Local Aggregation Scheme for Data Collection in Periodic Sensor Network

Neetu Verma, Dinesh Singh

Abstract: Data aggregation is an important technique for data collection & aggregation in WSN where sensor nodes sense the raw data and sends the aggregated data to the sink node. In a cluster based periodic network, sensor node senses the data on a specific time interval, performs local aggregation and send aggregated data to Cluster Head (CH). Various Local aggregation algorithms are used to remove redundant data at sensor nodes but local outlier detection problem is still unsolved. Therefore, a local aggregation algorithm has been proposed which uses the temporal correlation property of WSN to eliminate redundant and local outlier data which improves the data sent ratio and data quality. Sensor measurement is collected at different time interval of a sensor, exhibits temporal correlation because measurements varies with small or same difference (δ) and measurements are treated as similar measurements. In proposed local aggregation approach, each sensor node finds similar measurements of sensors with their frequency (number of occurrence) in a specific time interval (Temporal correlation). Set having higher frequency is selected and transmitted the average values of measurements that lie in the selected set to the cluster head. If sensors don't detect any reading between intervals it simple send a message 'data not found' instead of sending empty set. In this way we delete redundant and local outliers. The experimental result shows that algorithm improves the data quality and data sent ratio by eliminating redundant data and local outliers.

Key Words: Data aggregation, Data fusion, WSN, Temporal correlation.

I. INTRODUCTION

Data aggregation is the process of collecting and aggregating the data from various sensors [8]. Instead of sending raw data to sink node, aggregator nodes perform various aggregation techniques to eliminate the redundant data and send the fused information to the sink node. Data aggregation can be classify as Centralized Approach, Tree-Based Approach and cluster-Based, In-Network Aggregation [2][3] and Multi level aggregation[12]. In Centralized approach sensor node sends data using shortest possible route to a central node. Central node aggregates the data and send to the sink node. In Tree-Based Approach, tree like structure is establish where base station is considered as roots and sensor nodes are the leaves. Data from sensor node are aggregated by their parent node and forwarded to its successor and ultimately approaches to the root i.e base station.

In cluster-based approach, whole network is divided in to several clusters. For each cluster, one node is selected as cluster head (CH) while remaining node in cluster are work as sensor nodes [9][13]. Cluster head is responsible for aggregate data received from sensor node locally and then transmit the result to sink. In In-network aggregation data is processed in the network itself and send the required information to the sink. Size of data sent to sink node depends on two approaches of in- network aggregation: with size reduction in-network aggregation and without size reduction in network aggregation. In in-network size reduction aggregation data coming from different source is combined and compressed in order to reduce the information to be sent over the network while without size reduction technique, data coming from different nodes are merging into the same packet and sent over the network [1]. In multilevel aggregation, aggregation is performed at various levels. Local aggregation is performed at sensor node. In periodic network sensor node senses the data on a specific time interval, apply local aggregation to find out same or similar data or remove redundant data. The result of aggregated data is send to second level of aggregation [3]. Temporal correlated local aggregation is also used to delete local outlier which saves communication overhead [10]. This paper focus on local aggregation algorithm designed to delete similar value and considerably deviated values of sensed data using temporal correlation.

II. RELATED WORK

Jacques M. Bahi [5] define a data aggregation method for a tree based periodic sensors network that works in two level, at the node level aggregation called local aggregation and at the aggregator level. At every period *p,* sensor nodes send its aggregated data set to its appropriate aggregator. At aggregator level , aggregation method is used to aggregate all coming data set from various sensors nodes and forwarded to the sink node. In periodic network, the value of sensed data is similar to the previous value. Therefore, they use a link function to find out similar values.
Both the values are similar, if difference of measured sensed value by its previous value is less then threshold (δ).

$$link(y_{is_1}, y_{is_2}) = \begin{cases} 1 & if \quad \|y_{is_1} - y_{is_2}\| \le \delta, \\ 0 & otherwise. \end{cases}$$

The algorithm work as, when new sensed measurement is arriving in a period p , the sensor node check its similarity with existing measurements using link function. If a similar measure is found, it deleted the new measurements and incrementing by 1 in the frequency of corresponding measurement. After p period, each node sending *k* sets of measurements with their frequencies to the aggregator node.

Retrieval Number: B2602129219/2019©BEIESP
DOI: 10.35940/ijeat.B2602.129219
Journal Website: www.ijeat.org

3583

Published By:
Blue Eyes Intelligence Engineering
*& Sciences Publication*

Second level of aggregation is applied at aggregator node on data sets coming from different sensors and find similar sets using similarity functions. They use Jaccard similarity function to find the similarity between the sets of data measures at aggregator node. He also implemented In-network data aggregation algorithm [7] which is an effective technique to remove redundant data that preserve communication energy in sensor networks. Same local level aggregation is used, which he applied in their prior algorithm [5] but proposed a new frequency filtering approach which is applied on aggregator node to find similar set. Hassan Harb [6] proposed a cluster-based periodic sensor network (CPSN), in which sensor node periodically collect the data and sends its aggregated dataset to its Cluster head. Local aggregation is used at node level to delete similar or redundant data periodically. Aggregated dataset is then forwarded to its cluster head where second level aggregation is performed to find the similarity in collected datasets. They compare various similarity function used at second level. Very limited work is done on data aggregation at sensor's node for removal of local outliers and redundant data. Data collected or measured by WSNs may be affected by noise and atmospheric condition which leads to duplicate, missing and inconsistent data. This kind of unwanted data influence the quality of raw data and aggregated results. So, a local aggregation algorithm in periodic network is proposed to eliminate redundant data and local outliers.

III. PROPOSED NODE LEVEL AGGREGATION TO REMOVE REDUNDANCY AND OUTLIERS(ARRO)

The current work proposes a node level local aggregation works in a periodic sensor network. In a periodic cluster sensor network, sensor node senses the data on a specific time interval. Sensor nodes performs local aggregation to eliminate duplicate data and send the aggregated result to the cluster head or aggregator node. Cluster head perform second level aggregation on the data received by sensors nodes. The aggregated result is then finally transmitted to sink node. In the proposed algorithm, similarity of measured data in period 'p' is checked with the previous measured values. If the measured value is similar to the set of previous values then add the measured value in the set and its frequency count is incremented by 1, otherwise make new set which contain measured values with frequency count 1. At the end of specific interval, set having higher frequency is selected and average values of measurements that lie in the selected set is send to the cluster head. If no data found in a specific period then send a message 'No value' to the cluster head.
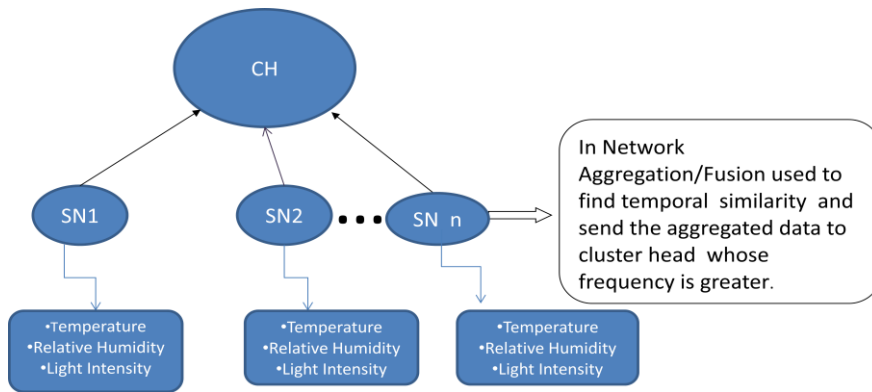
Figure 1: In network Aggregation/ Fusion at Sensor Node

Algorithm: Node level aggregation to remove redundancy and outliers

Require: new measure y_i, Set of previous measure M_i $(y_j…k)$.
Output: similar values having high frequency.
1. For every measure y_i in specific period p
2. If $|y_i-y_j|<= \delta$
3. Add y_i in M_i ,
4. $\quad$ F(M_i)= F(M_i)+1
5. Else
6. Add in to other set M_j
7. $\quad$ f(M_j)=f(M_j)+1
8. End for
9. If (M_i == 0) send a message " No value"
10. Else, Send the average of data in set s(i,j) whose frequency is higher among others.

Illustrative Example

For a time period of 5 minute, let us consider data sensed by sensor node 1 is

[10; 10.2; 10.3; 11; 11.1; 12; 11.3; 10.4; 12.1; 12.4].We use similarity threshold .5. Set is empty when first measurement 10 is arrive, no pervious values is there so it is placed in the first set and its frequency count is incremented to 1 .When next measurement arrives 10.2, its similarity values are smaller than threshold. Therefore it resides in the same set and increment the sets frequency. At the end of time period, result is [(10, 10.2, 10.3, 10.4, 4),(11,11.1,11.3,3),(12,12.1,12.4,3)], set having higher frequency count is selected and send average of all the data present in the set with their frequency count to cluster head. In this way similar data is removed. This algorithm also used to remove outlier. For example outlier is being placed by sensors node due to some error and the data sensed by sensor node is [10,10.2; 10.3; 11; 11.1; 12; 50; 10.4; 12.1; 12.4] ,sets will be shown as [10,10.2,10.3,10.4,(4)][11,11.1,(2)],[12,12.1,12.4,(3)] [50,(1)].

A set with higher frequency is selected to send the data. So we neglect outlier and those values which have low frequency. If frequencies of two sets are same then send the data of both sets to cluster head.

IV. EXPERIMENT RESULTS AND PERFORMANCE EVALUATION

- **Methodology**

To evaluate proposed local aggregation technique python based simulator is used & real data is obtained from Intel Berkeley Research Lab[14]. For this two files are being used , one file contain information about deployment of 54 sensors with their location while other file contains information about sensor's data of temperature, humidity, light and voltage for every 31 seconds shown as figure 2 . Local aggregation applied on sensor node where similarity of prior sensors data is checked with coming measurement. The collection of data contains values temperature, humidity, and light of single node for one day which is utilized for local aggregation.
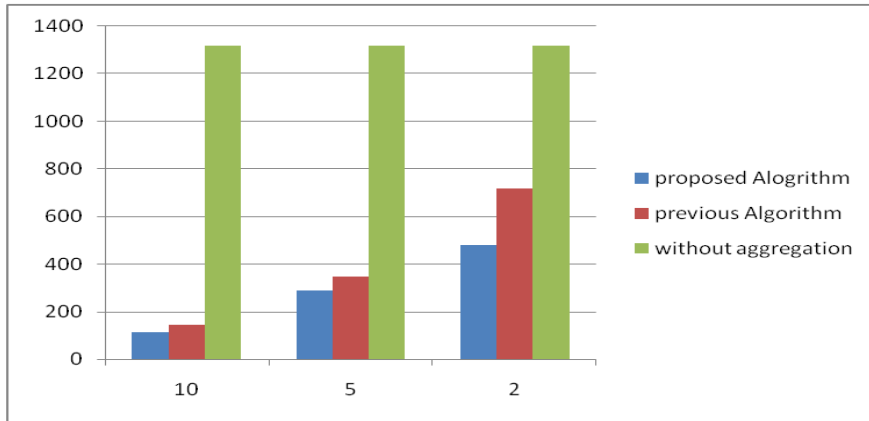
- **Simulation Model**

Cluster based periodic sensor network is used, where sensor nodes communicated with its corresponding CH and CH communicated with Sink node. All the member nodes collect their data and forward their data to CH in a periodic manner.

Simulation Parameters

Parameters	values
Number of nodes	1
Data in sensor node	1317
Outliers	50
Simulation duration	One day
Similarity Context (delta)	1,2 ,5,
Waiting time	2,5 and10 minutes

	Date	Time	Slot	Node id	Temperature	Humidity	Light	Voltage	
1	Date	Time	Slot	Node id	Temperature	Humidity	Light	Voltage	
2	03/03/2004	00:27,0	12005	1	18,3812	29,9691	1,38	2,63964	
3	03/03/2004	01:25,6	12007	1	18,4008	29,8271	1,38	2,63964	
4	03/03/2004	01:56,5	12008	1	18,4008	29,7561	1,38	2,63964	
5	03/03/2004	02:25,8	12009	1	18,4008	29,8271	1,38	2,63964	
6	03/03/2004	02:59,0	12010	1	18,391	29,8271	1,38	2,63964	
7	03/03/2004	03:27,7	12011	1	18,3812	29,7561	1,38	2,63964	
8	03/03/2004	03:57,7	12012	1	18,3812	29,7916	1,38	2,63964	
9	03/03/2004	04:28,7	12013	1	18,3812	29,8271	1,38	2,63964	
10	03/03/2004	05:57,4	12016	1	18,3714	29,8981	1,38	2,63964	
11	03/03/2004	06:28,1	12017	1	18,3616	29,8981	1,38	2,63964	
12	03/03/2004	07:00,1	12018	1	18,3616	29,8981	1,38	2,63964	
13	03/03/2004	07:55,2	12020	1	18,3616	29,8981	1,38	2,63964	
14	03/03/2004	08:25,3	12021	1	18,3616	29,8981	1,38	2,63964	
15	03/03/2004	08:59,9	12022	1	18,3714	29,8981	1,38	2,62796	
16	03/03/2004	09:55,3	12024	1	18,3616	29,8271	1,38	2,63964	
17	03/03/2004	10:25,3	12025	1	18,3616	29,7916	1,38	2,63964	
18	03/03/2004	10:55,8	12026	1	18,3518	29,8981	1,38	2,63964	
19	03/03/2004	11:27,0	12027	1	18,3616	29,9336	1,38	2,63964	
20	03/03/2004	11:58,4	12028	1	18,3518	29,9691	1,38	2,63964	
21	03/03/2004	13:27,0	12031	1	18,3518	29,9691	1,38	2,63964	
22	03/03/2004	14:01,4	12032	1	18,3518	29,9691	1,38	2,62796	
23	03/03/2004	14:25,9	12033	1	18,3518	30,0401	1,38	2,63964	
24	03/03/2004	14:59,1	12034	1	18,342	30,0046	1,38	2,63964	

Figure 2: Real Data Set

- Parameters used to evaluate the performance: Various parameters are used to evaluate the performance of our algorithm and also demonstrated the comparison of proposed algorithm with existing algorithm.
- (a) Waiting Time: sensor node wait (2, 5, 10 minutes) for sensor measurements and send aggregated data to cluster head.
- (b) Threshold: Similarity function δ(1,2 and 5) to check the similarity.

1. Amount of data sent to the CH:

Figure 3: Amount of Data sent to the CH

Sensor node sends their sensed data to CH. If data sent without aggregation and assumed that sensed data arrived at CH without packet drop then CH receives all data sent by the sensor nodes. Local aggregation is applied to deliver aggregated data to the CH where amount of sent data or aggregated data depends on the time interval (t). When time interval increases there are more chances to increase the similarity in measurement and reduction in sent data. If time interval increases, amount of aggregated data decreases. The proposed algorithm is compared with the algorithm defined by Jacques M. Bahi [4] shown in fig 3,which shows that fewer amounts of data was sent to the cluster head as compared to previous algorithm. In one day, one sensor node sensed 1317 measurements. The data sent to the CH via without aggregation, pervious algorithm and proposed algorithm are 1317, 143 and 115 when t=10.
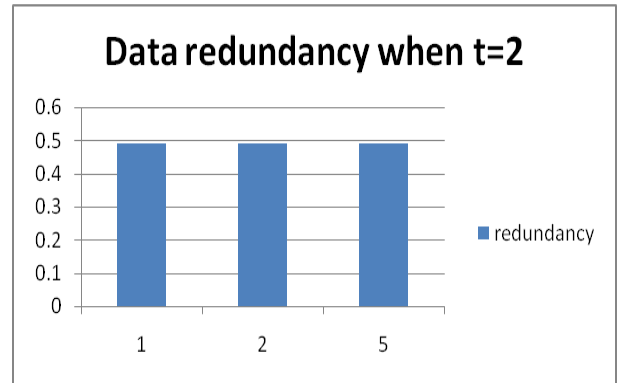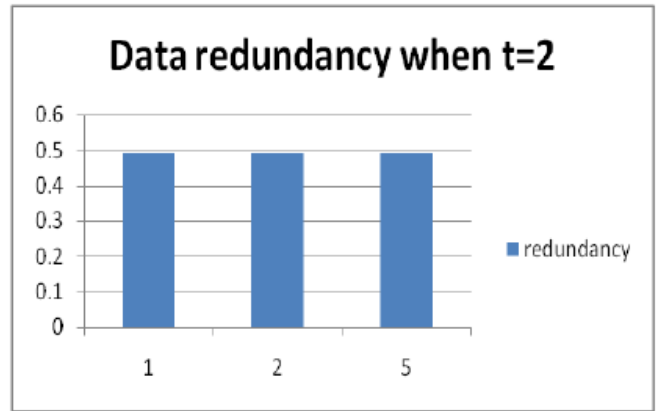
2. Redundancy factor:
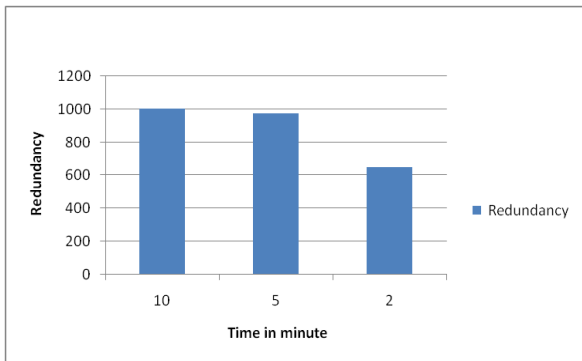
Figure 4: Redundancy varies with time
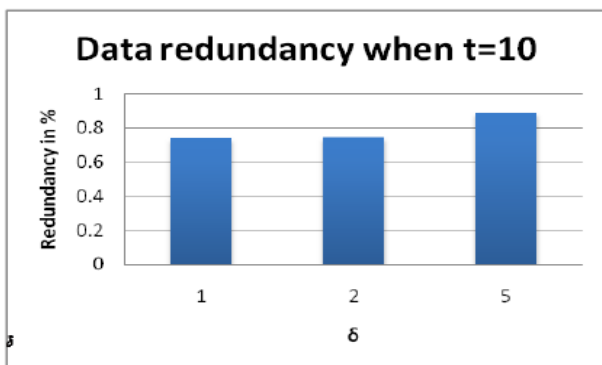
Figure5: Redundancy Varies with δ when t=10

Figure 6: Redundancy Varies with δ when t=2

Figure 7: edundancy Varies with δ when t=5

Redundancy means to find out same or similar values and it is affected by two parameters time and δ. If time period t is small than there is less amount of data to transmit therefore less chances to find similar values in contrast to this, if t is larger similarity of values is also larger as shown in figure 4. We also analyzed using figures 5,6,7 that when δ increases redundancy is also increases. So, redundancy is increased when t and δ increased.

3. Data Accuracy:

Data accuracy may be determined by data received at receiver side. When transmitting large sensor data over network, accuracy may increases.

In non aggregation algorithm, data are not aggregated and sink node received all data when there is no packet drop.

But, since when more data is transferred, it required more energy, so there is tradeoff between accuracy and energy consumption. Main motive of our algorithm is to reduce transmission between network and save energy as represented in figure 9. Data accuracy by previous algorithm is 10.8 whereas it is 8.7 % by our algorithm when t=10.
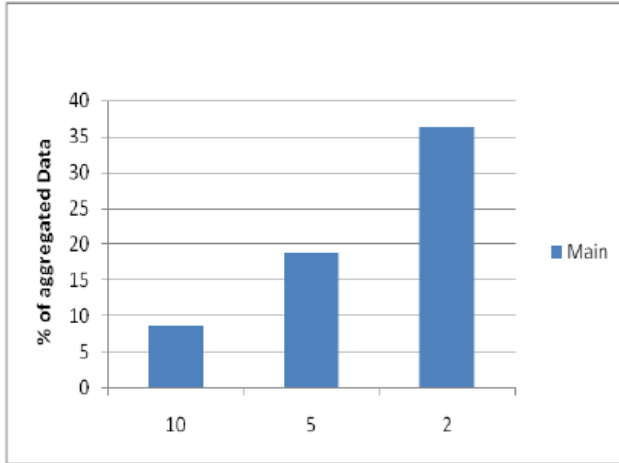

Figure 8: Percentage of aggregated data sent by Sensor node to CH
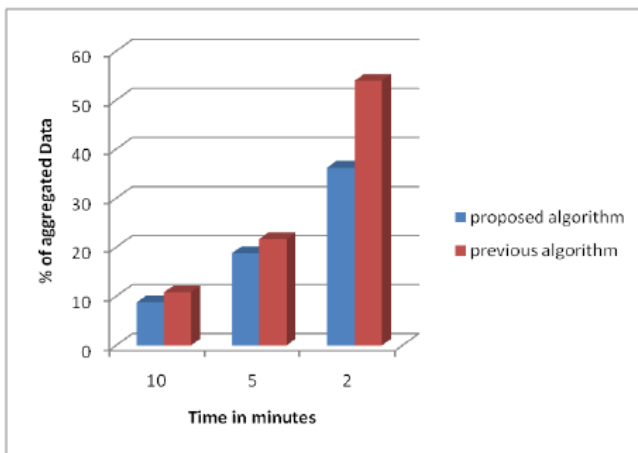

Figure 9: Comparison of aggregated data sent by Sensor node to CH

4. Outliers

This algorithm is also used to detect local outlier and send minimized data to the CH. Local outlier can easily removed by this algorithm, for example outlier is being placed by sensors node due to some error such that data sensed by sensor node is [10,10.2; 10.3; 11; 11.1; 12; 50; 10.4; 12.1; 12.4] ,sets will be shown as

[10,10.2,10.3,10.4,(4)][11,11.1,(2)],[12,12.1,12.4,(3)] [50,(1)]. A set with higher frequency is selected to send the data. So we neglect outlier and those values which have low frequency. If frequencies of two sets are same then send the data of both sets to cluster head. In this way outlier values are removed which also improves the data sent ratio and data quality.
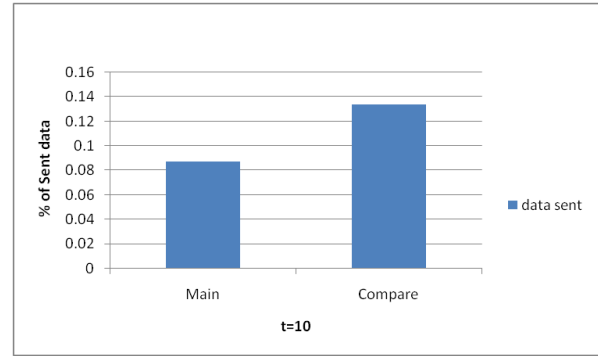

Figure 10: Percentage of data sent

V. CONCLUSION:

Local aggregation algorithm is sensor node based technique used to minimize the data transfer between sensor nodes to CH. In this paper, proposed local aggregation algorithm is used to remove redundant data as well as the local outlier and sent the minimized data to the CH. We applied experiments on real data measures and found that minimum number of data is transferred to the CH as compared to previous algorithm. We also analyzed that similarity between measurements depend on time and similarity function (δ). If the value of time and δ rises, there is possibility that large number of similar measurement comes in a single set, which increases the redundancy. Less data is transferred to the CH when more redundancy is found between sensor measurements. So redundancy also plays an important role to transfer minimized data. It has been also conclude that algorithm is better to find out local outlier, which is also used to improve the data quality and data sent ratio. In future work a cluster based algorithm is proposes which exploits spatial property of WSN to provide accurate decision about event to the Sink node.

REFERENCES:

1. Elena Fasoloy, Michele Rossiy, J¨org Widmer*?* and Michele Zorzi "In-network Aggregation Techniques for Wireless Sensor Networks: A Survey", IEEE communication, vol.14, issue 2,pp. 70 – 87, May 2007
2. Rocky Dunlap ,"In-Network Aggregation in Wireless Sensor Networks", College of computing. Georgia Institute of Technology. USA,2004
3. Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey" *Electrical Engineering and Computer Science*, 2006
4. Jacques M. Bahi et. al. "A Two Tiers Data Aggregation Scheme for Periodic Sensor Networks" *Ad Hoc & Sensor Wireless Networks*, Vol. 0 pp. 1–24, 2012
5. Runze Wan et.al, "Similarity-aware data aggregation using fuzzy c-means approach for wireless sensor networks", EURASIP Journal on Wireless Communications and Networking 2019
6. Hassan Harb et.al. "Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks" , IEEE Access,Vol 5,2017
7. Jacques Bahi, Abdallah Makhoul and Maguy Medlej, " An Optimized In-Network Aggregation Scheme for Data Collection in Periodic Sensor Networks", 11-th Int.Conf. on Ad Hoc Networks and Wireless, Serbia. pp.153 -166, 2012.
8. Gaurav Kawade et. al, " Confidentiality Protecting Hierarchical Concealed Data Aggregation for Wireless Sensor Network using Privacy Homomorphism", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019

9. P. Zurani, B. N. Mahajan, "Clustered Time Synchronization Algorithm for Wireless Sensor Networks", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-2, June 2012
10. Yang Zhang, Nirvana Meratnia, And Paul Havinga, "Outlier Detection Techniques For Wireless Sensor Networks: A Survey", IEEE Communications Surveys & Tutorials, Vol. 12, No. 2, Second Quarter, 2010
11. Vipul Narayan, A.K. Daniel," A Novel Protocol for Detection and Optimization of Overlapping Coverage in Wireless Sensor Networks", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019
12. Harsh Deep, Varsha, "Evaluation of Various Data Aggregations Techniques for Energy Efficient Wireless Sensor Networks", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-4 Issue-5, June 2015
13. Arvinda kushwaha, Mohd Amjad," Numerous Clustering Techniques with Application and Limitations in Wireless Sensor Network", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019
14. Http://Db.Csail.Mit.Edu/Labdata/Labdata.Html

AUTHORS PROFILE

Neetu Verma, is currently working as Assistant Professor in the Department of Computer Science & Engineering at Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat. She has more than 10 years of teaching & research experience in all. She has got Bachelors of Engineering & Masters of Engineering, both, from MDU Rohtak. She is in advanced state of her Ph.D from DCRUST Murthal. She has successfully guided numerous M.Tech scholars. She has published many research papers in reputed Journals & Conferences. Her area of expertise is in Wireless Sensor Network, Compiler design & Operating System.

Dinesh Singh, is currently working as Assistant Professor in the Department of Computer Science & Engineering at Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat. He has more than 15 year of teaching & research experience. He was successfully completed many research projects. He is also successfully guided numerous Ph.D & M.tech scholars. He has published many research papers in reputed international journals and conferences. His area of expertise is Wireless Sensor Network, Adhoc Networks.