An Extensive Survey on Clustering Algorithms for Data Aggregation in WSNs

Shailja Barmaiya [1], Akhilesh Bansiya [2],
[1,2]Department of Engineering Computer Science & Engineering,
RKDF University, Bhopal, India

Abstract: Now-a-days, WSN is considered as one of the developing technology since it incredibly helps individuals by offering sensing, figuring and networks abilities and empower people to have a nearly connection with nature wherever they go. This network is basically characterized as the gathering of hubs which are composed into an agreeable system Wireless sensor arrange (WSN) explore focuses on working with little, unassuming, multi-utilitarian sensor hubs that can detect, process, and impart. WSNs have various imprisonments appeared differently in relation to Ad-Hoc arranges with respect to its sensor nodes' ability of memory stockpiling, preparing and the accessible vitality source. These are light weight vitality compelled gadgets that work with little breaking point DC source. The energizing or substitution of vitality wellsprings of the sensor nodes is here and there difficult or even unrealistic.

Keywords: Wireless Sensor Network, Energy Clustering Algorithm, Distance, Density.

I Introduction

The term "wireless" has become a generic and all-encompassing word used to describe networks in which electromagnetic waves to carry a signal over part or the entire network path. Wireless technology can able to reach virtually every location on the surface of the earth. Due to tremendous success of wireless voice and messaging services, it is hardly surprising that wireless network is beginning to be applied to the domain of personal and business computing. Ad- hoc and Sensor Networks are one of the parts of the wireless network [1, 2].

In ad-hoc network each and every nodes are allow to communicate with each other without any fixed infrastructure. This is actually one of the features that differentiate between ad-hoc and other wireless technology like cellular networks and wireless LAN which actually required infrastructure based network like through some base station.

Wireless sensor network are one of the category belongs to ad-hoc networks. Sensor network are also composed of nodes as presented in Figure 1. Here actually the node has a specific name that is "Sensor" because these nodes are equipped with smart sensors. A sensor node is a device that converts a sensed characteristic like temperature, vibrations, pressure into a form recognize by the users. Wireless sensor networks nodes are less mobile than ad-hoc networks. So mobility in case of ad-hoc is more. In wireless sensor network data are requested depending upon certain physical quantity [3]. So wireless sensor network is data centric.

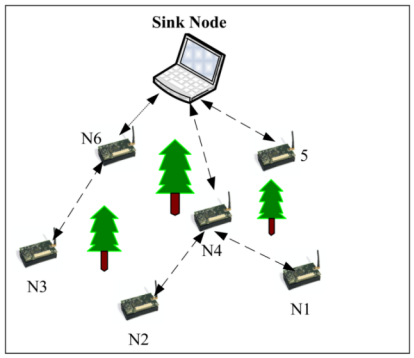

Figure 1: Basic Diagram of Wireless Sensor Networks

Clustering is the process of segregating the sensor nodes into virtual groups. Each one cluster is administered by a node called as cluster-head (CH) and different nodes are implied as member nodes. Clustered nodes do not communicate straightforwardly with the base station, but they need to transmit the gathered information through the cluster-head. The CH tries to aggregate the received data, received from the cluster members and forwards it to the BS [4]. Thus, it minimizes the energy utilization and a number of messages imparted to the base station. Likewise, the communications traffic in the network is lessened. The amazing result of clustering the sensors in a network helps in extending the lifetime of the network. Clustering is the hierarchical procedure followed in a network, made to streamline the communication process of the network. It prompts the presence of an incredible number of task-specific clustering protocols [5, 6].

In clustering as shown in Figure 2, the nodes are divided into different clusters based on certain heuristics, where one cluster-head is present for each cluster. All the member nodes transmit data to their respective cluster-heads, where the cluster-head performs the data aggregation and forward to the base station. As the nearby nodes inside one cluster may sense the same data, the duplicate data can be eliminated at the cluster-heads by the data aggregation technique. Subsequently, it helps in energy saving and re-utilizing the bandwidth in the process of clustering.

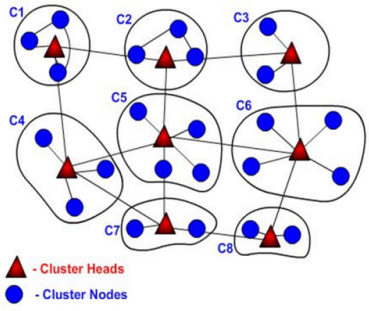

Figure 2: An Example of Clustering

Advantages of Clustering are:

- Transmit aggregated information to the BS
- Lesser number of nodes involved in transmission
- Valuable Energy utilization
- Versatility for vast number of nodes
- Diminishes communication overhead
- Productive use of resources in WSNs.

II Cluster-Based Data Aggregation

In cluster-based data aggregation, nodes are grouped into clusters, with one cluster head for each cluster. Members of a cluster send packets to their cluster head via single-hop or multi-hop network; the cluster head aggregates received data, and forwards the results to the sink, also via single-hop or multi-hop network. Generally the data aggregation can be classified on basis of network topology, quality of services, network basis and many more. For the current research, the techniques are being discussed based on the network topology. In this topology, the data aggregation technique is categorized into two parts which are flat and hierarchical network. Further hierarchical network is sub-classified into two parts which are cluster based, chain based, tree based and grid based [7–9].

2.1 Chain based Data Aggregation Techniques

It has been discussed that cluster members send the information to cluster head and further it transfer the aggregated information to sink. In case if the distance between cluster head and sink is far then it consumes more energy to communicate the sink whereas in the case of Chain based data aggregation the information is sent only to its closest neighbor.

2.2 Tree based Data Aggregations Technique

In tree based network, the nodes are organized in the form of tree topology where sink is considered as a root. Aggregation is carried out by constructing aggregation tree where source nodes are referred as leaves, and rooted at sink. Here data flow starts from leaves nodes and end at sink. So here all the intermediate nodes carry out the aggregation process and finally transfer to root (that is, sink). The main aim of the tree based approach is constructing an energy efficient tree.

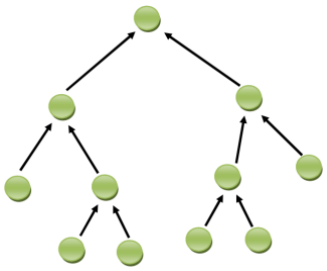

Figure 3: Tree Energy based Aggregation Technique

III Related Work

Lin *et al.* [10] Wireless sensor networks (WSNs) are wireless networks which consist of distributed sensor nodes monitoring physical and environmental conditions. Due to the energy limit of sensor nodes, prolonging lifetime of wireless sensor networks (WSNs) is a big challenge. In this paper, we propose a new clustering method called Density, Distance and Energy based Clustering (DDEC) to improve network performance. DDEC partitions the network into clusters with similar member number, so as to achieve load balancing. Then a cluster head is selected for each cluster based on three criteria: residual energy, distance and density, which achieves to minimize intra-communication cost and prolong cluster lifetime. In our performance analysis, they compared DDEC with another clustering method called DDCHS. The results showed that DDEC outperforms DDCHS in terms of alive node number and energy consumption.

Rajeswari *et al.* [11] In wireless sensor network (WSN), during fault detection and recovery, average energy loss, and message loss occurs including the link failure. Also, the number of faulty nodes and the traffic overhead is increased with the size of WSN. In order to overcome this issue, in this study, a cluster-based fault tolerance technique using genetic algorithm is proposed. Here the network is clustered according to energy-efficient distance-based clustering algorithm. For each cluster head, a set of backup nodes are selected using genetic algorithm based on the sponsored coverage and residual energy parameters. This helps in detecting the faults occurring in cluster members and cluster

Table 1: Related Work

S.No.	Title	Author	Year	Approach
1.	Density, distance and energy based clustering algorithm for data aggregation in wireless sensor networks	Lin *et al.*	2017	They proposed a new clustering method called Density, Distance and Energy based Clustering (DDEC) to improve network performance.
2.	Genetic algorithm based fault tolerant clustering in wireless sensor network	Rajeswari *et al.*	2017	A cluster-based fault tolerance technique using genetic algorithm is proposed. Here the network is clustered according to energy-efficient distance-based clustering algorithm.
3.	RRDVCR: Real-time reliable data delivery based on virtual coordinating routing for Wireless Sensor Networks	Venkatesh *et al.*	2016	They proposed a Real-Time Reliable Data delivery based on Virtual Coordinates Routing (RRDVCR) algorithm, based on the number of hops to the destination rather than geographic distance.
4.	Ferry based data gathering in Wireless Sensor Networks	Vanarotti *et al.*	2016	They are combining the concept of ferry to the WSN. In general, ferries are small commutator boats which carry or take the peoples from different places to reach to their destination within a particular region considering water as a medium
5.	A Clustering Algorithm for WSN to Optimize the Network Lifetime Using Type-2 Fuzzy Logic Model	Pushpalatha *et al.*	2015	They proposed a cluster head election algorithm using Type-2 Fuzzy Logic, by considering some fuzzy descriptors such as remaining battery power, distance to base station, and concentration, which is expected to minimize energy consumption.
6.	Localization algorithms based on hop counting for Wireless Nano-Sensor networks	Tran-Dang *et al.*	2014	Two ranging algorithms based on hop-counting methods are developed to estimate the location of every nanosensor within certain area and distance between nodes in the networks.

heads. Simulation results show that the proposed technique minimises the energy loss and overhead.

Venkatesh *et al.* [12] Real-time industrial application requires routing protocol that guarantees data delivery with reliable, efficient and low end-to-end delay. Existing Routing(THVR) [13] is based velocity of Two-Hop Velocity and protocol relates two-hop velocity to delay to select the next forwarding node, that has overhead of exchanging control packets, and depleting the available energy in nodes. We propose a Real-Time Reliable Data delivery based on Virtual Coordinates Routing (RRDVCR) algorithm, based on the number of hops to the destination rather than geographic distance. Selection of forwarding node is based on packet progress offered by two-hops, link quality and available energy at the forwarding nodes. All these metric are co-related by dynamic co-relation factor. The proposed protocol uses selective acknowledgment scheme that results in lower overhead and energy consumption. Simulation results shows that there is about 22% and 9.5% decrease in energy consumption compared to SPEED [14] and THVR [13] respectively, 16% and 38% increase in packet delivery compared to THVR [13] and SPEED [14] respectively, and overhead is reduced by 50%.

Vanarotti *et al.* [15] Wireless Sensor Networks are widely used in various applications like maintenance of ecosystems, in militaries, hospitals, etc Also WSNs are used with the data gathering methods to enhance the resources and energy which lack in some performance parameters. So here, we are combining the concept of ferry to the WSN. In general, ferries are small commutator boats which carry or take the peoples from different places to reach to their destination within a particular region considering water as a medium. Thus the ferries are mobile device which takes the data or information from different sensor nodes, aggregate it together and transmit to the base-station travelling through the shortest distance considering WSN as a medium. So an efficient data collection algorithm for the heterogeneous sensor nodes i.e., sensor nodes, cluster heads along with a ferry node is proposed. This proposed work can then be calculated in terms of parameters as network lifetime, total energy consumption and data travel time.

Pushpalatha *et al.* [16] In past few years the use of Wireless Sensor Networks (WSNs) are increasing tremendously in different applications such as disaster management, security surveillance, border protection, combat field reconnaissance etc. Sensors are expected to deploy remotely in huge numbers and coordinate with each other where human attendant is not practically feasible. These tiny sensor nodes are operated by battery power and the battery operated sensor node cannot be recharged or replaced very easily. So, minimization of energy consumption to prolong the network life is an important issue. To resolve this issue, sensor nodes are sometimes combined to form a group and each group is known as a cluster. In each cluster, a leader node is elected which is called as the cluster head (CH). When any event is detected, each node senses the environment and sends to the respective cluster heads. Then cluster heads send the information to the base station (BS). So, appropriate cluster head election can reduce considerable amount of energy consumption. In this paper, we propose a cluster head election algorithm using Type-2 Fuzzy Logic, by considering some fuzzy descriptors such as remaining battery power, distance to base station, and concentration, which is expected to minimize energy consumption and extends the network lifetime.

Tran-Dang *et al.* [17] Wireless Nano Sensor net-

works (WNSN) consist of nanosensors equipped with nanotransceivers and nanoantennas to operate in Terahertz frequency band (0.1-10THz). Due to the peculiarities of this communication channel (such as, very short range of transmission (under lm), high interference, high path loss) and limited capabilities of nano-nodes (such as, computing, sensing, memory, energy), the existing ranging techniques designed for traditional wireless sensor networks are not longer used in the WNSNs. In this paper, two ranging algorithms based on hop-counting methods are developed to estimate the location of every nano sensor within certain area and distance between nodes in the networks.

The first technique uses flooding mechanism to forward the packets to all nodes in the networks and count number of hops between two measured nodes. To overcome the problems of high overhead, duplication packets, and waste of consumption energy, the second algorithm based on clusters is developed. In this way, all sensor nodes are grouped into different clusters. Cluster heads will communicate together and count the number of hops. The performance of the algorithms are analyzed in terms of estimated distance, delay by taking also account the energy constrains of nanosensors. The simulation results show that, by being aware of the limitations of nanosensors, the proposed protocols are able to support WNSNs with very high density in ranging and localizing.

IV Problem Identification

Generally, lifetime of network is defined as the time whenever the first node fails to send its information to base station. This issue can be resolved by implementing data aggregation technique as it decreases data traffic and further saves energy by merging multiple.

One of the most important aspects in WSNs is the routing techniques that are used to relay data among the nodes in a WSN. Routing has a major effect on the performance and efficiency of WSNs. Energy efficiency is one of the main challenges in developing routing techniques since sensor nodes have limited amount of energy. A popular technique in saving energy and extending network lifetime is clustering, which has the advantage of being able to configure the network based on the nodes energy requirements.

V Conclusion

In this survey paper study of various types of data aggregation techniques. and brief survey of wireless sensor networks. WSN includes large number of sensor nodes which transfer the data from one system to another system without making use of any wires. All these sensor nodes in the network are resource constraint, so because of this reason the lifetime of the network is limited. Thus, various researchers propose numerous protocols or approaches for increasing the lifetime of the wireless sensor networks. In this means, the data aggregation concept has been introduced in this study as it is one of the important techniques that enhances the network lifetime. In this study, various data aggregation algorithms in WSN are discussed. Further a comprehensive study of different data aggregation protocols are presented under the network architecture. The data aggregation algorithms discussed in this study mainly focuses on three concepts which are efficient routing, organization and data aggregation tree construction.

REFERENCES

[1] A. Forster, *What are Wireless Sensor Networks?* IEEE, 2016. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7906235

[2] S. Misra and S. Goswami, *Routing in Wireless Sensor Networks*. Wiley, 2014. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8044231

[3] R. R. Selmic, V. V. Phoha, and A. Serwadda, *WSN Architecture*. Cham: Springer International Publishing, 2016, pp. 37–81. [Online]. Available: https://doi.org/10.1007/978-3-319-46769-6_3

[4] N. Gupta, P. M. Pawar, and S. Jain, "Benchmarks for evaluation of wireless sensor network clustering," in *Smart Trends in Information Technology and Computer Communications*, A. Deshpande, A. Unal, K. Passi, D. Singh, M. Nayak, B. Patel, and S. Pathan, Eds. Singapore: Springer Singapore, 2018, pp. 11–20.

[5] S.-S. Yang, J.-S. Shim, Y.-H. Jang, Y.-W. Ju, and S.-C. Park, "Design of clustering algorithm for efficient energy management in wireless sensor network environments," in *Advanced Multimedia and Ubiquitous Engineering*, J. J. J. H. Park, S.-C. Chen, and K.-K. Raymond Choo, Eds. Singapore: Springer Singapore, 2017, pp. 607–612.

[6] W. Zhang and D. Zhang, "Research on the algorithm of management clustering for wireless sensor network," in *Advances in Electronic Engineering, Communication and Management Vol.1*, D. Jin and S. Lin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 39–44.

[7] S. Siva Ranjani, S. Radha Krishnan, C. Thangaraj, and K. Vimala Devi, "Achieving energy conservation by cluster based data aggregation in wireless sensor networks," *Wireless Personal Communications*, vol. 73, no. 3, pp. 731–751, Dec 2013. [Online]. Available: https://doi.org/10.1007/s11277-013-1213-x

[8] M. Shoaib, W.-C. Song, and K. H. Kim, "Cluster based data aggregation in vehicular adhoc network," in *Communication Technologies for Vehicles*, A. Vinel, R. Mehmood, M. Berbineau, C. R. Garcia, C.-M. Huang, and N. Chilamkurti, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 91–102.

[9] W.-S. Jung, K.-W. Lim, Y.-B. Ko, and S.-J. Park, "Efficient clustering-based data aggregation techniques for wireless sensor networks," *Wireless Networks*, vol. 17, no. 5, pp. 1387–1400, Jul 2011. [Online]. Available: https://doi.org/10.1007/s11276-011-0355-6

[10] H. Lin, R. Xie, and L. Wei, "Density, distance and energy based clustering algorithm for data aggregation in wireless sensor networks," in *2017 IEEE/CIC International Conference on Communications in China (ICCC)*, Oct 2017, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ICCChina.2017.8330504

[11] K. Rajeswari and S. Neduncheliyan, "Genetic algorithm based fault tolerant clustering in wireless sensor network," *IET Communications*, vol. 11, no. 12, pp. 1927–1932, 2017. [Online]. Available: https://doi.org/10.1049/iet-com.2016.1074

[12] Venkatesh, C. S. Sengar, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Rrdvcr: Real-time reliable data delivery based on virtual coordinating routing for wireless sensor networks," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016, pp. 2227–2234. [Online]. Available: https://doi.org/10.1109/CompComm.2016.7925096

[13] Y. Li, C. S. Chen, Y. Song, Z. Wang, and Y. Sun, "Enhancing real-time delivery in wireless sensor networks with two-hop information," *IEEE Transactions on Industrial Informatics*, vol. 5, no. 2, pp. 113–122, May 2009. [Online]. Available: https://doi.org/10.1109/TII.2009.2017938

[14] T. He, J. A. Stankovic, T. F. Abdelzaher, and C. Lu, "A spatiotemporal communication protocol for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 10, pp. 995–1006, Oct 2005. [Online]. Available: https://doi.org/10.1109/TPDS.2005.116

[15] G. C. Vanarotti, U. M. Kulkarni, and H. H. Kenchannavar, "Ferry based data gathering in wireless sensor networks," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, July 2016, pp. 165–170. [Online]. Available: https://doi.org/10.1109/ICATCCT.2016.7911985

[16] D. V. Pushpalatha and P. Nayak, "A clustering algorithm for wsn to optimize the network lifetime using type-2 fuzzy logic model," in *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, Dec 2015, pp. 53–58. [Online]. Available: https://doi.org/10.1109/AIMS.2015.19

[17] H. Tran-Dang, N. Krommenacker, and P. Charpentier, "Localization algorithms based on hop counting for wireless nano-sensor networks," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2014, pp. 300–306.