{\mathrm{c}}^{\mathrm{CH}}$ is the energy consumed by CH for transmitting and receiving the control packets

$$E_T^{\mathrm{CH}} = E_D^{\mathrm{CH}} + E_{\mathrm{c}}^{\mathrm{CH}} \tag{37}$$

where $E_T^{\mathrm{CH}}$ is the total energy consumed by the CH

$$E_{avg} = \frac{E_T^{\mathrm{CM}} + E_T^{\mathrm{CH}} + E_T^{\mathrm{GN}} + E_{\mathrm{c}}^{\mathrm{SN}} + (E_{id} \times N_a)}{N_a} \tag{38}$$

where $E_{\mathrm{avg}}$ is the average energy consumption of the network, $E_{id}$ is the energy dissipation due to the node in idle state.

### 4.2 Control overhead

In each round, $N_a$ of HELLO packets are broadcasted for discovering the neighbour nodes (i.e. $O(T_a) = N_a$)

In cluster formation, there are $O(CH_{ca}) \simeq eqN_{\mathrm{m}} < N_{CH} \leq 2N_{\mathrm{m}}$ of CH awareness messages, $O(JR) = (N_a - N_{CH})$ of cluster join request messages and $O(CH_{sc}) = N_{CH}$ of schedule messages. Then, the total control overhead $O(T_c)$ involved in the cluster formation is determined by using the following equations:

$$O(T_c) = O(T_a) + O(N_a) + O(CH_{ca}) + O(JR) + O(CH_{sc}) \tag{39}$$

$$O(T_c) = 2N_a + N_{CH} \tag{40}$$

CSDGP has message complexity of O($T_c$) which is similar to that of VELCT protocol. Thus, the proposed CSDGP does not have message complexity in the system design. During the tree construction, $O(CH_{CSS-MSG}) = N_{CH}$ of layer separation messages, $O(GN_{rm}) = N_a - N_{CH}$ of gateway requisition messages, $O(GN_{am}) = N_{CH}$ of gateway approval messages, $O(PN_{rm}) = N_{tol} - 1$ of parent node requisition messages, and $O(PN_{am}) = N_{tol} - 2^{d_l} + 1$ of parent node approval messages are utilised, where $d_l$ is the depth of the layer. Because, the number of gateway nodes is $N_{GN} = \sum_{n=1}^{d_l} 2^{n-1}$ and the total number of nodes involved in tree topology is $N_{tol} = N_{GN} + N_{CH}$. The total control overhead of tree formation is computed, as given in the following equations:

$$\begin{aligned} O(T_t) = O(CH_{CSS-MSG}) + O(GN_{rm}) + O(GN_{am}) \\ + O(PN_{rm}) + O(PN_{am}) \end{aligned} \tag{41}$$

$$O(T_t) = N_{CH} + N_a + 2N_{tol} - 2^{d_l} \tag{42}$$

where $d_l$ is the depth of the layer and $N_{CH}$ is the total number of current CH. The overall network control overhead $O(C_{net})$ is determined by using the following equations:

$$O(C_{net}) = O(T_a) + O(T_t) + O(T_c) \tag{43}$$

$$O(C_{net}) = 2N_{CH} + 3N_a + 2N_{tol} - 2^{d_l} \tag{44}$$

### 4.3 Time complexity

The proposed CSDGP adapts distributed clustering strategy to find the optimum number of CHs by broadcasting the final CH awareness message based on $B_{\mathrm{wait}}(t)$. If it is assumed that all the deployed sensor nodes act as TCH in the CH selection, then it takes more time to end up the clustering process. To solve this problem, only $\chi N_m$ number of sensors is selected from $N_a$ which is participating as TCH in each round. Thus, the number of TCHs ($\hat{N}_{\mathrm{TCH}}$) depends on the $\chi$ and $N_m$, therefore $N_{CH}$ lies between $N_m$ and $\chi N_m$, i.e. $N_m \leq N_{CH} \leq \chi N_m$. It is understood that the time complexity for the proposed cluster formation depends on $\chi N_m$ and $t_{\mathrm{cf}}$, i.e. $O(\hat{N}_{\mathrm{TCH}} \,.\, t_{\mathrm{cf}})$. In CSDGP, $t_{\mathrm{cf}}$ is a fixed value ($c$) in the cluster formation, therefore the time complexity is $O(c \,.\, \hat{N}_{\mathrm{TCH}})$ which is approximately equal to $O(\hat{N}_{\mathrm{TCH}})$.

**Table 2** Simulation setup

| Simulation parameters | Values |
|---|---|
| targeted sensing network area | $800 \times 800\ m^2$ |
| number of sensor nodes | 100--500 |
| velocity of the sensor node | 0–50 m/s |
| transmission range of the antenna | 80 m |
| bit rate | 50 kbps |
| initial energy of sensor nodes | 10 J |
| data packet size [7] | 512 bytes |
| control packet size [7] | 25 bytes |
| $E_{elec}$ [7] | 50 nJ/bits |
| $E_{amp}$ [7] | $1.3\ fJ/bits/m^4$ |
| $E_{aggr}$ [7] | 5 nJ/bits/signal |

The features of the CSDGP are as follows:

(i) The minimum number of nodes ($N_m$) required to cover the entire network is calculated by using (1) and (2), and only twice the $N_m$ acts as a TCH, thus it achieves the optimum number of CHs (i.e. the number of final CHs lies between $N_m$ and $2N_m$) and also reduces the convergence delay.
(ii) The TCH selection in (3) helps in obtaining high residual energy node to act as a CH in each round and finds the optimum number of CHs.
(iii) The waiting time based CH selection ensures the uniform distribution of CHs. Therefore, all the sensor nodes receive the CHs awareness message, resulting in avoidance of the isolated nodes.
(iv) It helps in good load balancing of the CHs, as the CHs having the optimum number of neighbours are selected.
(v) The ratio of the remaining battery energy to the maximum energy of the sensor node at the time of deployment is considered in the TCH selection. This helps in avoiding the low energy sensor nodes to be selected as a CH, extending the lifespan of sensor nodes.
(vi) The ratio of the average connection time with neighbours to the maximum connection duration with neighbour helps in ensuring the cluster stability (i.e. connectivity between the CH and CMs).
(vii) As similar to LEACH, the control overhead of the CSDGP is $O(n)$, even the CH selection is performed in two phases. Thus, the designed protocol does not increase the complexity in the system design.
(viii) The control packet (CSS-MSG) alone is used to determine the tree topology and CSS, minimising the additional control overhead.
(ix) In CSDGP, the orphan CHs utilise the tree topology for transmitting its sensed data to the BS (multihop fashion) instead of the direct communication, conserving the significant amount of battery energy.
(x) The selection of GN and PN based on the residual energy, speed, and distance helps in forming the shortest path efficiently to reach the destination and also prolongs the network lifetime.
(xi) The consideration of speed as a routing metrics in PN and GN selection helps in enhancing the network stability and reduces the link failure in the route. Therefore, it lessens the packet loss and energy consumption.
(xii) The CSS-based data collection mechanism helps in allowing the concurrent data collection by the CHs without any data collision and loss.
(xiii) The TDMA schedule assigned to the CMs based on the connection time in ascending order helps in avoiding simultaneous data transmission and collision.

## 5 Results and discussion

The Network Simulator version 2 (NS-2) is used to validate the performance efficacy of the proposed CSDGP along with other existing cluster based protocols like ARBIC, VELCT, MBC, and LEACH-M. The simulation parameters of our proposed model are mentioned in Table 2.

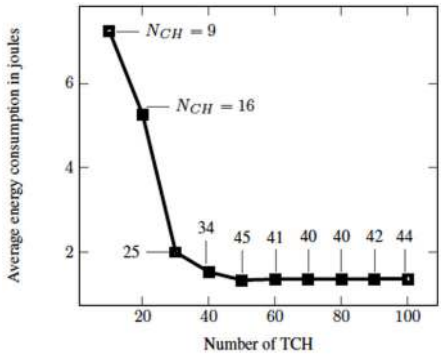


**Fig. 16** *Average energy consumption versus number of TCH*

### 5.1 Analysis of energy consumption versus TCH

The result shown in Fig. 16 is obtained for 300 deployed nodes by setting the simulation time of 180 s and the node's speed is varied from 0 to 30 m/s.

Fig. 16 shows that the proposed CSDGP is the most energy efficient, i.e. it maintains the minimum number of sensor nodes required for covering the entire sensing area. And also, it can provide consistent performance when there is an increase in the number of TCH. It means that the sufficient number of CHs is maintained which is needed to cover the entire area even when the number of TCH increases. Initially, the energy consumption is high as shown in Fig. 16 which is due to the insufficient number of CH causing the coverage problem and isolated nodes. Further, the results are analysed by increasing the number of TCHs. It is clear that the proposed protocol maintains the optimum energy level of 1.37 J when the number of CHs ranges from 40 to 45. However, the complexity of the CSDGP to setup the cluster depends on $t_{cf}$ and the number of TCH. In order to reduce the time complexity, the multiplication factor ($\chi$) is introduced in the CSDGP which is used to set the number of TCH based on the requirement.

### 5.2 Numerical analysis of CSDGP with existing protocols

Table 3 shows the various network parameters related to the impact on the network performances. In general, the energy consumption of the networks mainly depends on three parameters. (i) High CM: due to the unbalanced workload of the CHs depletes their battery energy very quickly; (ii) Transmission range: energy consumption of the sensor nodes is directly related to the distance. For example, LEACH-M protocol directly communicates with BS and MBC performs the data communication through CH to CH, thus both LEACH-M and MBC consume higher energy as compared to ARBIC, VELCT and the proposed CSDGP; (iii) Size of the network area: as the size of the network increases, either the

**Table 3** Numerical analysis of the proposed CSDGP with existing clustering protocols

| Protocols | Number of control packets used in cluster formation | Number of nodes involved in the inter-cluster communication | Energy consumption of CH for inter-cluster data communication | Energy consumption of CH for intra-cluster data communication | Time complexity for constructing the cluster topology |
|---|---|---|---|---|---|
| Proposed CSDGP | $2N_a + N_{CH}$ | $N_{CH} + \sum_{n=1}^{d_l} 2^{n-1}$ | $E_r(k, d)$ | $N_{\mathrm{CM}} \times E_r(k)$ | $O(\hat{N}_{\mathrm{TCH}})$ |
| ARBIC [9] | $2(N_a + N_{CH})$ | $N_{CH} + \sum_{n=1}^{d_l} 2^{n-1}$ | $E_r(k, d)$ | $N_{\mathrm{CM}} \times E_r(k)$ | $O(n_{\mathrm{gen}})$ |
| VELCT [26] | $2N_a + N_{CH}$ | $N_{CH} + \sum_{n=1}^{d_n} 2^{n-1}$ | $E_r(k, d)$ | $N_{\mathrm{CM}} \times E_r(k)$ | $O(N_a)$ |
| MBC [25] | $2N_a$ | $N_{CH}$ | $E_r(k, 2d)$ | $N_{\mathrm{CM}} \times E_r(k)$ | $O(N_{CH})$ |
| LEACH-M [18] | $2N_a$ | $N_{CH}$ | $E_r(k, d_{CB})$ | $N_{\mathrm{CM}} \times E_r(k)$ | $O(N_{CH})$ |

transmission range or number of node also increases which may deplete the battery energy quickly.

And also, the control overhead of the network depends on five parameters: (i) methodology applied in the cluster formation: the utilisation of the optimum number of CHs minimises the control packets; (ii) isolated nodes: due to the left out, i.e. uncovered by the CHs, the sensor node continuously searches a CH by broadcasting join request message or directly communicates with the BS which increases the number of control packets and energy consumption; (iii) mobility: as the speed of node increases the usage of the number of control packets also increases which is due to often occurrence of the link failure and network instability and also due to the searching of new CH to send the data by broadcasting join request message; (iv) node density: the utilisation of the number of control packets is directly proportional to the number of nodes; (v) the number of nodes in the tree topology: due to the mobility environment, the nodes in the tree topology often check their link connectivity by sending the message causing additional control overhead. From the simulation result, LEACH-M, MBC, and VELCT show high control overhead which is mainly due to the isolated nodes and link failure.

The number of nodes involved in the proposed CSDGP and the existing ARBIC depends on the number of CHs and depth of the layer ($d_l$) based on the connectivity between the CHs, whereas in the VELCT, it depends on the depth of the layer ($d_n$) based on the size of the network and the transmission range of the sensor node.

The time complexity of the ARBIC was $O(\max_{\mathrm{gen}} . n_{\mathrm{gen}})$ which depends on the maximum number of iterations ($\max_{\mathrm{gen}}$) and number of candidates ($n_{\mathrm{gen}}$). The $\max_{\mathrm{gen}}$ is assumed to be fixed value $c$, therefore time complexity is $O(c . n_{\mathrm{gen}})$ that is, approximately equal to $O(n_{\mathrm{gen}})$. All the deployed nodes in VELCT calculated the threshold value and transmitted the CH awareness message based on the CSMA/CA. And, only fixed number of nodes was allowed to calculate the threshold value to act as CH in each round in MBC and LEACH-M.

### *5.3 Impact of the speed of sensor nodes*

For the simulation analysis in Figs. 17–20, the speed of sensor nodes is varied from 0 to 10, 20, 30, 40 and 50 m/s, the number of sensor nodes is 300 and the simulation run time is 300 s.

From Fig. 17, it is noted that the average end-to-end delay of CSDGP is 15.64, 17.21, 19.17, and 28.27% lower than ARBIC, VELCT, MBC, and LEACH-M when the speed of the sensor node reaches the maximum, i.e. 50 m/s. Since the CSDGP includes the connection time for electing the CHs and fault tolerance mechanism utilised in the tree topology ensuring the stability of the routing topology. Also, the data communication uses a cluster switch mechanism which effectively reduces the contentions in the MAC level. In addition, the estimation of the minimum number of sensor nodes ensures the link available to all the nodes.

In Fig. 18, the average energy consumption of CSDGP is reduced by 10.03, 12.11, 14.1, and 17.64 % than the existing ARBIC, VELCT, MBC, and LEACH-M, respectively, when the speed of the sensor node reaches the maximum. This is because of the proposed mechanism primarily considers the residual energy of the sensor node while electing the TCH. In addition, the CHs and nodes involved in the tree topology are made to change in each and every round which ensures the uniform energy consumption.

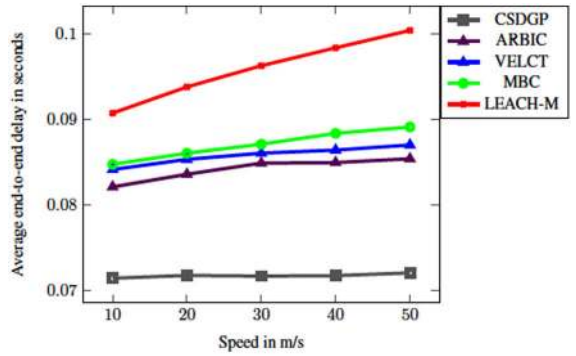


**Fig. 17** *Average end-to-end delay versus speed in m/s*

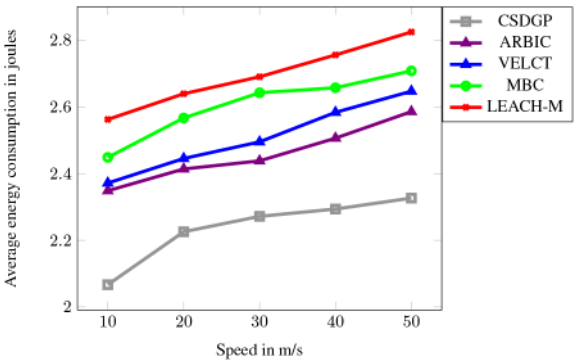


**Fig. 18** *Average energy consumption versus speed in m/s*

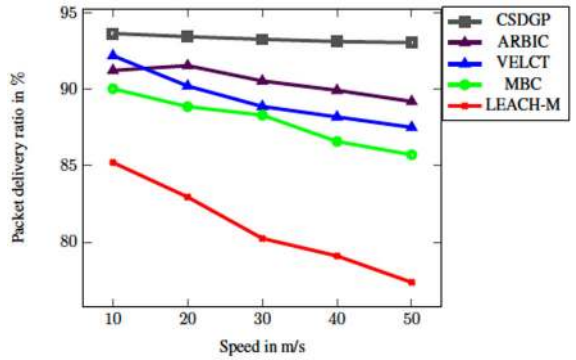


**Fig. 19** *Packet delivery ratio versus speed in m/s*

Moreover, effective utilisation of the maximum transmission power is ensured by the consideration of $\vartheta_{w3}$, thereby reducing the number of nodes involved in the tree formation. And also, it minimises the number of data transmission and reception resulting in reduced interference and energy consumption. From Fig. 19, it is observed that the PDR of CSDGP is 4.14, 5.97, 7.89, and 16.88% higher than the existing ARBIC, VELCT, MBC, and LEACH-M, respectively, when the speed of the sensor node reaches the maximum. This is because of CSDGP ensures a stable routing path between the CM to CH and CH to BS. Furthermore, the CSDGP uses a cluster switch mechanism for minimising the data

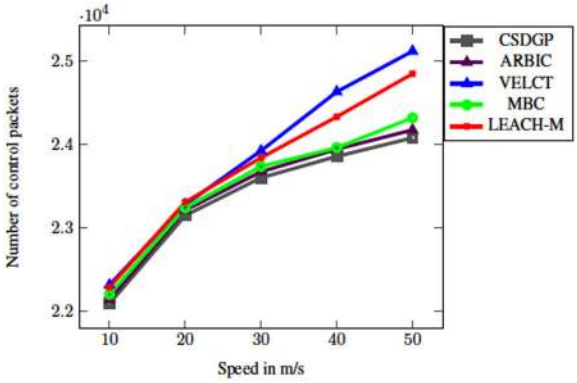


**Fig. 20** *Control overhead versus speed in m/s*

retransmission. And also, it efficiently collects the data from the MNs by using the dynamic slot adjustment. And also, the speed of the sensor node is considered as one of the parameters in tree topology construction in CSDGP which ensures the topology stability and avoids the occurrence of frequent link failures, which in turn results in reduced packets loss. Thus, the CSDGP consistently provides a PDR above 93%.

Fig. 20 depicts that the proposed CSDGP used a less number of control packets as compared to the existing ARBIC, VELCT, MBC, and LEACH-M. The estimation of connection time in the proposed protocol based on the $x$, $y$ coordinate values, moving speed and direction ensures the network stability. And also, the node's speed and distance to the BS are utilised to select the parent and gateway nodes in the tree topology that avoids the link failures in the routing and finds the shortest path to reach the BS. Thus, it minimises the number of control packets utilised in the CSDGP.

### *5.4 Impact of the number of sensor nodes*

The results shown in Figs. 21–23 and Table 4 are obtained for the maximum of 500 deployed nodes by setting the simulation time 180 s with the node's speed is varied from 0 to 30 m/s.

The proposed clustering algorithm guarantees that the selected CHs are uniformly distributed over the network area and also avoids the multiple CHs to lie in the same cluster. The CSDGP uses a cluster switch based data gathering which blocks the simultaneous data transmissions of its neighbour clusters. Thus, only one CH is allowed to receive the data from the CM according to the allotted TDMA time schedule minimising unwanted signal from the neighbour clusters and collision. Moreover, the cluster switch mechanism collects the data from the migrated sensor nodes, whenever the CH has a free time slot and avoids isolated node. And also, an effective fault tolerance mechanism applied in tree topology easily finds the link failure and reestablishes the routing with minimal control overhead. This is reflected in the simulation results found in Figs. 21–23 that the proposed CSDGP has high PDR by 4.54, 5.11, 6.93 and 16.93%, low average end-to-end delay by 16.98, 17.55, 18.31 and 22.26% and reduced average energy consumption by 10.29, 9.92, 12.5 and 24% as compared to ARBIC, VELCT, MBC, and LEACH-M, respectively.

It is seen from Table 4 that the control overhead linearly increases with respect to the number of nodes. This is because of, the number of control packet used is directly proportional to the number of deployed nodes.

### *5.5 Impact of the simulation run time*

The simulation results are taken at 60, 120, 180, 240, 300, and 360 s for 300 deployed sensor nodes and the speed is varied from 0 to 10 m/s. Table 5 depicts that, whenever there is an increase in the runtime of the network, the workload (the number of packets processed by the networks) also increases, resulting in increased energy consumption and control overhead. In the proposed clustering mechanism, the residual energy of sensor node is an important parameter in CH election to avoid the low energy node to be elected as the final CH. In addition, the CSDGP takes care of the network stability by calculating the exactness of the connection

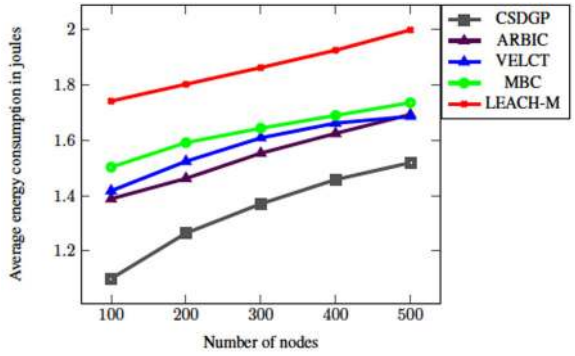


**Fig. 21** *Average energy consumption versus number of nodes*

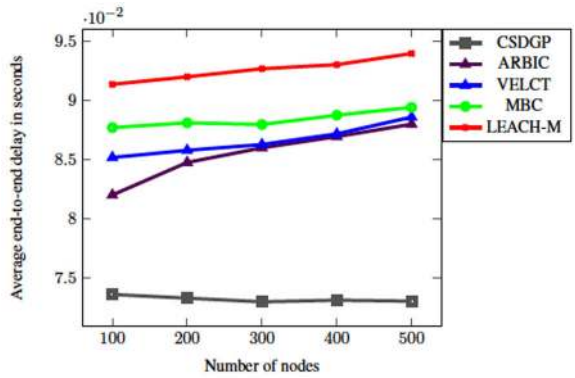


**Fig. 22** *Average end-to-end delay versus number of nodes*

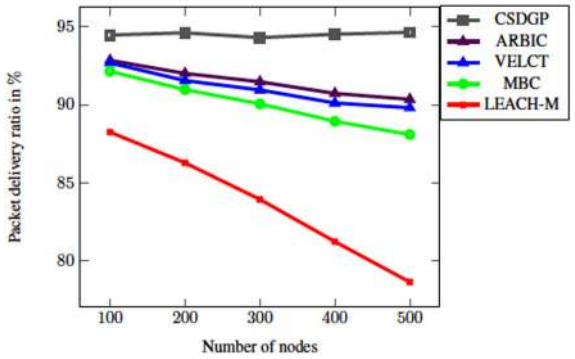


**Fig. 23** *Packet delivery ratio versus number of nodes*

**Table 4** Overhead analysis with respect to the number of nodes

| Protocols | 100 Nodes | 200 Nodes | 300 Nodes | 400 Nodes | 500 Nodes |
|---|---|---|---|---|---|
| proposed CSDGP | 5466 | 10,222 | 14,292 | 18,191 | 21,999 |
| ARBIC [9] | 5508 | 10,296 | 14,402 | 18,238 | 22,059 |
| VELCT [26] | 5759 | 10,587 | 14,700 | 18,684 | 22,408 |
| MBC [25] | 5541 | 10,386 | 14,533 | 18,271 | 22,122 |
| LEACH-M [18] | 5624 | 10,406 | 14,632 | 18,396 | 22,365 |

time between the nodes. Thus, it ensures the link availability and avoids the occurrence of link failures. And also, it establishes the stable routing path between the CH and the BS.

## 6 Conclusion

This paper presented a CSDGP which is completely operated in a distributed manner by eliminating the drawbacks in the random selection of CHs. In addition to this, the cluster switch mechanism utterly eludes the data collision at the receiver, thus minimising the data retransmission, improving the energy efficiency of the networks. Also, the CSDGP allows receiving the data from the

**Table 5** Network performances at different simulation run time

| Protocols | Performance metrics | 60 s | 120 s | 180 s | 240 s | 300 s | 360 s |
|---|---|---|---|---|---|---|---|
| proposed CSDGP | AEC in Joules | 0.38523 | 0.82620 | 1.24104 | 1.64314 | 2.06667 | 2.44018 |
| | PDR in % | 96.4486 | 96.1106 | 95.2733 | 92.1809 | 93.6743 | 93.7991 |
| | AEED in seconds | 0.06965 | 0.07018 | 0.07022 | 0.07041 | 0.07138 | 0.07113 |
| | control overhead | 4571 | 9057 | 13,537 | 17,935 | 22,095 | 26,596 |
| ARBIC [9] | AEC in Joules | 0.43148 | 0.96847 | 1.4461 | 1.96438 | 2.3484 | 2.87265 |
| | PDR in % | 92.608 | 91.964 | 92.476 | 91.842 | 91.2516 | 91.563 |
| | AEED in seconds | 0.081874 | 0.082134 | 0.081456 | 0.081664 | 0.082081 | 0.082264 |
| | control overhead | 4608 | 9182 | 13,599 | 17,992 | 22,176 | 26,702 |
| VELCT [26] | AEC in Joules | 0.446912 | 0.97242 | 1.45926 | 1.93564 | 2.37156 | 2.89813 |
| | PDR in % | 93.7289 | 92.7522 | 92.6234 | 93.1369 | 92.2162 | 91.768 |
| | AEED in seconds | 0.08247 | 0.08329 | 0.08167 | 0.08266 | 0.08411 | 0.08376 |
| | control overhead | 4861 | 9271 | 13,758 | 18,144 | 22,316 | 26,813 |
| MBC [25] | AEC in Joules | 0.45073 | 0.97976 | 1.48285 | 1.97181 | 2.44863 | 2.91328 |
| | PDR in % | 92.559 | 91.681 | 92.1457 | 88.711 | 90.0508 | 90.1345 |
| | AEED in seconds | 0.08298 | 0.08142 | 0.08370 | 0.08390 | 0.08472 | 0.08468 |
| | control overhead | 4634 | 9217 | 13,660 | 18,021 | 22,206 | 26,756 |
| LEACH-M [18] | AEC in Joules | 0.48572 | 0.99279 | 1.53362 | 2.03262 | 2.56241 | 3.06052 |
| | PDR in % | 86.2541 | 86.6487 | 86.0249 | 85.9543 | 85.2161 | 84.347 |
| | AEED in seconds | 0.09054 | 0.08924 | 0.08995 | 0.09002 | 0.09072 | 0.09122 |
| | control overhead | 4698 | 9232 | 13,702 | 18,112 | 22,290 | 26,786 |

MNs which minimise the data packet loss. Moreover, the selection of parent nodes and gateway nodes depend on the cost value which consequently lessens the workload of the sensor node. The simulation results of the CSDGP prove its efficacy over ARBIC, VELCT, MBC, and LEACH-M in terms of average energy consumption, average end-to-end delay, packet delivery ratio, and control overhead.

## 7 References


[1] Walravens, C., Dehaene, W.: 'Low-power digital signal processor architecture for wireless sensor nodes', *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, 2014, **22**, (2), pp. 313–321

[2] Bellasi, D.E., Rovatti, R., Benini, L., *et al.*: 'A low-power architecture for punctured compressed sensing and estimation in wireless sensor-nodes', *IEEE Trans. Circuits Syst.*, 2015, **62**, (5), pp. 1296–1305

[3] Gungor, V.C., Hancke, G.P.: 'Industrial wireless sensor networks: challenges, 'design principles, and technical approaches'', *IEEE Trans. Ind. Electron.*, 2009, **56**, (10), pp. 4258–4265

[4] Bhanumathi, V., Kalaivanan, K.: 'Application specific sensor-cloud: architectural model', in Mishra, B., Dehuri, S., Panigrahi, B., Nayak, A., Mishra, B., Das, H (Eds.): '*Computational intelligence in sensor networks. studies in computational intelligence*', vol. 776 (Springer, Berlin, Heidelberg, 2019), pp. 277–306

[5] Qi, Y., Cheng, P., Bai, J., *et al.*: 'Energy-efficient target tracking by mobile sensors with limited sensing range', *IEEE Trans. Ind. Electron.*, 2016, **63**, (11), pp. 6949–6961

[6] Meghanathan, N.: 'Link selection strategies based on network analysis to determine stable and energy-efficient data gathering trees for mobile sensor networks', *Ad Hoc Netw.*, 2017, **62**, pp. 50–75

[7] Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Trans. Wirel. Commun.*, 2002, **1**, (4), pp. 660–670

[8] Muruganathan, S.D., Ma, D.C.F., Bhasin, R.I., *et al.*: 'A centralized energy-efficient routing protocol for wireless sensor networks', *IEEE Commun. Mag.*, 2005, **43**, (3), pp. S8–S13

[9] Sabor, N., Ahmed, S.M., Abo-Zahhad, M., *et al.*: 'An adjustable range based immune hierarchy clustering protocol supporting mobility of wireless sensor networks', *Pervasive Mob. Comput.*, 2018, **43**, pp. 27–48

[10] Chen, G., Li, C., Ye, M., *et al.*: 'An unequal cluster-based routing protocol in wireless sensor networks', *Wirel. Netw.*, 2009, **15**, pp. 193–207

[11] Chiwewe, T.M., Hancke, G.P.: 'A distributed topology control technique for low interference and energy efficiency in wireless sensor networks', *IEEE Trans. Ind. Inf.*, 2012, **8**, (1), pp. 11–19

[12] Neamatollahi, P., Abrishami, S., Naghibzadeh, M., *et al.*: 'Hierarchical clustering-task scheduling policy in cluster-based wireless sensor networks', *IEEE Trans. Ind. Inf.*, 2018, **14**, (5), pp. 1876–1885

[13] Hanzalek, Z., Jurcik, P.: 'Energy efficient scheduling for cluster-tree wireless sensor networks with time-bounded data flows: application to IEEE 802.15.4/zigBee', *IEEE Trans. Ind. Inf.*, 2010, **6**, (3), pp. 438–450

[14] Yang, Y., Xu, Y., Li, X., *et al.*: 'A loss inference algorithm for wireless sensor networks to improve data reliability of digital ecosystems', *IEEE Trans. Ind. Electron.*, 2011, **58**, (6), pp. 2126–2137

[15] Han, G., Liu, L., Jiang, J., *et al.*: 'Analysis of energy-efficient connected target coverage algorithms for industrial wireless sensor networks', *IEEE Trans. Ind. Electron.*, 2017, **13**, (1), pp. 135–143

[16] Wilhelm, M., Lenders, V., Schmitt, J.B.: 'On the reception of concurrent transmissions in wireless sensor networks', *IEEE Trans. Wirel. Commun.*, 2014, **13**, (12), pp. 6756–6767

[17] Koubaa, A., Severino, R., Alves, M., *et al.*: 'Improving quality-of-service in wireless sensor networks by mitigating hidden-node collisions', *IEEE Trans. Ind. Inf.*, 2009, **5**, (3), pp. 299–313

[18] Kim, D.S., Chung, Y.J.: 'Self-organization routing protocol supporting mobile nodes for wireless sensor network'. Proc. of the First Int. Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06), Hanzhou, Zhejiang, China, 2006, 2, pp. 622–626

[19] Younis, O., Fahmy, S.: 'HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', *IEEE Trans. Mob. Comput.*, 2004, **3**, (4), pp. 366–379

[20] Malathi, L., Gnanamurthy, R.K., Chandrasekaran, K.: 'Energy efficient data collection through hybrid unequal clustering for wireless sensor networks', *Comput. Electr. Eng.*, 2015, **48**, pp. 358–370

[21] Taheri, H., Neamatollahi, P., Younis, O.M., *et al.*: 'An energy-aware distributed clustering protocol in wireless sensor networks using fuzzy logic', *Ad Hoc Netw.*, 2012, **10**, pp. 1469–1481

[22] Lee, S., Choe, H., Park, B., *et al.*: 'LUCA: an energy-efficient unequal clustering algorithm using location information for wireless sensor networks', *Wirel. Pers. Commun.*, 2011, **56**, (4), pp. 715–731

[23] Chao, S., Ru-Chuan, W., Hai-Ping, H., *et al.*: 'Energy efficient clustering algorithm for data aggregation in wireless sensor networks', *J. China Univ. Posts Telecommun.*, 2010, **17**, pp. 104–109

[24] Liu, T., Li, Q., Liang, P.: 'An energy-balancing clustering approach for gradient-based routing in wireless sensor networks', *Comput. Commun.*, 2012, **35**, pp. 2150–2161

[25] Deng, S., Li, J., Shen, L.: 'Mobility-based clustering protocol for wireless sensor networks with mobile nodes', *IET Wirel. Sensor Syst.*, 2011, **1**, (1), pp. 39–47

[26] Velmani, R., Kaarthick, B.: 'An efficient cluster-tree based data collection scheme for large mobile wireless sensor networks', *IEEE Sens. J.*, 2015, **15**, (4), pp. 2377–2390

[27] Kalaivanan, K., Bhanumathi, V.: 'Reliable location aware and cluster-tap root based data collection protocol for large scale wireless sensor networks', *J. Netw. Comput. Appl.*, 2018, **118**, pp. 83–101

[28] Stojmenovic, I., Lin, X.: 'Power-aware localized routing in wireless networks', *IEEE Trans. Parallel Distrib. Syst.*, 2001, **12**, (11), pp. 1122–1133