J Electr Eng Technol.2018; 13(2): 998-1007
http://doi.org/10.5370/JEET.2018.13.2.998

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

A Honey-Hive based Efficient Data Aggregation in Wireless Sensor Networks

Nandhakumar Ramachandran† and Varalakshmi Perumal*

Abstract – The advent of Wireless Sensor Networks (WSN) has led to their use in numerous applications. Sensors are autonomous in nature and are constrained by limited resources. Designing an autonomous topology with criteria for economic and energy conservation is considered a major goal in WSN. The proposed honey-hive clustering consumes minimum energy and resources with minimal transmission delay compared to the existing approaches. The honey-hive approach consists of two phases. The first phase is an Intra-Cluster Min-Max Discrepancy (ICMMD) analysis, which is based on the local honey-hive data gathering technique and the second phase is Inter-Cluster Frequency Matching (ICFM), which is based on the global optimal data aggregation. The proposed data aggregation mechanism increases the optimal connectivity range of the sensor node to a considerable degree for inter-cluster and intra-cluster coverage with an improved optimal energy conservation.

Keywords: Data aggregation, Honey hive clustering, Intra cluster min max discrepancy, Inter cluster frequency matching, Energy consumption

1. Introduction

Wireless sensor network is an emerging technology that has been used in all types of real-time applications since the last decade. Some of the major issues in wireless technology are the location of sensors [1] based on coverage of the nodes, resources and energy constraints. Normally sensor nodes are tiny lightweight devices that have less resources and have frequent energy depletion.

Due to the limited resources of sensor nodes, a data aggregation technique to aid in the transmission of sensed data with minimum energy consumption is needed, as discussed in [2]. To obtain an optimal energy conservation, nodes are scheduled at local intervals without any interference or data loss during transmission. To achieve this, a time slot is usually allocated for each sensor node within a locally reachable neighborhood that avoids the interference among local neighbors. This method helps to overcome the problem of data loss during data transmission. Data conflicts occur when two nodes have the same time slot for the transmission.

Honey hives are generally composed of hexagonal wax cells. A possible reason for honey hives being composed of hexagons is that hexagonal tiling maximize the total perimeter of the cells. Hexagonal wax cells are perfect for honeybees to store honey and providing a secure storage structure. Similarly, hexagonal tiling is the densest way to arrange sensor nodes that have a circular transmission range. Honey hives provide a data summarization and query analysis for fast and efficient data transmission from one cluster to another through Assembly Hubs and Cluster Heads (CH). The frequency of occurrence of data above the threshold level is identified at each Assembly Hub (AH).

In our work, raw data are collected from multiple sensors. Data are aggregated using the proposed data aggregation technique to improve the conservation of the sensor nodes energy. Based on their reachable range, collected information is aggregated using intra-cluster aggregation within a honey hive and inter-cluster aggregation among different honey hives. Simulation results demonstrate the feasibility of the proposed Intra-Cluster Min-Max Discrepancy analysis and Inter-Cluster Frequency Matching techniques for minimizing the energy consumption and extending the lifetime of sensor nodes compared with other existing data aggregation techniques.

The rest of the paper is organized as follows: Section 2 discusses research related to various energy conservation mechanisms in WSN; Section 3 presents the proposed honey-hive mechanism; honey-hive cluster formation, the Intra-Cluster Min-Max Discrepancy (ICMMD) technique for intra-cluster data aggregation, and the Inter-Cluster Frequency Matching (ICFM) technique for inter-cluster data aggregation are discussed in sections 4, 5 and 6, respectively; Section 7 presents the performance evaluation of the proposed mechanism. The conclusion follows in Section 8.

2. Literature Survey

[3] discusses a routing algorithm by introducing an

† Corresponding Author: Dept. of Computer Technology, MIT Campus, Anna University, Chennai, India. (nandhakumarr03@gmail.com)
* Dept. of Computer Technology, MIT Campus, Anna University, Channei, India. (varanip@gmail.com)
Received: June 7, 2016; Accepted: November 15, 2017

Energy Delay Index for Trade-off (EDIT) to enhance the network lifetime, optimize energy and reduce delay. The cluster heads and next hop are selected using EDIT by considering the energy and/or delay requirements of a given application. In this approach, the Euclidean distance and hop count are the two primary parameters used for finding the distance between a node and a Base Station. The proposed algorithm consists of two phases. The first phase is the cluster setup phase and the second is the steady state phase. Before executing these two phases, a neighbor discovery phase is executed once, at the time sensor nodes are deployed in the network area. In the neighbor discovery phase, the sink sends a hello packet to the nodes that lie in its transmission range. Each node in the cluster setup phase is allowed to wait for a period of time called the wait time. The wait time is the inverse of the node's remaining energy. Cluster heads are elected by comparing the received messages with energy details. Before final announcement, eligible cluster heads are allowed to wait 1/EDIT time. The second phase is the steady state phase. According to the TDMA schedule, data are transmitted to CHs by non-CH nodes.

[4] Concerns the grouping of sensor nodes in clusters of varying size. After the cluster formation, each cluster communicates with the fusion center is formed in an interleaved manner. The authors have proposed a technique that optimizes intra-cluster communication distance, and data fusion occurs within the network via the compression ratio. Sensor nodes in the network area are organized into k clusters, where k is a predefined value. If the data obtained by cluster heads from cluster members are fusible, the cluster heads perform data fusion and reduce the size of outgoing data.

[5] Proposes a mechanism called Energy Aware Sink Relocation (EASR). EASR is a technique used for mobile sinks in Wireless Sensor Networks. This technique makes use of the sensor node`s residual battery energy and then adjusts the sensor node`s transmission range by the sink relocation scheme. The EASR technique contains two components, namely, energy-aware transmission range adjustment and the sink relocation mechanism. For finding the first component, the sensor nodes are classified on the basis of remaining battery level and the transmission range is adjusted accordingly. The second component determines the sink relocation and determines the direction of the sink and the relocation distance.

[6] Presents the An Energy-Balanced Routing Method based on Forward-Aware Factor (FAF-EBRM) technique, in which routing is based on the forward-aware factor, which balances the energy and overcomes the communication range of the sensor node. In this paper, the forward energy density is defined by the forward transmission area, which combine slink weight with the forward aware factor. In this technique, the next-hop node is chosen based on the forward energy, density and link weight.

[7] Discusses Load-Balanced Data Aggregation Tree (LBDAT) construction in the Probabilistic Network Model (PNM). In this paper, three problems are identified and investigated. The first is the Load-Balanced Maximal Independent Set (LBMIS) problem, with LBMIS constructed by considering a graph and an independent subset. The second is the Connected Maximal Independent Set (CMIS) problem. In CMIS, when making the LBMIS connected, an attempt is made to find the LBMIS connector set that consists of a set of nodes of minimum size. Load Balanced Parent Node Assignment is conducted after the formation of CMIS. The direction to each link is given after the parent nodes are assigned in the constructed tree structure and the LBDAT is obtained.

[8] Discusses a clustering technique that uses hybrid compressive sensing for sensor nodes in wireless sensor networks. The sensor nodes forward data to the cluster head without using compressive sensing. Cluster heads send data to the sink using compressive sensing. The proposed work consists of three stages. The first one is the analytical stage, which finds the optimal cluster size and the relationship between the number of transmissions required for the cluster size. The second stage is the centralized clustering algorithm, in which sensor nodes are divided into clusters that have the optimal cluster size. The third one, in which the sink divides the field into C cluster-areas, calculates the geographic central point of each cluster-area and broadcasts the information to all sensor nodes to elect CHs.

[9] Proposes data aggregation scheduling using the signal-to-interference plus noise ratio(SINR) constraints of the sensor nodes in wireless sensor networks. This paper presents the construction of a routing tree and proposes two algorithms that generate collision-free link schedules based on a reduced graph. The reduced graph is obtained from the original communication graph. Compressive scheduling is used for maximizing the throughput where the maximum links are scheduled in each time slot.

[10] Presents a virtual hexagon-based clustering algorithm that extends the lifetime of sensor networks. The network is partitioned into virtual hexagons to avoid the overlapping of circular clusters. A new method is proposed for cluster head election to prevent frequent cluster head selection. Each sensor node forms a cluster head order list based on a weight value function, and the cluster heads are elected on the basis of the distance between sensor nodes and the center of the virtual hexagon.

[11] Proposes a technique for the contiguous link scheduling problem. In this paper, time slots are assigned to each node, and each node can wake up only once for data collection in its scheduling period. Interference-free link scheduling with the minimal time slot is the aim of this technique. This contiguous link scheduling can help reduce the energy consumed by the node`s state transmissions.

[12] Presents a load-balanced clustering algorithm. In this paper, the authors propose a technique for random distribution of wireless sensor nodes based on residual

energy. The proposed technique is implemented by considering two parameters, distance and density distribution of the neighbor node. The technique consists of three stages. The first one is the cluster head selection phase. The second one is the cluster building phase and the third is the cycle phase.

[13] Aggregates the data based on packet attributes. This process identifies samples of data from different sensors over different applications. The proposed Attribute-aware Data Aggregation (ADA) improves data aggregation efficiency with minimum energy consumption even as network size increases.

[14] Proposes an efficient algorithm that minimizes the delay among sensor nodes during data aggregation. Scheduling is based on the network radius and the maximum node degree in the communication graph. Dominator nodes at each level are scheduled based on their communication radii. Each data aggregate from the dominator nodes is finally transmitted to the sink.

[15] presents the Power Efficient Data Gathering and Aggregation Protocol (PEDAP), which is another form of centralized clustering scheme in which the sensor nodes are organized into a single minimum spanning tree using Prim`s algorithm. The tree also effectively reduces the total path length and the number of connections to lower energy nodes. One disadvantage of PEDAP is that the lifetimes of those nodes that are closest to the base station diminish rapidly compared to all other nodes.

[16] Proposes a method called Distributed Self-Organization for Wireless Sensor Networks (DSBCA) to balance the load of the clusters based on the distribution of nodes. A node's distance from the BS and number of neighbors are used for determining the cluster radius. The weight of the CH is a measure based on residual energy, number of neighbors and the duration for which the same node is elected as the CH. This algorithm has the negative characteristic of consuming maximum energy for communication and weight calculation during the cluster head selection process.

All the discussed literature is focusing on network life time maximization in WSN. The energy maximization is achieved through different considerations such as topology construction, link stability of the sensor nodes, relocation of the sink and data aggregation procedures. The energy conservation of WSN based on topology construction is discussed in [3, 4, 6, 8, 10, 12] and [16], as they are using clustering techniques. The energy consumption is reduced through load balancing among the neighbors. The member of the cluster is constructed based on their link stability, bandwidth and energy consumption for data transmission. Link stability is determined based on coverage of the neighbor nodes using RSSI which is discussed in [9]. In [5], sink is relocated periodically based on their energy of the reachable neighbors in order to conserve energy of the network. [13, 14] and [15], are using data aggregation based on their attributes, during different period for different type of data, and aggregation is done at each intermediate node from the source to the sink.

In most of these papers, CH is selected based on the residual energy, distribution strategy and distance from the base station. In our proposed work, the CH is selected not only based on the maximum number of neighbors, and energy. The proposed honey hive clustering algorithm helps to improve the coverage range and also improves the data aggregation process. The process such as Intra-Cluster Min-Max Discrepancy (ICMMD), an intra cluster data aggregation process and Inter-Cluster Frequency Matching (ICFM), an inter cluster data aggregation help to reduce the transmission messages during the data aggregation process. These improves the energy conservation and in turn, improves the network lifetime.

3. Proposed Work

The proposed honey-hive model collects the sensed data from the sensor nodes within a hive coverage region and transmits the data to the base station with minimal energy. This process includes: (1) the honey-hive cluster formation with the deployment of the sensor nodes, (2) Identifying the CH and assembly hub for data forwarding, (3) Quantifying the sensor information inside the honey hive using the Intra Cluster Min Max Discrepancy (ICMMD) technique, and (4) Aggregating the quantifying information to the AH using the Inter Cluster Frequency Matching (ICFM) technique for forwarding the collected information to the base station. The proposed architecture, shown in fig 1, consists of components such as base station, sensor nodes, cluster heads and assembly hubs, which are discussed in the following section.

3.1 Base station

The base station has sufficient energy. It is responsible for collecting the data from all the sensor nodes in the wireless sensor network. It initiates the data collection process by sending hello packets to its one hop neighbors.

3.2 Sensor Node(SN)

Each sensor node senses the environment condition

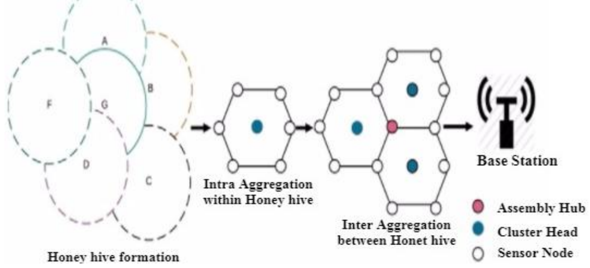

Fig. 1. Proposed Architecture

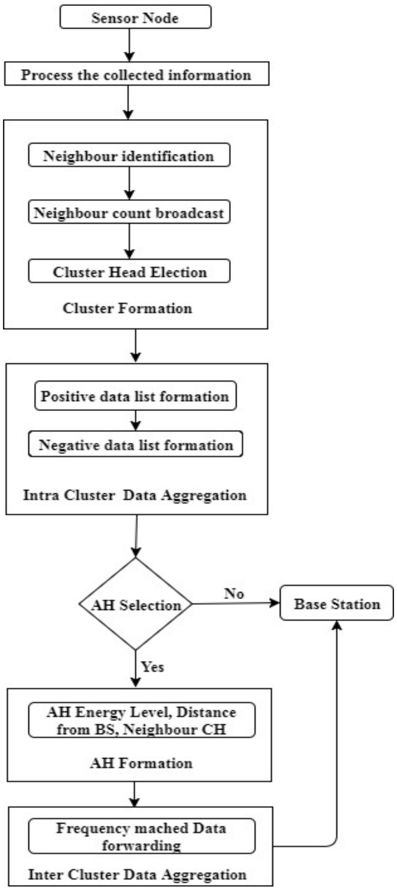

Fig. 2. Flow diagram of the proposed system

which contains its own id and the number of hops needed to reach the sink. In addition, each sensor node maintains a list of neighbors. It has its own transmission range and limited energy.

3.3 Cluster Head(CH)

Each honey-hive cluster has its own cluster head responsible for collecting the data from the member sensor nodes within the cluster. The node with the maximum number of neighbors and the most energy compared to its neighbors becomes the cluster head of that honey-hive cluster.

3.4 Assembly Hub (AH)

The assembly hub is responsible for collecting the data from the eligible cluster heads and performs the data aggregation and forwards the result to another assembly hub or directly to the sink. The flow diagram of the proposed system is shown in Fig. 2.

4. Honey-hive Cluster Formation

Initially, all the sensor nodes have the same energy level.

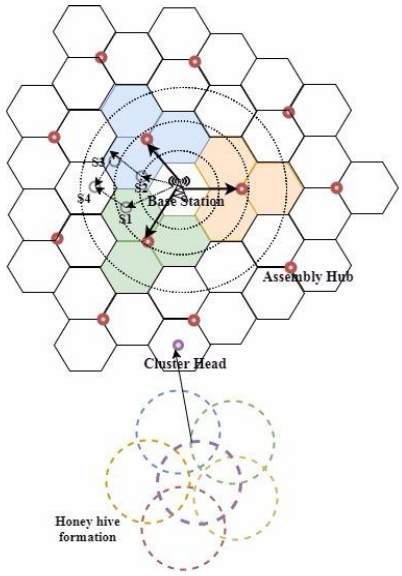

Fig. 3. Honey hive topology formation

To form the honey-hive structure among the sensor nodes, the BS broadcasts its hello message, with its id and initial BS hop count (Bh) set to 1, to its one hop neighbors. A node that receives the broadcast message from the base station sets its id and hop count, increments the BS hop count by one, and forwards the message to its one hop neighbors. Each downstream node receives the hop count information from BS or their reachable upstream nodes. When a sensor node receives Hop count from the Base Station (Bh) and also from more than one upstream neighbor node, it will set the minimal Bh and increment the Bh by 1 and forward the incremented Bh value with its id to the upstream neighbor nodes, as shown in Fig. 3. For example Bh hop count of S4 is set as 2 even though it receives Bh hop count from S3 as well as S1.

This process will be continued until all nodes receive the Bh distance. Once all the nodes receive Bh, a node i starts identifying its reachable-neighbors count (Si(nc)), as given in Eq. (1) , and the energy consumption (Si(E)) of node i for eligible honey-hive cluster formation is computed in Eq. (2).

Whenever the current (Si(E)) level falls below the application-specific threshold, then the algorithm searches iteratively for the next eligible CH to become the CH of the honey-hive cluster. The proposed honey hive cluster formation algorithm in given in Fig. 4.

$$S_i(n_c) = \sum_{j=1}^{n} (S_{ij}),\ i > 0, j > 0 \tag{1}$$

$$S_i(E) = \left\{ R_e + \left(\frac{1}{B_h}\right) * S(n_c) * C_e(S) \right\} \tag{2}$$

where,

$S_i(n_c) \rightarrow$ Neighbour count of ith sensor nodes

Input- Set of sensor nodes and sink node
Output- Set of Honey-hive clusters with CHs
Each node sends its neighbor count to its one hop neighbors. The node with the highest neighbor count compared to its one hop neighbors will be elected as CH.
Each CH will send a request message to its one hop neighbors.
for *each of the neighbors of CH* **do**
  if *a node receives the request message from more then one CH* **then**
    A neighbor node replies with join message only to CH with highest hop count from BH
  else
    A neighbour node will reply with a join message to CH
  end
 end
end
Cluster Head announces its election to its one_ hop_ neighbors
Honey-hive edges are formed with neighbors, with the exception of the Cluster Head.

Fig. 4. Honey hive cluster formation

S_{ij} → Reachable neighbour(j) of node i
R_{ei} → Remaining energy of node i
B_{hi} → Base station hop count of the node i
n → number of reachable nodes within the transmission coverage range of node.

For honey-hive cluster formation, each node sends its neighbor count to its one hop neighbors. As each node receives the counts from its one hop neighbors, it compares its neighbor count with that of each of its neighbors, as given in Eq. (1). If any of its neighbors has a count greater than its own count, the node recognizes its ineligibility to act as a cluster head. Then, it declares its ineligibility to its neighbor nodes. If a node neighbor count is more than the received neighbor counts of its neighbors, the node will consider itself to be able to act as CH and send a request message to all its neighbors. The neighbors in turn, reply with a join message to the CH. The CH will then announce its selection to its neighbors. If a neighbor node receives a request message from more than one CH, it will send the join message only to the CH with the maximum hop count from the sink (i.e., away from the sink). Therefore, an interior CH of a honey-hive cluster (closer to the sink) has a smaller number of neighbor nodes than an exterior CH and will use less energy for aggregation compared to a CH farther away from the sink. Moreover, the inner level CHs will forward more packets compared to CHs of outer levels. The honey-hive cluster formation is described in Fig. 3. During the construction phase, all the members of the honey hive set their threshold value as (half the value of the RRE of CH).

In the example shown in Fig. 5, node 1 is the cluster head. Every node shares its count with each of its neighbors. Node 1 is elected as cluster head as it has the maximum number of neighbors. Node 2 has three in its neighbor list, and node 3 has two in its neighbor list. In this way, finally, nodes 2, 3, 4, 5, 6 and 7 form the honey-hive cluster with node 1 as the cluster head.

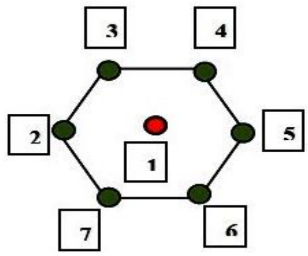

Fig. 5. Honey hive cluster formation

Input - Set of sensor nodes.
Output - Set of possible Assembly Hubs(AH)
if *the energy level of current Assembly Hub < threshold* **then**
  if *the next nearby eligible AH is available and more than two CH in its coverage range and the residual energy > threshold* **then**
    then that node become the new AH and set AH bit as 1
    Set the old Assembly Hub as passive node and set AH bit as 0.
    Return
  end
end

Fig. 6. Assembly Hub re-election algorithm

A node in the transmission range of more than two CHs is eligible to act as an Assembly Hub (AH), as given in Eq. (3). Each eligible AH announces its eligibility by changing an AH bit from 0 to 1. Among the received AH eligibility status messages, each downstream CH sends replies only to its optimal AH to forward the aggregated packets. For successive iterations, the AH will be reelected when the current AH becomes a passive node (i.e., it acts as a node that transfers data to the cluster head), which occurs when the energy of the current AH, given in Eq. (4), drops below the threshold value. The next eligible (AHi(E)) will be elected for data forwarding, based on downstream CH interest as given in Eq. (3). The newly elected AH then makes the old AH a passive node. The AH reelection algorithm is given in fig. 6.

$$S_i(AH) = \sum_{k=1}^{n} (S_{ik}), \quad i > 0 \tag{3}$$

where,

n is the number of CHs within the coverage range of sensor nodes

$S_i(AH)$ → number of CHs within the coverage range of node i
S_{ik} → k^{th} CH of i^{th} node.
$AH_i(E)$ → the energy consumption of i^{th} assembly hub

R_{aei} → the residual energy of i^{th} assembly hub
B_{ahi} → the hop count of i^{th} AH from sink

Each node in the honey hive holds 3 bits such as the Data Collection(DC) bit, Eligible for Assembly Hub (AE) bit, and currently acted as Assembly Hub (AH) bit.

After the selection of CH, if a node receives acknowledgement(ACK) from two or more CHs, then the node will set the AE bit as 1 to indicate that it becomes an eligible assemble hub. With this procedure, all AH nodes will set their AE bit to 1. No need to select all the AH as AH, even the AE bit is set. We should select the minimal number of AH to cover the entire topology to conserve energy and improve the network lifetime. For this, first we should select 3 AH, which are having the maximum coverage range from the BS and each is approximately 120^0 apart from each other to cover the 360^0 from the BS. From each of the AH, next level AHs will be chosen which are having the maximum coverage range from each of the AH in the current level and would have set AE bit as 1 already.

Then, these selected AHs will set their AH bit as 1. The same way all the current AHs will be chosen and the entire network topology is constructed.

5. Intra Cluster Data Aggregation

Initially, all the CHs will aggregate the sensed data from their member node using TDMA in order to avoid the collision problem. For example, as shown in Fig. 7, for CH 1, 4 member nodes are there such as 2, 3, 6 and 7. Then, the CH will maintain 4 DC bits for 4 member nodes and initially set as 0. For each time slot in TDMA, the CH will send the request message to each member.

Then, the member node will send the sensed data to the CH node. After receiving the sensed data from all member nodes in 4 time slots, the CH will compare its sensed data with each of the member data by keeping its value as reference, and deviation of +/-2 as threshold value. The CH will construct the Positive Data List (PDL) if the deviation of the member value is $\leq +2$ and > 0, and their frequency also stored in CH along with their member IDs. Same way, the CH will construct the Negative Data List (NDL), if the deviation of member value < 0 and ≥ -2.

In the next time instance, when the CH sent the request to each of the members, the members will compare the current sensed data with the previous collected data. If there is a change, then the DC bit is set to 0 and sends the reply message with new sensed data to CH, similarly to the earlier case. If there is no change between the current value and pervious value, the DC bit is set as 1 in member node and send this information to CH. CH also will set its corresponding member DC bit as 1.

The CH will send only 3 values such as CH sensed data as (reference data), mode of the PDL and mode of the NDL to its corresponding AH as Intra Cluster Min Max Discrepancy (ICMMD) list. This procedure is given as

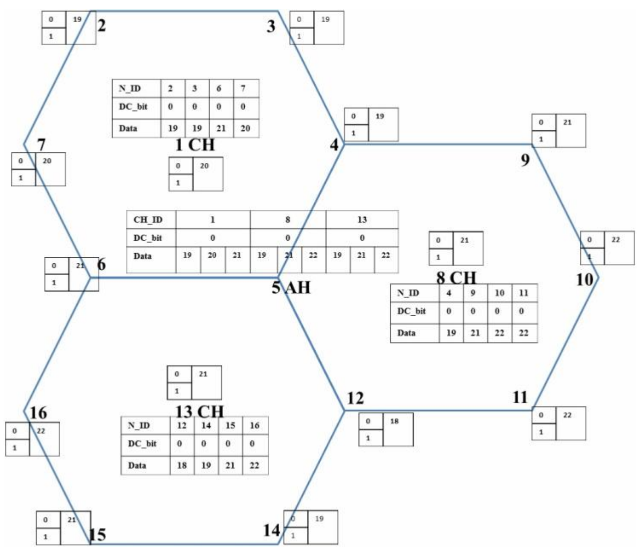

Fig. 7. Intra cluster data aggregation

```
INPUT : CH, Cluster member of honey-hive cluster
and threshold
OUTPUT : NDL, PDL, CH data as reference, ICMMD
List
Each CH maintains a binary variable Data
Collection(DC) bit for each of the member sensor
nodes
for each CH do
    if RRE > threshold
        // Remaining Residual Energy(RRE) then
        Data sensed by CH maintained as CH
        reference data
        for each received data from the member sensor do
            Variation is calculated as
            (CH reference data − received _ data)
            if variation > 0 and < +2 then
                Sensed data is indexed with id in PDL
                (Positive Data List)
                else if variation < 0 and abs(variation ≥
                -2) then
                    Sensed data is indexed with id in
                    NDL(Negative Data List)
                end
            end
        end
    end
    Mode for sensor data with their ids is retrived
    from both PDL and NDL
    Mode from PDL, CH data, mode from NDL is
    considered in ICMMD List and it is used for Inter
    Cluster data Troving
    if variation > -2 and < -2 then
        The abnormal sensed data is forwarded
        directly to next CH or Assembly Hub or BS.
        else if RRE ⇐ threshold then
            Announce its inability to its member nodes
            Select the next optimal CH and announce
            its ability as CH to its member nodes
        end
    end
end
```

Fig. 8. Intra Cluster Min Max Discrepancy (ICMMD)

Intra Cluster Min Max Discrepancy (ICMMD) algorithm in Fig. 8.

6. Inter cluster data aggregation

For inter-cluster data aggregation, the assembly hub runs the Inter-Cluster Frequency Matching (ICFM) algorithm using the ICMMD lists received from other cluster heads. The AH maintains a separate AH Eligible (AE) bit to represent a node`s eligibility to be an assembly hub for inter-cluster data aggregation and AH bit as 1 to represent that this node is the currently chosen AH. If the ICMMD lists from different CHs are the same, a single ICMMD list with the ids of the corresponding CHs and frequency are placed in ICFM List and forwarded to the next AH. ICFM List will be forwarded continuously until it reaches the sink. Thus, redundancy removal in forwarded messages helps to conserve energy. Otherwise, different ICMMD lists from different CHs are placed in the ICFM list by the AH and as such forwarded to the next AH continuously until it reaches the sink. The assembly hub is a node present in the transmission region of more than two cluster heads.

The ICMMD lists, along with their frequencies in ICFM, are forwarded to the sink using the Inter-Cluster Frequency Matching (ICFM) technique. The packet size and the transmission time are reduced by limiting the data (ICFM) forwarded to the next AH. This reduces the energy consumption of the AH, thereby increasing the lifetime of the network. The AH maintains a separate bit sequence for its data collection from each of the CHs within its coverage range. The aggregated data from each CH is transmitted to the AH within its assigned slot after comparing the collected packet information to determine the variation between the current and the previous data (ICMMD). If there is no change between the current and previous ICMMD list, no transmission from CH to AH is necessary, thus conserving energy. Otherwise, the CH transmits the ICMMD list to the AH in its allotted slot.

After constructing ICFM list from ICMMD lists from CHs within its AH coverage range, it compare the current ICFM list with earlier ICFM list. IF there is a change, then only, the new ICFM list forwarded to next AH and it will be continued till it reaches the sink. IF there is no change between the earlier ICFM and current ICFM, no need to send ICFM to next AH or to the sink. Thus, by avoiding the data transmission, to next AH or sink, energy conservation is improved.

In inter Cluster data aggregation, AH will aggregate the data (ICMMD) list, from each of its member CH using TDMA strategy to avoid the collision during data aggregation. For example, as shown in Fig. 9, node 5 will act as AH, and will aggregate the data (ICMMD) list from 3 CHs such as node 1, 8 and 13, in three different times slot using TDMA and will create ICFM list which contain 3 data such as the mode of positive data list from 3 CHs ICMMD list and the mode of negative data from 3 CHs ICMMD list and the mode of reference data from 3 CHs ICMMD list, that will be aggregated into a ICFM list of AH, and it will be forwarded to the next AH or to the BS.

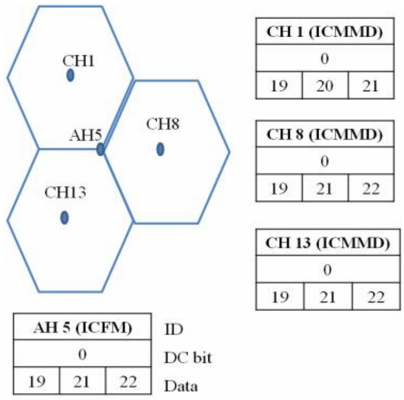

Fig. 9. ICFM list calculation

```
Input: ICMMD list of three Cluster Head, AE bit, RRE
Ouput: ICFM list
if (AE = 1 and RRE > threshold) then
    if (ICMMD list(CH1 data == CH2 data == CH3
    data)) then
        Generate ICFM with the same ICMMD list
        along with its own id
        else
            AH reference data = Mean(Ch1 data, CH2
            data, CH3 data)
            Find the PDL and NDL with maximum
            mode value with respect to AH reference
            data and ICMMD list
            Generate ICFM list at AH and forward to
            next AH or CH towards BS.
        end
    end
    else if  AE = 1 and RRE <= threshold then
        Change its AE bit to 0 and announce it to
        reachable CH
        Select next optimal AH and change its AE bit
        to 1 and announce to its reachable CH
    end
end
```

Fig. 10. Inter Cluster Frequency Matching (ICFM)

Here, AH (node 5) will maintain 3 DC bit for each of the member CH along with their ICMMD list and their IDs. If there is no change in all CHs data (ICMMD) list compared to the previous collected ICMMD lists, no need to send the ICFM list to the next AH or BS, and only send the reply message with its ICFM list to AH or BS.

If there is a change in any of the CHs ICMMD list, then updated ICFM list should be forwarded to the next AH or BS and the algorithm is shown in Fig. 10.

7. Performance Evaluation

The proposed system is simulated using NS-2.35. The sensing terrain is centered at (50 m, 50 m), the sensor nodes are distributed in a sensing field of 100m*100m, and the sink is located at (50 m, 90 m). Initially 200 nodes are deployed in the sensor network environment. Transmission range of each sensor node is 18m. Each node finds its hop count from the sink. Cluster heads are elected by comparing the neighbor counts of each node. Honey-hive clusters are formed and data are forwarded to the sink by performing inter-cluster and intra-cluster data aggregation.

In the proposed method, initial honey-hive cluster formation consumes substantial energy; with subsequent incremental iterations, the energy consumed is greatly reduced when compared to the existing approaches of in-network data fusion and LEACH approaches, as shown in Fig. 11. The result shows that the energy consumption of the proposed system is lesser than that of the other existing methods. At the initial stage, for constructing of honey hive topology, with CH and AH election will consume some energy. After successive iterations, the proposed technique consumes the minimum energy for computation and transmission compared to the existing methods.

Data aggregation duration time for the proposed technique ICMMD-ICFM is lesser compared to the existing approach such as In-network data fusion and LEACH approaches as shown in Fig 12. In-network data fusion allocates time for their predecessor and successor node. It increases the delay depend upon network size. Even though increase in size of network topology, the proposed honey-hive consumes the minimum time for data aggregation. The energy consumed for data aggregation is also lesser for the proposed system compared to in-network data fusion technique and LEACH.

Simulation results for the proposed intra cluster data aggregation technique with different variations are shown in Fig. 13. The data aggregation, with (Mean=2 and sigma = 2) scenario, the average number of data aggregation message is lesser compared to (Mean=2, Sigma=1) scenario as the network size increases.

Similiarily, with (Mean=2, Sigma=2) scenario the energy consumption also lesser due to the lesser number of messages transmitted compare to the (Mean=2, Sigma=1) which is shown in Fig. 14.

Due to the proposed honey hive cluster formation, minimum transmission range is efficient for data aggregation process from the member nodes to CH. So minimum energy consumption is possible for data aggregation process, in turn the network lifetime is also improved.

The radius of each honey hive is fixed throughout the network topology. Each sensed data item is transmitted to the sink based on the ICMMD list in a CH and the ICFM list from AH until all the Aggregated data reach the sink. The proposed intra-cluster (ICMMD) and inter-cluster (ICFM) data aggregation techniques minimize the aggregation ratio compared to in-network data fusion, as

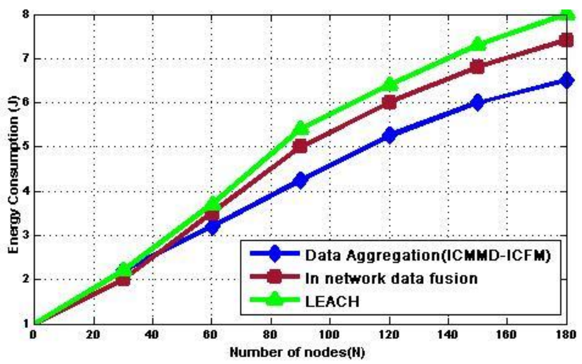

Fig. 11. Energy consumption according to the number of iterations

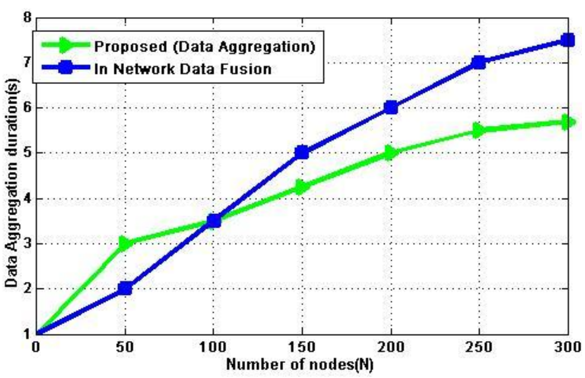

Fig. 12. Data aggregation duration

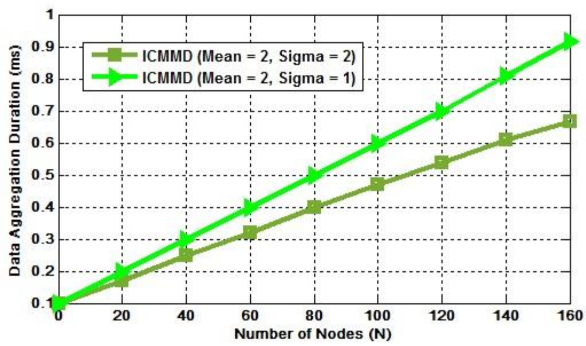

Fig. 13. Data aggregation duration with difference variance of ICMMD

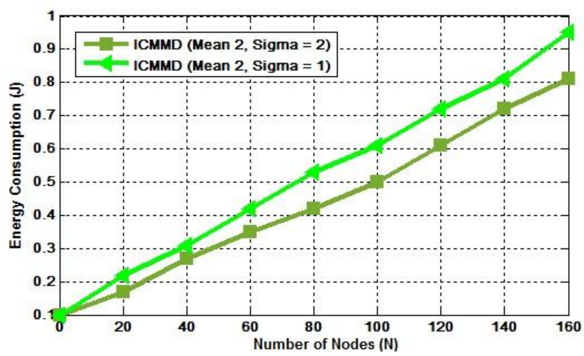

Fig. 14. Energy consumption with difference Mean=2 (σ = 1, σ = 2) of ICMMD

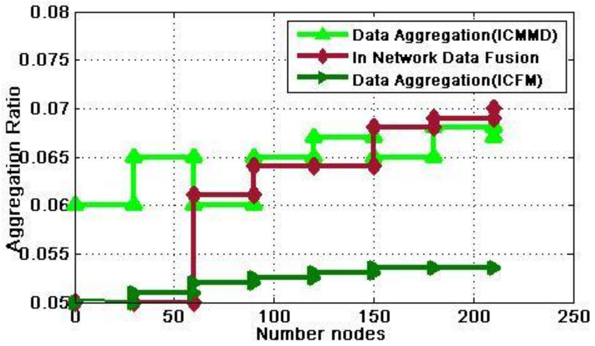

Fig. 15. Aggregation Ratio

shown in Fig. 15. The data aggregation ratio of the proposed work itself considered through as two cases. In case 1, data aggregation is done only with ICMMD list alone through CH to BS. In case 2, data aggregation is done with ICMMD list and the ICFM list from CH to AH and to the BS. Anyway, the performance of two cases of the proposed method is better than the In-network data fusion existing method. Intra-cluster data aggregation depends upon the member sensor nodes. The aggregation rate increases and decreases, but inter-cluster data aggregation in AH will maintain the same aggregation rate, which is determined by the number of CHs in each honey-hive communication to the AH. ICFM list maintains in each AH based on the ICMMD list of their member CHs. Even though increase in size of the network depends upon the need of the user, the aggregation rate remains same in AH Each AH aggregate the frequency item among CH, so the amount of data forwarded from AH to BS will be minimum compared to CH alone as in In-network data fusion method

8. Conclusion

The proposed honey-hive topology based clustering mechanism minimizes the data transmission time by optimizing the intra- and inter-cluster data aggregation techniques. The data difference maintained in the ICMMD list and ICFM list optimizes the energy consumption and number of transmissions using the proposed Intra-Cluster Min-Max Discrepancy (ICMMD) and Inter-Cluster Frequency Matching (ICFM) techniques. Thus, the proposed honey-hive cluster procedure and data aggregation techniques increases the lifetime of the network compared to existing data aggregation techniques.

References

[1] Xingfa Shen, Zhi Wang, Peng Jiang, Ruizhong Lin and Youxian Sun, "Connectivity and RSSI Besed Localization Scheme for Wireless Sensor Networks," *Advance in Intelligent Computing Lecture Notes in Computer Science*, vol. 3645, pp. 578-587, 2005.

[2] Haeyong Kim, YonghoSeok, Nakjung Choi, Yanghee Choi and Taekyoung Kwon, "Optimal Multi-sink Positioning and Energy-Efficient Routing in Wireless Sensor Networks," *Information Networking, Con-vergence in Broadband and Mobile Networking Lecture Notres in Computer Science*, vol. 3391, pp 264-274, 2005.

[3] Ankit Thakkar and KetenKotecha, "Cluster Head Election for Energy and Delay Constraint Appli-cations of Wireless Sensor Network," *IEEE sensors journal*, vol. 14, no. 8, pp. 2658-2664, August 2014.

[4] Chi-Tsun Cheng, Henry Leung, and Patrick Maupin, "A Delay-Aware Network Structure for Wireless Sensor Network With In-Network Data Fusion," *IEEE sensors journal*, vol. 13, no. 5, pp. 1622-1631, May 2013.

[5] Chu-Fu Wang, Jau-Der Shih, Bo-Han Pan, and Tin-Yu Wu, "A Network Lifetime Enhancement Method for Sink Relocation and Its Analysis in Wireless Sensor Networks," *IEEE sensors journal*, vol. 14, no. 6, pp.1932-1943, June 2014.

[6] Degan Zhang, Gaung Li, KeZheng, Xuechao, and Zhao-Hua Pan, "An Energy-Balanced Routing Method Based on Forward-Awares Factor for Wireless Sensor Networks," *IEEE Transactions On Industrial Infor-matics*, vol. 10, no. 1, pp. 766-773, Feburary 2014.

[7] Jing He, ShoulingJi, Yi Pan, and Yingshu Li, "Constructing Load-Balanced Data Aggregation Trees in Probabilistic Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 25, no. 7, pp. 1681-1690, July 2014.

[8] RuitaoXie and XiaohuaJia, "Transmission-Efficient Clustering Method for Wireless Sensor Networks Using Compressive Sensing," *IEEE Transactions on Parallel and Distributed systems*, vol. 25, no. 3, pp. 806-815, March 2014.

[9] XingfaShen, Zhi Wang, Peng Jiang, "Ruizhong Lin and Youxian Sun, Connectivity and RSSI Besed Localization Scheme for Wireless Sensor Networks," *Advance in Intelligent Computing Lecture Notes in Computer Science*, vol. 3645, pp. 578-587, 2005.

[10] Xin Guan, Lin Guan and Xingang Wand, "A Novel Energy Efficient Clustering Technique based on Virtual Hexagon for Wireless Sensor Networks," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 4, pp. 1891-1904, 2011.

[11] Junchao Ma, Wei Lou, and Xiang-Yang Li, "Contiguous Link Scheduling for Data Aggreagtion in Wireless Sensor Networks," *IEEE Transactions on Parellel and Distributed systems*, vol. 25, no. 7, pp. 1691-1701, July 2014.

[12] Han Zhang, Liang Li, Xin-fang Yan, and Xiang Li,

"A Load Balanced Clustering Algorithm With for Data Gathering," *2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce* (*AIMSEC*), pp. 915-918, 2011.
[13] Fengyuan Ren, Jiao Zhang, Yongwei Wu, Tao He, Canfeng Chen and Chuang Lin, "Attribute - Aware Data Aggregation using Potential-Based Dynamic Routing in Wireless sensor Networks," *IEEE Transactions on parallel and Distributed System*, vol. 24, no. 5, pp. 881-891, May 2013.
[14] Xiaohua Xu, Xiang-Yang Li and Min Song, "Efficient Aggregation Scheduling in Multihop Wireless Sensor Networks with SINR Constraints," *IEEE transactions on mobile computing*, vol. 12, no. 12, pp. 2518-2528, December 2013.
[15] HuseyinOzgur Tan and Ibrahim Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," *ACM SIGMOD*, vol. 32, no. 4, December 2003.
[16] Ying Liao, Huan Qi, and Weiqun Li, "Load Balanced Clustering Algorithm With Distributed Self- Organization for Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1498-1506, May 2013.

Nandhakumar Ramachandran is currently pursuing his Ph.D. in Anna University Chennai. He received B.Tech. degree in Information Technology and M.E. degree in computer science and Engineering from Anna University. His research interest includes Wireless Sensor Networks, Mobile Ad hoc Networks, IoT and Machine Learning.

Varlakshmi Perumal is an Associate Professor in Madras Institute of Technology, Anna University. She received her Ph.D. degree under Information & Communication Engineering-Trust Management in Grid Computing at Anna University. Her research interest includes Grid & Cloud Computing, Compiler Design, Network security and IoT.