ENHANCEMENT OF VARIOUS COMPUTATIONAL CLUSTERING TECHNIQUES IN WIRELESS SENSOR NETWORKS

R.Tamilselvan
Assistant Professor
Department of Computer Science and Engineering,
Chaitanya Bharathi Institute of Technology,
Email: tamilselvancse@gmail.com

K.Sankar
Associate Professor
Department of Computer Science and Engineering,
GITAM University, Bangalore Campus,
Email: sankarkaliappan@gmail.com

Abstract: In recent years, the Wireless Sensor Network (WSN) clustering technique has proven to be the most effective at reducing packet loss and communication costs. In particular, the extension of the nodes' lifetime in WSN is significantly improved by clustering using a distributed technique. The Energy efficient Clustering Hierarchy, which is based on evolutionary algorithms, forms clusters centrally (GAECH). Energy Efficient Coverage aware Data Collection are used to centrally build clusters (EECDC) eliminates this redundant data by forming clusters whose members' coverage regions do not cross. The most significant design issues for WSNs are with Quality of Service (QoS) and energy efficiency because of the frequent topological changes and absence of a charging option or battery replacement. The two primary steps of the unsupervised learning technique K-Means are clustering observations and computing cluster centres up till there is no change in the estimated centroids. Weight-based clustering is examined by the Mobility Based Clustering Algorithm (MBA). The MK-means algorithm, which is used in this research, incorporates energy, location, load balancing, and distance to CH with various operational modes and top query data aggregation models. MK-means picks several CHs in one.

***Keyword**: Cluster, WSN, GAECH, EECDC, K-Means.*

1. INTRODUCTION

The wireless sensor network (WSN) is the connection of numerous sensor nodes that are detecting the necessary environmental physical occurrences. Even though WSN technology is constantly being improved, a key element influencing how well it works in many practical applications is still the network's lifetime. The WSN data collection mechanism that works best is clustering.

A massive number of sensor nodes that run on batteries make up a wireless sensor network (WSN). To transfer sensed data from the end sensor nodes to the base station, nodes in the WSN cooperate with one another (BS). WSN was initially developed for heavy industry, environmental monitoring, and military applications (tracking the faraway target). The development of electronics, material science, and communication protocols allowed for the wide-scale deployment of WSN in non-critical applications. The fundamental criteria for WSN-based applications include load sharing, energy efficiency, increased operational life, and network scalability.

The working group of the WSN is made up of just a few thousand sensor nodes that are connected to one another. Due to the expense of connection maintenance and the complexity of the technology, the number of participating sensor nodes is restricted. Wireless LAN technologies like Wi-Fi (IEEE 802.11) and Bluetooth are used by existing WSN (IEEE 802.15.1). The ZigBee (IEEE 802.15.4) and Ultra-Wide Band (IEEE 802.15.4a) communication technologies show promise for lowering the cost of acquiring more nodes. Once low-cost nodes are accessible, the complexity of extending the network and the portability of software to handle a wide range of applications become the bottlenecks instead of node pricing. The applications like RFID (Radio Recurrence Distinguishing proof) labels and cell phones which support gigantic measured dispersed remote hubs couldn't measure up to WSN because of the asset limitations engaged with the sensor hubs plan.

A WSN comprises of a gathering of remotely interconnected, spatially circulated PCs utilizing sensors and actuators to survey and interface with their general climate. These PCs, known as sensor hubs, are typically asset obliged, little structure factor gadgets intended to run on battery power. Normally, they speak with a far off door through an uplink association with an organization spine or by means of their nearest neighbor in a remote lattice network figure 1.1

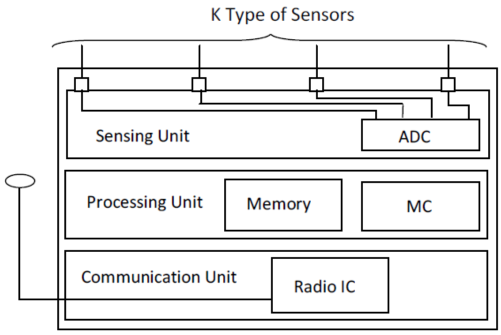

Figure 1.1: Block diagram of a sensor node

Clustering is the technique of assembling sensor nodes that are geographically close to one another [1]. A cluster head (CH) is a node that manages a cluster and may start the clustering process. Cluster members are the remaining nodes in the cluster (CM). These CM nodes will continuously perceive their surroundings and transmit data to the corresponding CH nodes. Since all member nodes in a cluster are close to one another, the information they produce will also be redundant. In the majority of application instances, sending this redundant information to the BS is unnecessary, and it also shortens the network's lifespan. Therefore, the CH nodes combine the data they have obtained from the CMs into a single piece. This single piece of information alone will be communicated to the BS.

The grouping plan tends to the self-association and adaptability properties with rotational based CH determination. Subsequently a gathering is driven by CH hub with energy preservation worldview to send the information to Base Station (BS) in single bounce or multihop. Bunching further develops the energy advancement by spatial reuse of resources[2]. It decreases the steering above and directing postponement. It expands the hub the executives successfully and can have a planned climate between the hubs. Consequently bunching follows its own execution way in all WSN applications.

The bunching methodology suggested in a few computations takes into account various boundaries mentioned above and various grouping methods used in WSNs [3]. According to the bunch head capacity, hub type, procedure used (model), and grouping methods, the available calculations can be categorised. As shown in Figure 1.2, this system starts with handling and decision-making at hubs, whether they are unified or distributed in style.

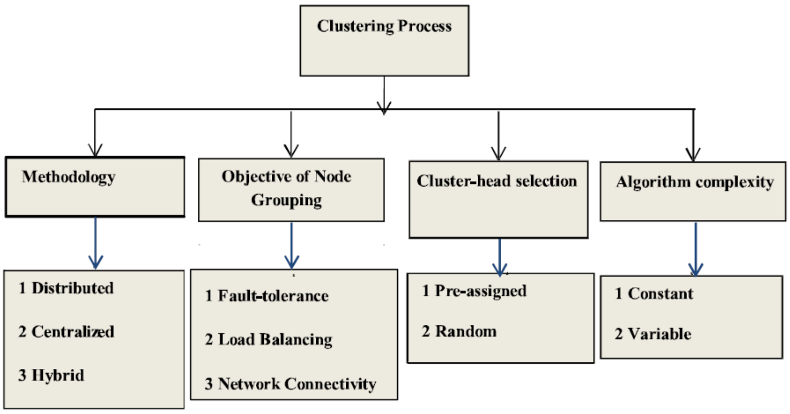

Figure 1.2 Taxonomy of Clustering Process

Higher abilities fall under heterogeneous organisation, or on the other hand, on the off chance that the hubs are identical, they are classed as homogeneous organisation. Hub kinds describe the interoperability of hubs in an upward direction over the various available correspondence advancements. In heterogeneous sensor networks, simple sensor nodes are used for sensing activity while sensors with more processing power and transmission range are utilised for data gathering, processing, and forwarding to another node.

2. LITERATURE REVIEW

There are two sorts of nodes in a clustering architecture: I Cluster member (CM) nodes that sense external events in their environment and (ii) Cluster head (CH) nodes that gather data from CMs and transmit it to the central base station (BS). The creation of stable balanced groups and extending the network operating lifetime in WSN have attracted researchers' interest lately. Numerous earlier studies in this field concentrated on improving energy efficiency, latency, throughput, and balanced load distribution across nodes to lengthen network lifetime.

Hybrid Energy-Efficient Distributed

clustering (HEED), a reliable distributed processing clustering method. With a clear focus on energy, HEED offers a multihop communication channel that carries an energy-efficient clustering routing. The fact that HEED does not choose the cluster head at random sets it apart from LEACH. The remaining energy of each node and the cost of intra-cluster communication are two metrics that are combined during the clustering process.

Classification of clustering algorithm

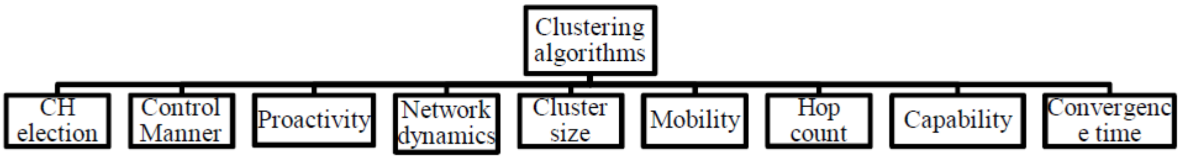

Fig 2.1 clusteting algorithm Classification

The various categorization strategies for WSN clustering algorithms are shown in Figure 2.1. The suggested classification scheme divides clustering algorithms into different categories based on size, CH election, and control style. The clustering algorithms are initially split roughly into two groups: I distributed algorithms and (ii) centralised methods. Once more, each kind is split into I clusters of identical size and (ii) clusters of unequal size. Nearly all equal-sized clustering methods need the CH nodes to send the aggregated data straight to the BS. The energy required for this direct data transfer will be higher from the CH nodes far from the BS. The architectural solution to the remote CH node problem is unequal sized clustering. Both SECA-M and ECRA take on unified bunching way to deal with make energy effective group tree directing construction for the sensor hubs.

EAUCF (energy aware unequal clustering using a fuzzy technique) [4] makes a dispersed determination of the CH hubs. The fluffy info factors lingering energy and distance to BS are utilized for CH political race. Arbitrary number age is utilized to pick the potential CHs hubs. A fluffy derivation approach is utilized to decide the opposition span for the potential CH hubs (FIS). Energy-efficient data collection(EEDC): The spatial connection between's the sensor hubs is the establishment for EEDC [5]. Since the part hubs of a group are neighboring each other, the information created by these hubs will be repetitive. To forestall repetitive information creation and to save the hubs' leftover energy, EEDC just allows one hub to be dynamic in a group.

EECPL: The Energy Efficient Clustering Protocol (EECPL) comprises of two stages: arrangement and consistent state [6]. During arrangement, the BS is educated regarding the excess energy and the area. The BS currently chooses hubs for a bunch from two gatherings. There are two distinct geographies: bunch source and CH. During the TDMA consistent state period, the CH will plan and broadcast to the individuals likewise to different calculations. As indicated by the timetable, the part hub sends the information to the following ring hub.

Conventional Weight based CH Election:

Typically, CH nodes in this type are chosen using a weight value. A single parameter or a collection of related parameters may be used to determine the weight. A pioneering centralised clustering technique for WSN is LEACH-C: Low Energy Versatile Bunching Progressive system - Unified (Drain C) [7]. There are two stages in it: the arrangement stage and the consistent state stage. During the arrangement stage, every hub sends the BS data about its situation and remaining energy. Drain C guarantees all the more equally dispersed groups with minimal measure of room between the CHs. Drain F: The decent bunching procedure from the Filter family is called Filter Fixed [41]. Like Drain C, the working is static once the bunches are delivered and keeps on being so all through their lives. The CH obligation is turned among the individuals inside a similar bunch during each round.

CGC: Centralized genetic based clustering (CGC) is a recently developed genetic algorithm-based clustering algorithm for WSN. Along with the idea of a genetic algorithm, CGC employs a novel strategy called onion layering.

GCA: The genetic clustering algorithm (GCA) increases longevity by controlling two crucial variables. The total transmission distance (also known as intra-cluster distance) within a cluster is the first parameter. By summing each CM's distance from the CH node, the overall transmission distance is determined. The total number of CHs in the network is the second parameter. Since CH nodes consume more energy than CM nodes do, fewer CH nodes will result in a significantly longer network lifetime. The fitness function of GCA is displayed in equation (2.1).

A centralised dynamic clustering algorithm for WSN is called CDC (centralised dynamic clustering). The clusters are created by BS, just like in other centralised clustering. However, for each cycle, the same cluster is retained rather than re-clustering.

The two phases of the adaptive decentralised re-clustering protocol (ADRP) are the initial phase and the cycle phase. Similar to LEACH-C, the initial phase of ADRP includes communication to the BS about the node's position and remaining energy.

The startup phase and the steady state phase are both parts of the ERP-SCDS protocol. It is a dynamic routing technique with a static cluster that uses less energy. Again, there are three sub-phases to the initialization process: The Energy Efficient Clustering Algorithm Based on Neighbors, or EECABN, is the name of the centralised weight-based method. The BS initially divides the nodes into strong and weak categories.

Another centralised evolutionary computing approach for WSN is called the Evolutionary Aware Energy Efficient Routing Protocol (EAERP). ERP: Another genetic-based clustering approach for WSNs is the evolutionary routing protocol (ERP). For its fitness function, ERP takes three factors into account. The first parameter is intra-distance or compactness, which is the total of all clusters' clusters' lowest distances between any CM and its CH. In-depth explanations of the various clustering methods' goals are provided. Depending on the requirements of the final application, the objectives will vary.

3. SENSOR NETWORK

Clustering in WSN comprises assembling sensor nodes for easy management of nodes, resources, network stability, collaborative sensing, and energy-efficient multihop data forwarding. The graph G = (V, E) can be used to visualise the wireless sensor network (WSN) that this dissertation assumes [91]. V is the group of sensor nodes in the network, and E denotes the links that connect the nodes. If two nodes x and y can be connected together directly, an edge Exy will be added to the set E. A sensor node's communication radius Rc is constant and the same for every other node in the network. Any two nodes can directly communicate with each other if their distances are smaller than Rc.

The region of interest (ROI) is the area where the sensor nodes are installed for monitoring. The communication path between two neighbouring nodes can be visualised as an undirected edge in the set E. Both inside and outside of the ROI are possible locations for the Base Station (BS). Figures 3.1, 3.2, and 3.3 show the sample network that has been set up in the 200 m by 200 m square. The blue spots stand in for the sensor nodes, whereas the red point denotes the BS. BS is situated near the corner of Figure 3.2's ROI, in the middle of Figure 3.1's ROI, and outside of Figure 3.3's ROI.

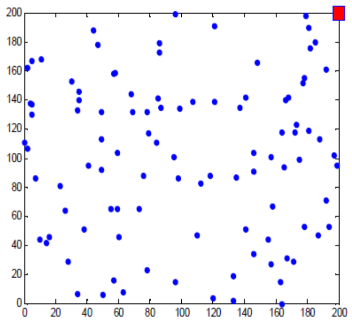

Figure 3.2: BS at the corner of ROI

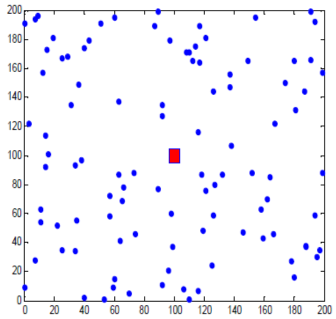

Figure 3.1: BS in the middle of ROI

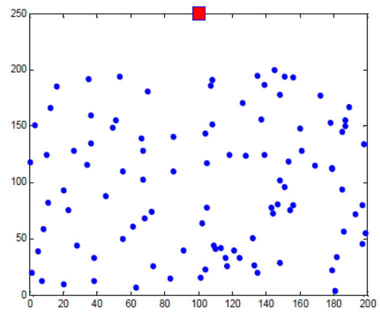

Figure 3.3: BS outside to the ROI

While GAECH, EEGTP, and EECDC fall under a centralised method, DUCF and FLECH are distributed clustering algorithms. However, the execution of the algorithm is split into two separate stages: I the Cluster construction phase, and (ii) the Data collecting phase. A round symbolises the phases of cluster creation and data gathering connected to it. A round in the WSN clustering techniques is shown in Figure 3.4.

Cluster Formation phase:

In the case of centralised methods, the BS will be informed of the necessary information from the nodes at the beginning of the cluster formation phase using control messages, such as residual energy, the number of neighbours, etc. Based on the data gathered, the BS will create clusters, elect a Cluster Head (CH), and announce the results to every node. However, when using distributed techniques, neighbouring nodes exchange control messages rather than sending information to the BS. Based on the information exchanged, the neighbouring nodes elect their CH, creating distributed clusters. None of the above distributed or centralised techniques create clusters that overlap. In the network, there is only one CH to which all Cluster Members (CM) belong.

K-Means Clustering In Wireless Sensor Networks

By dividing the network into one level of partitions, the K-Means clustering algorithm is one of the most used clustering algorithms. K cluster centres are initially selected at random, and each node is given a point for the centre that is closest to it. The mean of various member patterns is then used to update the clusters, and the process is repeated until the algorithm converges. The most iterations in a row, a difference in the value of the distortion function, etc. are examples of common convergence criteria.

3.1 Implementing Centralized & Distributed Clustering

Sending the positions and energies of each node to the central node: The decision-making location should have access to the positions and energies of each node. The sink node, or central node, in centralised clustering serves as the decision-making entity. The central node can therefore access the positions and energy of every node. Accessing resources and positions: Each node's position can be obtained from the object files that were created for it. In a specified work plane, nodes are deployed using a random deployment approach. The processor node receives position and energy by contacting the pointers used for location storing and energy handling.

In distributed clustering, each organization hub partakes in the dynamic cycle. Consequently, every hub ought to approach the position and energy of each and every other hub. The partaking hubs make a construction as a connected rundown in "node.h" and proclaim its initialised pointer in class Hub to store values for hub id, position, and energy upsides, everything being equal. Thus, a trigger hub communicates a question in the organization requesting data about its hub id, position, and energy to get this information. To keep away from bigger bunch gatherings, got control messages are hashed on every individual hub prior to being sent to different hubs with a limited number of bounces. Hubs after t_initial will call hashed information to choose CH among themselves and announces themselves after a few emphasess.

3.2 Fuzzy Logic based Energy efficient Clustering Hierarchy (FLECH) for WSN:

Three information boundaries - leftover energy, hub centrality, and distance to base station - were utilized by FLECH to choose CHs (BS). The result variable is possibility. At long last, CHs are chosen utilizing this opportunity factor. The Low Energy Versatile

Grouping Order (Drain), Bunch Head Political decision Instrument Utilizing Fluffy Rationale (Culinary specialist), Energy Mindful Conveyed Grouping Utilizing Fluffy Rationale Approach, and ECPF are stood out from the FLECH (EADC-FL). While Filter is a notable standard bunching procedure that might be differentiated, ECPF and Cook both utilize fluffy rationale based grouping. It serves as a benchmark because EADC-FL was developed specifically for non-uniform networks like FLECH. The three researchers compared FLECH to various clustering algorithms in order to confirm its effectiveness.

Distributed Cluster Head Election Mechanism based on Modified K-Means in a Wireless Sensor Network

K cluster centres are initially selected at random, and each node is given a point for the centre that is closest to it. Once the mean of the various member patterns is determined, the clusters are changed, and the process is repeated until the algorithm converges. The most iterations in a row, a difference in the value of the distortion function, etc. are examples of common convergence criteria.

3.3 Modified K-Means Algorithm

Based on the original K-Means algorithm, the Modified K-means method (MK-means) has gone through significant change to incorporate burden adjusting and network lifetime augmentation. The strategy adds a last step and continues in a progression of legitimate advances like K-implies subsequent to arriving at stable bunches.The proposed approach beats k-implies by consolidating information total, top-k questions, and various bunch head choice in a solitary group. The CHs in a group successfully share the responsibility by diminishing how much control messages expected for bunching during political race in the request for the quantity of CHs chose during political decision.

In this way, as per the recommended MK-implies calculation, the energy expected to choose a solitary CH for a group is identical to the energy expected to choose various CHs without a moment's delay. The bunch head political race stage and group development energy consumption during re-decisions will be decreased because of this bunch size and CHs working for a more drawn out timeframe, up until the last CH from CHs picked in the bunch actually passes the information to sink. For example, in the event that three group heads (CHs) are picked in a solitary political decision, the CHs will be given TDM openings to fill, which will consolidate the three CH races into a solitary CH political decision. The group configuration utilizes around a third less energy, and the information total and top-K question have a huge impact in dealing with the information move through conglomeration and limit based information sending. In contrast with ECRA, SECA-M, K-means, Notice, and Filter, the proposed strategy MK-implies with top-K question has higher remaining energy, greater steadfastness, and lower energy use, as per the after effects of the simulations.

3.4 Graph Based Clustering Hierarchy In WSN

A wireless sensor network (WSN) can be represented by the graph G = (V, E), where V is the set of sensor nodes and E is the set of links linking neighbouring nodes. Consider two nodes u and v from the set V. If u and v are both within communication range of one another, an edge Euv can be added to the set E. The cost of Euv is equal to Evu since the symmetric link is taken into account in this dissertation. The value of Euv might be either the distance between u and v or the energy needed to send a data packet from u to v.

Maximal Independent Set Based Clustering Hierarchy In Wsn

A set with non-adjacent elements is referred to as an independent set in the context of graph theory. Non-adjacent elements are those in the graph that do not have edges connecting them. The term "maximum independent set" refers to an independent set that contains the greatest number of such non-adjacent items (MIS). The non-adjacent elements in a network view are those that are not directly related to one another (i.e., there is no direct communication link).

Redundant Data in a Cluster:

The Cluster Members (CMs) are situated near to one another in the traditional clustering architecture. These individuals will produce data that is more redundant in nature. At the Cluster Heads (CHs), this redundant data are aggregated using the proper methods to create a single data item. One of the key determinants of energy usage in CHs is aggregate cost. Avoiding the creation of redundant data is the best strategy to address this issue rather than aggregating it.

4. CONCLUSION

A potential technique for monitoring the environment when human interaction is either not possible or not necessary is Wireless Sensor Networks (WSN). In this paper, the time complexity, node mobility, cluster overlap, and topology followed of the algorithms DUCF, FLECH, GAECH, EEGTP, and EECDC are thought about. assessed the presentation of the MK-implies calculation in contrast with the ECRA, SECA-M, k-means, Regard, and Drain calculations. To appropriately look at the viability of the MK-implies calculation and the k-implies approach, various boundaries are assessed. MK-implies fundamentally broadened the sensor organization's lifetime when contrasted with ECRA, SECA-M, K-means, Regard, and Filter. When compared to other algorithms, MK-means transfers significantly more data packets per round, enhancing throughput, network stability, and communication connection reliability. MK-means considerably reduces the quantity of control message transmissions required during clustering by reducing the number of clustering processes that must be carried out during network operation. To address the WSN's new network requirements and challenges, significant research and standardisation activities will be needed in the coming ten years.

REFERENCES

1. C. Alcaraz, P. Najera, J. Lopez, and R. Roman. Wireless Sensor Networks and the Internet of Things: Do We Need a Complete Integration? 1st International Workshop on the Security of the Internet of Things (SecIoT10) 2010.
2. Sourabh Bharti, K.K. Pattanaik. Task Requirement Aware Pre-processing and Scheduling for IoT Sensory Environments. Ad Hoc Networks 2016; 50: 102 –114.
3. PratyayKuila, PrasantaK.Jana. Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach. Engineering applications of artificial intelligence 2014; 33: 127-140.
4. Bajaber, F. and I. Awa (2014). An efficient cluster-based communication protocol for wireless sensor networks, Telecommunication Systems, Vol. 55, No. 3, pp. 387-401.
5. Chang, J.Y. and P.H. Ju (2014). An energy-saving routing architecture with a uniform clustering algorithm for wireless body sensor networks, Future Generation Computer Systems, Vol. 35, pp. 128-140.

6. Nooshin Nokhanji, Zurina Mohd Hanapi, et al. An Energy Aware Distributed Clustering Algorithm Using Fuzzy Logic for Wireless Sensor Networks with Non uniform Node Distribution. Wireless Personal Communication 2015; 84:395-419.

7. L. Malathi, R.K. Gnanamurthy, et al. Energy efficient data collection through hybrid unequal clustering for wireless sensor networks. Computers and Electrical engineering 2015; 48: 358-370

8. R. Logambigai, A. Kannan. Fuzzy logic based unequal clustering for wireless sensor networks. Wireless Networks 2015; 22: 945-957.

9. Nokhanji, N., Z.M. Hanapi, S. Subramaniam and M.A. Mohamed (2015). An Energy Aware Distributed Clustering Algorithm Using Fuzzy Logic for Wireless Sensor Networks with Non-uniform Node Distribution, Wireless Personal Communications, Vol. 84, No. 1, pp. 395-419.

10. Rao, P.C.S. and H. Banka (2015). Energy efficient clustering algorithms for wireless sensor networks: novel chemical reaction optimization approach, Wireless Networks, pp. 1-20.

11. Wu, Y.C. and C.C. Tuan (2015). K-Hop Coverage and Connectivity Aware Clustering in Different Sensor Deployment Models for Wireless Sensor and Actuator Networks, Wireless Personal Communications, Vol. 85, No. 4, pp. 2565-2579.

12. Yu, T., S. Liu and Z. Zhang (2015). A Cluster-based Uneven Grid Data Aggregation in Wireless Sensor Networks, Journal of Computational Information Systems, Vol. 11, No. 19, pp. 7055-7062.