RESEARCH ARTICLE OPEN ACCESS

I.

A Study on Multi-Sink Data Aggregation in Wireless Sensor Network

Manjunath C R [1], Nagaraj G S [2], Abhishek G [3]
Research Scholar [1] & [3]
Department of Computer Science and Engineering
SET, Jain University
Professor [2]
RVCE, VTU
India

ABSTRACT
Wireless sensor network are mainly used for the purpose of monitoring large scale applications, like physical or environmental conditions, such as temperature, sound, pressure, etc. and to collectively data from different source node to sink. Aggregation is the process of collecting the data from sensor nodes to the destination which reduces the network traffic helps to reduce energy consumed by sensor nodes. According to the survey it is tedious job to aggregate the data with the single sink node for large area of network. The proposed paper is a study on issues related to the single sink and behavior of network when more than one sink is considered. With the increase in the number of sink and their placement in the network of sink plays a major rule in efficient data aggregation.
Keywords:- Data Aggregation, single-sink, Multi-Sink, Wireless Sensor Network, Clustering.

I. INTRODUCTION

A Wireless Sensor Network (WSN) comprises of a few spatially disseminated self-sufficient gadgets (sensor nodes) with detecting and correspondence capacities that agreeably sense physical or natural conditions, for example, temperature, sound, vibration, weight, movement or poisons at various areas and utilized as a part of uses, for example, ecological observing, country security, basic base frameworks, interchanges, fabricating and so on. WSNs are information driven systems that for the most part deliver a lot of data that should be steered over the systems. As sensor nodes are vitality obliged gadgets and the vitality utilization is by and large connected with the measure of accumulated information. Since vitality protection is a key issue in WSNs, Data combination and Data aggregation is misused keeping in mind the end goal to spare vitality. A methodology to enhance the steering assignment for the accessible handling limit can be given by the middle of the road sensor nodes along the directing ways. Information aggregation is characterized as the way toward accumulating the information from different sensors to take out excess transmission and give melded data to the base station. The primary objective of information aggregation calculations is to accumulate and total information in a productive way so that lifetime of the system increments by diminishing the quantity of bundles to be sent to sink or base station, assistant lessens the Correspondence expenses and vitality utilization. The steering convention of sensor systems is commonly apportioned into two sub routings: (1) level directing convention and (2) progressive (tree-based or group based) directing convention. In level steering conventions, information aggregation is expert by information driven directing where the sink for the most part transmits a question message to the sensors, through flooding though in the Hierarchical directing convention information aggregation and information combination is performed with a specific end goal to diminish the quantity of transmitted messages to the sink hub. Quantities of calculations have been Proposed to give information aggregation amid the directing in WSNs, greater part of them falls in either tree-based or group based calculations. Bunch based calculations with information aggregation and In-system preparing can accomplish huge vitality funds in WSNs and will be viable in drawing out the system Lifetime, can be either static grouping or element grouping sort. Static grouped sort systems partition the system proactively into numerous bunches while dynamic bunched sort systems make a

group responsively in the region of the occasion detecting nodes. Every one of the information are gathered and amassed by the bunch head (CH), and afterward sent to the sink. The primary favorable position of element sort over static sort that exclusive the fundamental nodes, will partake in the information aggregation, saving vitality of the other sensor nodes. Subsequently, aggregation rate for element grouped information aggregation is high.

A. Data Aggregation

The principle reason for the information aggregation is to decrease the force utilization by minimizing the quantity of information transmissions. Information aggregation is characterized as the way toward totaling the information from numerous sensors to dispose of repetitive transmission and give melded data to the base station. All the aggregation nodes gather information from their kids nodes and figure the aggregation esteem. At that point just the totaled qualities are sent towards the information sink. The total quality might be normal, most extreme (least), summation, and so forth which is figured by application necessities. .Information created from neighboring sensors is frequently repetitive and very associated. What's more, the measure of information created in vast sensor systems is normally colossal for the base station to prepare. Information aggregation normally includes the combination of information from different sensors at middle of the road nodes and transmission of the totaled information to the base station rather than "don't." The serial comma is favored: "A, B, and C" rather than "A, B and C." The sensors occasionally sense the information, process it and transmit it to the base station. The recurrence of information reporting and the quantity of sensors which report information ordinarily relies on upon the particular application. The proficiency of information aggregation calculations relies on upon the connection among the information created by various Data sources (sensor units). A connection can be either spatial or worldly. Aggregation addition is characterized as the measure of decrease in the correspondence movement because of the aggregation. The most vital element for aggregation is an all-around composed steering convention, delegated Classic (location driven) directing conventions normally forward information along the briefest way to the destination and Data driven steering conventions forward information in view of the parcel content and pick the following bounce keeping in mind the end goal to advance in-system aggregation and minimize vitality consumption.

B. Classification of Data Aggregation Mechanisms

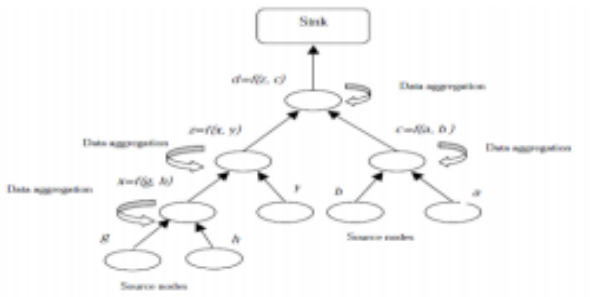
Agreeing the WSN, Data aggregation instruments can be named without structure, structure-based and half breed structure. At the point when sensor nodes are haphazardly sent in the earth, by nature, they require a sans structure component. At the point when sensor nodes are sent at an extensive scale, it gets to be troublesome as far as information aggregation and administration the WSNs. though the structure-based information aggregations are characterized with an arrangement of calculations, which isolates the system into gatherings and/or levels. This gathering oversees independently their information aggregation and diminished perspective of the whole system. However the structure-based instruments require an additional charge to sort out the system and to keep up association amid the system lifetime. Crossover structure consolidates attributes of both structure free and structure-based is relying on application.

Centralized Approach: Utilizing this convention, every node in the most limited conceivable course to the multi-jump remote node sends a location where the information driven methodology. Just the most grounded node is a pioneer in sensor nodes, sends information parcels. It gets the saint of the questioned information. Every middle of the road node has a youngster node to send information bundles tended to the pioneer. The quantity of messages that every node is equivalent to the total of the length of the outer way to an inquiry amid the telecast must be the best.

In-Network Aggregation: In-system multi-jump steering data through a system of worldwide procedure of gathering and asset utilization (specifically vitality) not exactly with the end goal of transitional nodes, and expansions the lifetime of the handling of information through the system. Without a lessening in the size decrease and size: organize the aggregation of the two measures. In-system size diminishment of the transmitting or sending the

parcel to the sink to decrease the length of the information bundles from one node to its neighbors in Balance and compacting process. In-system without reducing the size alludes to the way toward converging with an information parcel information bundles got from various neighbors, yet the benefit of handling the information.

Tree-Based Approach: It is conceivable to set the base crossing tree-based methodology, established in the leaves is viewed as a sink and source units, to keep up the building. Every node has a guardian node to forward the information. The surrenders are sink over to the guardian nodes of the nodes and the information stream starts.



Cluster-Based Approach: In Cluster-based methodology, the whole system is isolated into a few bunches. Every group has a bunch of select individuals from the bunch. The gatherer of the part of the bunch individuals from the group head locally and afterward transmit it to sink as an aftereffect of data got generally speaking.

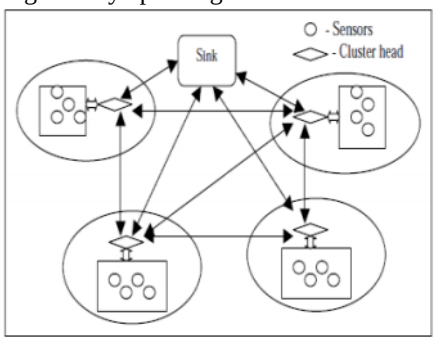


Fig: Cluster based approach

C. Issues related to the Single Sink Based Data Aggregation

1. The sensor information assembling, the "numerous to-one" sort of system movement that can prompt awkward nature in vitality expenses could bring about early end of the lifetime of the whole system to a more noteworthy degree. So open test to the vitality utilization in the meantime keeping up the equalization and vitality proficiency to develop the lifetime of the system.

2. The system engineering adaptability issue in the sink. In Single sink design, with its expansive number of sensors the single sink will have less capacity to gather every one of the information. Moreover, the normal additional jumps between the source and sink the sink close to the radio channel limit will get to be over-burden.

3. On account of single-sink system outline has been intensely sink, then you can't achieve the destination is stacked as an aftereffect of the disappointment of the information transmission.

C. Data aggregation in one sink

When there is a small area to be monitored for example an area where we can deploy 30-50 sensor nodes, the data collected by the sensor nodes will be aggregated to the sink node, the data aggregation process will go smoother for the small area. Single sink holds all the data related to the monitoring area

D. Data collecting techniques

D1. Structure

Information aggregation in sensor system depends on without structure and structure-based models. Without structure information aggregation conventions use arbitrary and probabilistic techniques to total the information. Structure-based information aggregation conventions use group based, tree-based and anchor based strategies to total the information.

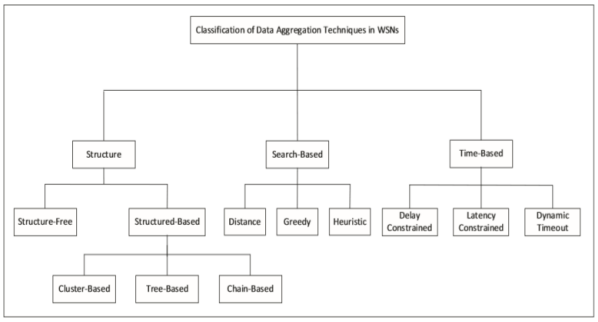


Fig: data aggregation techniques

D2. Structure-free

Structure-based information aggregation conventions bring about support cost because of topology

structures (like tree, cross section, and so on.) and differing system movement conditions. Sans structure information aggregation convention in light of spatial meeting and worldly joining use Data-Aware Anycast (DAA), and Randomized Waiting (RW) calculations to total the information. RTS and CTS messages taking into account the aggregation personalities (AIDs) of neighboring nodes are utilized as a part of the aggregation procedure. DAA utilizes anycast for the following jump transmission at the MAC layer. RW utilizes the counterfeit postponement for every information bundle along the course way and minimizes the impedance. Limit based without structure information aggregation depends on (i) holding up arrangement - holding up condition of information bundles to experience aggregation and (ii) constant anycast - moment directing choices taking into account the node move states and vitality levels.

D3. Structure-based

Cluster-Based: Vitality mindful occasional group information aggregation utilizes the deterministic strategy to choose the CH node as opposed to utilizing progressive or randomized techniques. The remaining vitality and scattered vitality are ascertained for every node in the system. The CH node with less vitality shows the Cluster_Head_Release message to the part nodes inside the bunch. On the off chance that the remaining vitality of CH node is more than the dispersed vitality, then the CH node proceeds with its dynamic part in the system, else the node with most extreme vitality is chosen as the CH node. For new sessions, the current groups are adjusted taking into account the vitality levels of CH nodes and dynamic sensor nodes. Vitality effective casing work for group based information accumulation use spatial and fleeting connections to spare the node vitality and decrease information excess in the system. This convention utilizes forecast and versatile methodologies which differ with system activity conditions. Information gathering in view of in-system aggregation minimizes the information excess in the system. Productive bunch based information aggregation utilizes (i) half breed group based information aggregation, that joins various aggregation conventions in light of node area and (ii) versatile bunch based information aggregation in which the sensor node switches its aggregation technique as per the fluctuating system conditions.

D4. Search-based

Greedy approach: Eager based strategies utilize the guideline of finding the best arrangement in light of the estimations of the present arrangement. The ravenous methodology is principally utilized as a part of little scale sensor systems and use separation based ascribes to discover the CH node in the area range. The node with greatest vitality is chosen as the CH node. Since, the part of CH node continues changing the covetous methodology chooses the best ideal CH node in the system. Vitality productive avaricious schemes defeat the issue of inactive nodes with less lingering vitality and least number of neighbor nodes.

Distance approach: Information aggregation conventions utilize short separation and multi-bounce ways to total the information in the system. As the separation between the node expands, the vitality utilization in information aggregation process increments. Delay-Aware methodology [9], intermittently finds the base separation between the nodes in the system. For every session, the non-dynamic nodes along the course ways are debilitated, and the nodes with most extreme vitality in the area of base station are chosen as the CH node. Extra repeaters are sent in the system to set up network with the CH nodes that are far from the base station. This technique minimizes the vitality of information authorities that utilization most brief course ways to total the information. This strategy arranges the sensor nodes and builds up the enhanced connections in the system.

Heuristic: Heuristic based Ant Colony Optimization (ACO) finds the dynamic node in the system in light of the inquiry procedure of the subterranean insect settlement. Ants lay a compound substance known as pheromone along the way traversals and the trailing ants utilize the pheromone power levels to cross towards the destination (looking for nourishment, province support, and so forth.). Vitality proficient scope utilizes the subterranean insect state and probabilistic location techniques to quantify the instability continuously sensor systems. Information aggregation procedure is performed taking into account the base course cost (pheromone check)

between the sensor nodes and the CH node in the system. The dynamic and inactive nodes are added to the disjoint set to cover the system zone in light of the leftover node vitality.

D5. Time-based

Delay-constrained: Delay compelled information aggregation utilizes the door node furnished with the vitality and handset signal quality that are more than the sensor nodes. The entryway nodes inside every group performs data handling and totals the information in the system. Information aggregation is performed as for the limit bundle size and the normal message delayin the system.

Latency-constrained: Latency constrained aggregation depends on the deferral of information bundles that are moving towards the sink node. The correspondence expense is a component of vitality utilization and deferral element over the connections in the system. This convention utilizes appropriated synchronous and offbeat strategies to evaluate the most extreme and least time to reroute the amassed information towards the sink node. Information aggregation procedure is dynamic when the information parcels in past session are sent to the following level towards the destination node.

Dynamic Timeout: Dynamic timeout information aggregation depends on the quantity of information bundles amassed at the information aggregation node. This convention utilizes neighborhood seek strategy to total the information and minimize the control overhead. This component Minimize the intricacy of information aggregation conventions. In the event that the information collected at the node is past as far as possible, then the convention times-out the information aggregation process and sits tight for the following cycle to total the information.

II. PROPOSED WORK

E. MULTI-SINK

The battery asset of the sensor nodes ought to be overseen proficiently, with a specific end goal to drag out system lifetime in remote sensor systems. In addition, in extensive scale systems with an expansive number of sensor nodes, various sink nodes ought to be conveyed, not just to expand the reasonability of the system, additionally to decrease the vitality dispersal at every node. We concentrate on the numerous sink area issues in huge scale remote sensor systems. Diverse issues relying upon the outline criteria are introduced. We consider finding sink nodes to the sensor environment, where we are given a period limitation that expresses the base required operational time for the sensor system. We utilize reproduction methods to assess the nature of our answer.

The whole and the exact zones of the sink nodes clearly impacts the lifetime of the sensor framework. In this way, for a monetarily conceivable theory, the fashioner should focus on right course of action of the sink nodes. Dependent upon the diagram criteria, there might be a couple of procedures for this issue.

E1. Locate the Best Sink Locations (BSL)

In numerous genuine sending situations, the originator will have a predefined spending plan allowed for the venture. Hence, the quantity of sink nodes is known preceding the arrangement stage. Since we know the quantity of sink nodes, which speaks to the quantity of bunches in the system, the main issue remains is the effective grouping of these sensor nodes. We call this issue as finding the "Best Sink Locations" issue (BSL). There are numerous great grouping calculations in the writing, which are basically named various leveled and non-progressive techniques. The non-various leveled techniques are normally alluded to as k-means grouping. Another bland technique is the self-sorting out maps, which is a universally useful unsupervised learning calculation. None of these techniques give the ideal number of groups that ought to be framed. The quantity of groups ought to be given as a choice parameter to the calculations. Precise area of the sink nodes are effortlessly found when the grouping calculation finishes. At whatever point the Euclidean separation is utilized as the bunching metric, then the focal point of mass of the nodes inside a group would give the area of the sink nodes. Contingent upon the needs of the steering calculation, power mindful separation measurements could likewise be utilized.

E2. Minimize the Number of Sinks for a

Predefined Minimum Operation Period (MSPOP)
In a few applications, the financial specialist may ask for the sensor system be operational for a predefined term. For instance, in farming applications, the field must be checked until the harvest. In this way, the sensor system ought to be dependable until the harvests grow up, and are procured from the field. The agriculturist is certainly going to send another sensor system in the following season. We call this issue as "Minimization of the quantity of Sink nodes for a Predefined least Operation Period" (MSPOP). With a specific end goal to take care of this issue, we need to compute the sensor system lifetime for any number of sink nodes. At that point, just the arrangement will be chosen, where the system lifetime surpasses the predefined restricting imperative, with the base number of sink nodes. The real issue in this issue is to choose the right number of sink nodes. The savage power strategy is to begin with stand out sink node, as expressed in [5]. While augmenting the quantity of sink nodes by one, the system lifetime is assessed. The hunt will stop, at whatever point the sought lifetime is come to.

E3. Locate the Minimum Number of Sinks while Maximizing the Network Life (MSMNL)
We may likewise attempt to augment the system lifetime however much as could reasonably be expected with the most practical speculation. We call this issue as "Minimization of the quantity of Sink nodes while amplifying the Network Lifetime" (MSMNL). When we don't have an earlier learning on the quantity of sink nodes nor the lifetime imperative, then we raise a combinatorial advancement issue. For this situation, the goal ought to consolidate the two options inside one capacity, where the monetary allowance saved for the sink nodes ought to by one means or another be associated with the lifetime of the sensor nodes. The underlying speculation for the sensor nodes ought to be used for the longest period. Here, the "expense per unit time" metric could be utilized. Keeping in mind the end goal to come to the required time period, we may need to play out some redeployment inside the system. We ought to include the expense of the redeployment to the underlying venture cost, which incorporates both the sensor and the sink nodes. The close ideal answer for this issue could be discovered utilizing some heuristic methods, as reenacted tempering or hereditary calculations.

III. A SOLUTION TECHNIQUE FOR THE PROBLEMS

F1. Deployment of the Sensor Nodes
Contingent upon the hidden application, there might be a few options on sensor system sending. In an agrarian application, they may be scattered by the agriculturist, in an about uniform way. In an ecological application like woodland flame identification, they may be dropped from a flying machine. In-house applications, they may be introduced by the development specialists by hand. By the by, the bunching calculation utilized ought to have the capacity to manage these arrangement situations.

F2. Finding Location Information
To compute sink areas, we should know the area of every individual sensor node. Area data can without much of a stretch be determined, utilizing focal or conveyed strategies. In the event that we utilize a focal methodology, the focal operator can play out disconnected area estimation.

F3. Collecting the Location Information from the Field
After the organization of the sensor nodes, we can utilize a portable terminal to gather the area information, which crosses the field. On the off chance that we don't have such an extraordinary reason terminal, then we need to introduce a sink node briefly. We can take this sink node back amid the last arrangement stage. When this is impractical because of physical confinements of the application, then we can utilize this node as an altered sink node in the last arrangement, and consider this node in the bunching calculation, without moving it.

F4. Finding the Best Location for K Sink Nodes
In the event that the quantity of sink nodes is known, The grouping calculation is dependable to find the sink nodes ideally. Contingent upon the organization dissemination, we can pick distinctive grouping calculations at this progression. In our usage, we have utilized the understood k-implies grouping

calculation.

IV. CONCLUSION

More number of sensor nodes are used to cover the sensing area a sensor node can cover up to 2 sq foot of distance so according to that node placement has to be done. Cluster head collects the data from the sensor node and send the data to the nearest sink node as the number of sink nodes are increased the energy efficiency will be more because the chances of saving the energy will be more.

For the small application single sink node and 30 sensor nodes are considered(in this project we have considered 30 sensors for a small area) For the small application two sink node and 50 sensor nodes are considered(in this project we have considered 50 sensors for a medium area) For the small application four sink node and 100 sensor nodes are considered (in this project we have considered 100 sensors for a large area).

REFERENCES

[1] Preethi Y. R, Manjunath C. R and Manohar M, "Data Routing in In-network Aggregation in WSN: a Cluster Based approach", International Journal of Modern Engineering Research, Vol. 3, Issue. 3, May.-June.2013 pp-1636-1640.

[2] VaibhavPandey, Amarjeetkaur&Narottam Chand, "A review on data aggregation techniques in wireless sensor network", in Journal of Electronics & Electrical Engineering, ISSN:0976-8106 vol.1 issue 2010,pp 01-08.

[3] Leandro Villas, AzzedineBoukerche, Heitor S. Ramos, " A Lightweight and Reliable Routing Approach for inNetwork aggregation in WSN", in Jounal of IEEE Tranactions on computer networks ,vol 38,Mar 2011,pp 393-422.

[4] Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey", (2006), Electrical Engineering and Computer Science. Paper22.

[5] Jamal N. Al-KarakiRazaUl-Mustafa Ahmed E. Kamal , "Data Aggregation and Routing In wireless sensor networks: optimal and heuristic algorithms" ,A survey IEEE wireless communications,vol 11,Dec 2004,pp 6-28.

[6] Woo-Sung Jung, Keun-Woo Lim, Young-BaeKo, "A Hybrid Approach for Clustering-based Data Aggregation in Wireless Sensor Networks", in IEEE transactions on networking,vol 11,pp2-16,Feb 1999.

[7] SonamPaldenBarfunga, PrativaRai, "Energy Efficient Cluster Based Routing Protocol for Wireless Sensor Networks", in International Conference on Computer and Communication Engineering (ICCCE 2012), 3-5 July 2012, Kuala Lumpur, Malaysia

[8] AratiManjeshwar and Dharma P. Agrawal , "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks", Center for Distributed and Mobile Computing, ECECS Department, University of Cincinnati, Cincinnati, OH 45221-0030,2001.

[9] Manjeshwar and D.P. Agarwal, "APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks", in proceedings of Int'l. Parallel and Distributed Processing Symposium (IPDPS), Fort Lauderdale, USA, 2002, pp 195-202.

[10] S. Lindsey and C.S .Raghavendra ,"PEGASIS: Power- Efficient Gathering in Sensor Information Systems",in proceedings of IEEE ICC 2001, vol. 3, Jun 2001, pp.1125-1130

[11] Shirshanka Das, AlokNandan, Michael G. Parker, Giovanni Pau and Mario Gerla, "Grido- An Architecture for a Grid-based Overlay Network"[1], Proceedings of the 2nd Int'l Conf. on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'05) 0-7695-2423-0/05 $20.00 © 2005 IEEE

[12] Shucheng Dai, Changjie Tang, ShaojieQiao, KaikuoXu, Hongjun Li and Jun Zhu, "Optimal Multiple Sink Nodes Deployment in Wireless Sensor Networks based on Gene Expression Programming[2]". College of Computer Science, Sichuan University, Chengdu, China, 610065 978-0-7695-3961-

4/10 $26.00 © 2010 IEEE.

[13] [13] RuiTeng and Bing Zhang , "Distribution of Sink-Node's Operation for On- Demand Information Retrieval in Wireless Sensor Networks"[3], National Institute of Information and Communication Technology, Kyoto, Japan.

[14] Haicheng Li, "Research on Distribution of Sink Nodes in Wireless Sensor Network"[4], Information Technology College Eastern Liaoning University Dandong ,China,118003, 978-0-7695-4360-4/10 $26.00 © 2010 IEEE.

[15] Ai-li Zhang and Yong-zhen Li, "BasedonMoreSink Nodes of Routing Protocolfor Wireless Sensor Networks", Network & Information Security Lab. Department of Computer Science and Technology Yanbian University Yanji, Jilin, China.

[16] Josnajose, manojkumar s and joycejose, "Energy Efficient Recoverable Concealed Data Aggregation in Wireless Sensor Networks"[6], 978-1-4673-5036-5/13/$31.00 © 2013 IEEE.

[17] James C. Bezedek, "FCM: the fuzzy c-means clustering algorithm[7]". Computers & Geosciences Vol. 10, No. 2-3, pp.191-203,1984.