IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

IMPACT OF EMERGING TECHNOLOGY TO IMPROVE THE NETWORK AGGREGATION FOR BUSINESS ORGANIZATIONS

PanduRanga Rao Arnepalli[1], Srinivas Aditya Vaddadi[2], AdithyaPadthe[3], Ramya Thatikonda [4]

Doctor of Philosophy, Information Technology, University of the Cumberlands, USA[1-4]

Abstract: In the current days, huge number of data are processed by the user in different emerging technologies with respect to data aggregation, it aggregates the bulk of confidential data to be more accurate and efficient as it is handled by business organization. In this process, the problem of data interruption and modification occurs and leads to data inaccuracy and unreliability as there are various confidential information persist in the various business organization. In the data aggregation, the data are divided into several group and with a role provisioning. To overcome the problem of data interruption and modification, propose a MG based network aggregation algorithm with the integration of emerging technology to gather and aggregate the data, energy efficiency can be improved and enhanced based on the input data. Then the data are gathered and routed to the destination where resource utilization is diminished and network lifetime gets expanded using a grouping-based central entity. In the performance analysis, the ranking the data and similarity are determined along with the evaluation of energy consumption and scheduling length compared with existing DICA and DICA extension.

Keywords: Organization, Data Aggregation, Fuzzy, Neural, data interruption, Resource utilization.

I. INTRODUCTION

As WSN will play a managing role for deploying the automated wireless networks and as the network consists of small number of nodes supported with low power consumption and less price [1]. These nodes are denoted as sensors where nodes can send the information using certain set of activity carried out by the sensor nodes and receives the information from the transmitting nodes in the netwok. This sensor wireless networks are used to perform various activity for different real-time applications such as, smart homes, Electronic health care, etc. Mainly, the sensor systems will perform node identification and monitor those activity, which has been used for various applications like, defence, smart based applications, etc [2-4]. There are several sensors will be used to detect various activity like, noise creation, pressure, body temperature, image extraction, etc., through which the information is extracted that are useful for various activity [5].

Generally, the global positioning system used several protocols for performing different activity and there is information get extracted for performing different activity. Sensors will perform several activities and process those data through the establishment of channel communications. There are different factors needed to perform sensor nodes in the network such as, reducing the cost, reduce the consumed energy, minimized data quantity, etc. while considering the sensor nodes, which is equipped with information processing by using any energy optimization model, which plays an important role in the network.

Application related to WSN to perform certain process such as, initially data collection process is performed and then those data are processed by considering the data near to the network application [6]. Mainly, the clustering algorithm to achieve resource imbalance factor to minimize the energy consumption among the sensor nodes. To provide the sensor node placement, node identification and cluster head selection are performed by minimizing the information quantity and network congestion. (Li et. al., 2001) perform Low Energy Adaptive Clustering Hierarchy (LEACH) clustering algorithm based on energy minimization through data aggregation in the network. The LEACH algorithm has two phases to complete the process: establishing the cluster phase and the network steady phase. In the initial phase, every node present in the network uses distributed environment and check the node to act as the header selection or not. Here the node uses certain threshold value to be defined as 0's and 1's. Then the node is selected as header and the listen to the selected node

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

to information broadcast. Then steady phase gets initiated by sending the cluster node to send the information by performing certain operation using Time Division Multiple Access (TDMA) [7].

While sending the information to the Base Station (BS), the data fusion algorithm is used along with LEACH protocol. This helps to balance the consuming energy and network node lifetime in the network. In this wireless network, i.e., Wireless Sensor Networks (WSN), the data aggregation process plays an important role and faces several challenges.

While processing the data aggregation, many works has been carried out on neural networks and fuzzy logic in WSN.

Comparative study of various data aggregation techniques based on strategy, delay, redundancy, average energy consumption, bandwidth overhead and proposed a new data aggregation model [8] which considers all these issues as represented in figure 1.

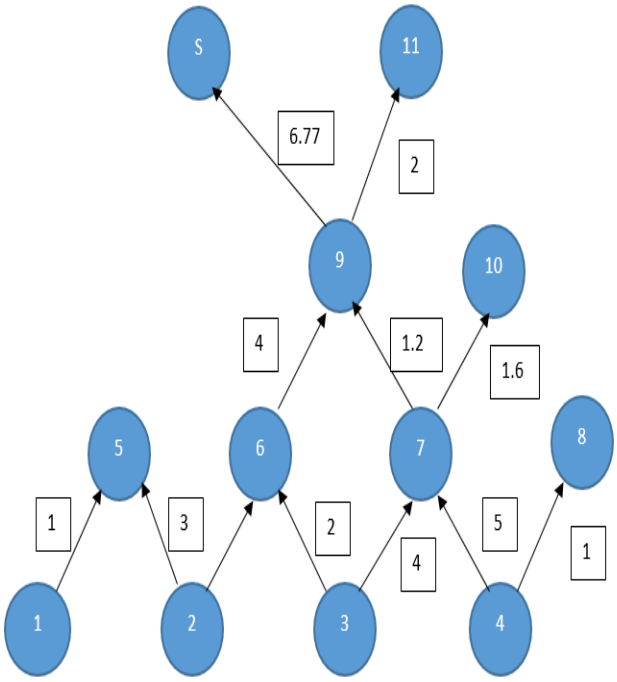

Figure 1. Scheduling and Tree Formation

We did not focus on secure data aggregation, but proposed secure data aggregation in WSN which can be referred for more information [9-12].

In this study, a model of energy efficient data aggregation i.e. MG based Network Aggregation (MG-NA) Algorithm is to be discussed and essential components are considered. Here we have applied cross layer genetic based artificial intelligence technique, where the data are gathered and routed to the destination where resource utilization is diminishing and network lifetime gets expanded using cluster based central entity.

The paper is organized as follows, literature of the existing works are discussed in section 2 and then formulation of the problem is discussed in section 3.

The proposed research methodology in section 4 and the performance metrics are analysed in section 5 Finally, the conclusion in section 6.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

Table 1. Existing Approaches applied based on the data aggregation

S.No.	Title of the paper	Proposed Methodology	Performance Metrics	Research Gaps
1	A dynamic hierarchical fuzzy neural network for a general continuous function [13]	Two stages genetic approach is proposed as it is associated with a dynamic-based hierarchical fuzzy neural network and FNN merged.	B-Spline Membership Function	Apply the proposed approach in to the real world applications
2	Clustering in Wireless Sensor Networks: Performance Comparison of LEACH & LEACH-C Protocols Using NS2 [14]	The clustering algorithm is proposed on the wireless sensor networks as it performs effective routing strategy.	Average latency, energy dissipation, network lifetime	Apply centralized approach and deterministic, the network lifetime and cluster number also gets increased
3	The Applications of Genetic Algorithms in Medicine [15]	Genetic algorithm is proposed as it provides some implications on various medical domains	Clinical Laboratory based on scheduling rotation	Involving the real-time applications like health care and hospital.
4	A new QoS routing algorithm based on self-organizing maps for wireless sensor networks [16]	Directed diffusion and Energy-Aware Routing is compared with the proposed SIR as it applies neural networks	Average dissipated energy, average delay	Including Artificial Intelligence inot the wireless sensor networks
5	Tool-supported design of data aggregation processes in cloud monitoring systems [17]	Z3 SAT solver is used as it aggregates the data as it uses the feature selection and composition.	CPU Usage, packet drop, open stack results,	Multi-level approach can be applied as it helps to manage and transaction

II . LITERATURE SURVEY

Base data aggregation functions as performed four methods: Sum, Minimum, Maximum and Averaging. These aggregate functions make to perform. Some operations based on ‘n’ number of rows.

As business organization is concern, there different level are applied in the organization and based this hierarchy, the data aggregation is summarized easily using database [18].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

In the minimum, the smallest value in the database table is identified using the function MIN.
SELECT MIN (`Marks`) FROM `students`;

The maximum value in the database table is identified using the function MAX.

SELECT MAX (`Marks`) FROM `students`;

In the summation, total sum of the values are calculated and printed using SUM function as it works with the numeric field [19].

There are four ways to do data aggregation in WSNs [20]:

- Centralized
- In-network
- Tree based
- Cluster based.

Based on table 1, it is observed that the multi-level approach has to be applied.

III. PROBLEM FORMULATION

As the data are collected and transmitted to main node i.e., base station in the network. Here, there is a lack of effective approach to aggregate the data from various nodes and as it takes longer to perform the aggregation, which leads to energy inefficiency.

As the different node components are used, it is difficult to schedule, scheduling load imbalance and select the node in the network. Finally, this may lead to data interruption and modification during the data communication.

IV. RESEARCH METHODOLOGY

The proposed MG-based NA algorithm as it is associated with scheduling and graph node formation. In this network, there are various heterogeneous as it contains different node attributes.

In this algorithm, the scheduling is performed along with parent selection based on a bottom-up approach, as it provides node aggregation in the network. In this algorithm,

a. Assign the node label
b. Determine how many packets are transmitted and packet type.
c. Adding the weight to the edge in the network

The weight of the link helps to determine the distance between the two nodes in the network. The aggregation factor of the network graph G is determined as,

$$\beta = \sum_{i=1}^{n} \frac{\beta_i}{n}, n = 1,2,3 \ldots \quad (1)$$

Then after applying the min max normalization approach, network is fuzzified as,

$$D = \frac{D - \min(A)}{\max(A) - \min(A)} [(\text{newmax}(A) - \text{new}\min(A)] + \text{newmin}(A) \quad (2)$$

Algorithm: MG Based Network Aggregation (MG-NA)

Input: Node Graph; Root Node (R_N)

Output: Graph Formation; R_N → node selection

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

1. Create the node graph by representing the R_N
2. If (Root ≠ 0)
3. {
4. Assign R_N in the graph node formation
5. Else
6. Create a new node 'N_N' into the graph 'G'.
7. While N_N is True then
8. {
9. Apply AddEdge as it initializes with (Source, Destination, weight)
10. (Source & Destination → Nodes 'N' belongs to G
11. Weight → Distance between the two nodes belongs to G)
12. 'G' Graph formation → Nodes are directly connected with Child node 'C_N'
13. }
14. For N_i = 1 to N Do
15. {
16. If (N_i> N_TS && N_i<= N_TM) then
17. {
18. N_i belongs to N_TS
19. }
20. }
21. Apply Min-Max Normalization
22. For N_i = 1 to N Do
23. {
24. While(N_i> N_TS) do
25. // Perform Scheduling and select the parent
26. }
27. Parent node Selecting Formation
28. For N_i = 1 to N Do
29. {
30. // Distance between the two nodes
31. }

Based on algorithm 1, there are specific steps are followed as follows,

A. Graph Formation:

i. Based on R_N, the node graph gets an initialization
ii. The node label is assigned
iii. The edge of node is assigned with a weight
iv. Based on the weight, the distance between the two nodes in the network

B. Parent node selection:

i. While the node is connected directly with the child node, the parent node candidate is determined.
ii. Based on the appropriate patent node-set, the neighbor node can be determined based on the candidate node dynamically
iii. The parent node is considered when the node has less number of neighbors dynamically.
iv. While checking the parent node, if the neighbor are equal based on the candidate,
    i. In the node graph of the wireless sensors, fuzzify the node formation
    ii. Based on the graph with node edge, weight of the node gets normalized by applying min-max normalization
    iii. The weight of the node is determined based on the edge membership
    iv. The membership gets determined based on the degree of the node.

In Algorithm 1,

Dataset of references (urls) to news web pages

TAGS: web pages, news, aggregator, classification, clustering

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111204

LICENSE: Public domain - Due to restrictions on content and use of the news sources, the corpus is limited to web references (urls) to web pages and does not include any text content.

The references have been retrieved from the news aggregator through traditional web browsers.

FILE ENCODING: UTF-8

FORMAT: Tab-delimited CSV files.

DATA SHAPE AND STATS:

Categorizing the business based on 152746

Fetching the information based on science and technology on 108465 115920 →news of business category

Information on the categorization on health care for
Information on categorizing the entertainment to determine the similarity in the cluster for 2076

Information on science and technology categorization to determine the cluster similarity for 1789 2019 →clusters of similar news for business category
Based on the categorization of healthcare to determine the cluster similarity.

The collection on the web pages is analysed as it contains the web links and some reference information. The web page URLs includes some web browsing history of two pages. The information collection includes 15516 of two pages of web history sessions as it divides the web into 946 cluster data.

The categorizing the business information for each page session for 6091
Categorization of entertainment information for the next web session history for two pages.

CONTENT

FILENAME #1: 1_Dataset.csv (102.297.000 bytes)
DESCRIPTION:
News pages

where:
ID → Numeric datatype
TITLE → web information title
URL → web URL
PUBLISHER→ Name of the web publisher

STORY Alphanumeric ID of the cluster that includes news about the same story

HOSTNAME Url hostname

TIMESTAMP→ information are published as it si based on the time approximation based on the number of milliseconds as there are variation of iteration 00:00:00 GMT, January 1, 1970

V. PERFORMANCE ANALYSIS

Based on figure 2 and 3, the Energy consumption and Average scheduling length are calculated based on the variation in the number of iteration.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

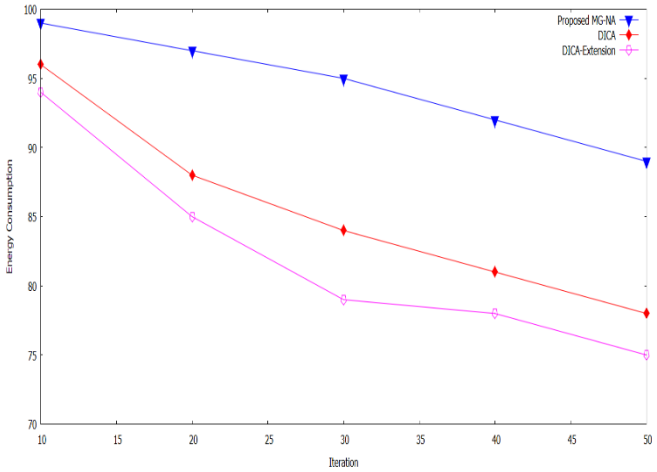

Figure. 2 Data sensitivity based on number of epcohs

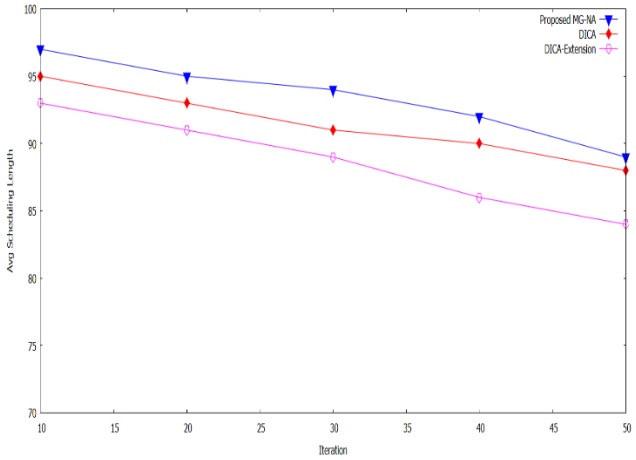

Figure 3. Data Specificity based on number of epochs

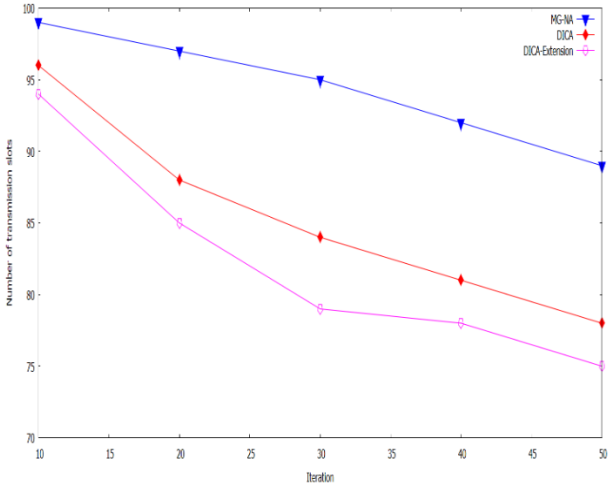

Figure 4. Number of transmission slots Vs Iteration

In the figure 4, the number of transmission slots are determined based on the variation in the number of iteration.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

Table. 1 Data Aggregation factor based on number of attributes

S.No.	Proposed MG-DA	DICA	DICA-Extension
10	0.985	0.968	0.945
20	0.965	0.952	0.915
30	0.954	0.947	0.895
40	0.925	0.898	0.865
50	0.918	0.875	0.845

Table. 2. Control overhead time based on number of iteration

S.No.	Proposed MG-DA	DICA	DICA-Extension
10	0.0045	0.0024	0.0018
20	0.0041	0.0023	0.0015
30	0.0038	0.0019	0.0013
40	0.0036	0.0018	0.0011
50	0.0031	0.0014	0.0009

Based on figure 5 and 6, the data sensitivity and data specificity are calculated based on the variation in the number of epochs.

VI. CONCLUSION

As the data are collected and transmitted to the main node i.e., base station in the network. Here, there is a lack of effective approach to aggregate the data from various nodes and as it takes longer to perform the aggregation, which leads to energy inefficiency. As the different node components are used, it is difficult to schedule, scheduling load imbalance and select the node in the network.

Finally, this may lead to data interruption and modification during the data communication.In the performance analysis, the ranking the data and similarity are determined along with the evaluation of energy consumption and scheduling length compared with existing DICA and DICA extension.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111204

REFERENCES

[1] M. Bagaa, M. Younis, D. Djenouri, M. Derhab& N. Badache. (2015). Distributed low-latency data aggregation scheduling in wirelesssensor networks. ACM Trans Sensor Netw 11(3), 1–36.
[2] M. Balakrishnan, E.E. Johnson. (2005). Fuzzy diffusion for distributed sensor networks. In Proceedings MILCOM IEEE Military CommunicationsConference,(pp 3023–3028).
[3] P. Chithaluru, F. Al-Turjman, M. Kumar, T. Stephan. (2020). IAREOR: an energy-balanced clustering protocol for implementinggreen IoT in smart cities. In Porceedings of Sustain Cities Soc.
[4] D. De Guglielmo, S. Brienza, G. Anastasi. (2016). IEEE 802.15. 4e:a survey. Computer Communication 88, 1–24.
[5] W. Fang, W. Zhang, Q. Zhao, X. Ji, W. Chen, B. Assefa. (2019). Comprehensive analysis of secure data aggregation scheme for industrial wireless sensor network.Computers, Materials & Continua, 61(2), 583-599
[6] F. Fanian&M.K. Rafsanjani. (2020) A new fuzzy multi-hop clustering protocol with automatic rule tuning for wireless sensor networks.Appl Soft Comput 89, 106115.
[7] C. Gherbi, Z. Aliouat,M. Ben mohammed. (2016).An adaptive clustering approach to dynamic load balancing and energy efficiencyin wireless sensor networks. Energies. 114, 647–662.
[8] S. Halawani, A.W. Khan. (2010). Sensors lifetime enhancement techniquesin wireless sensor networks-a survey.Journal Of Computing, 12(5), 34-47.
[9] Z. Han ,J. Wu, J. Zhang, L. Liu &K. Tian. (2014). A general self organized tree-based energy-balance routing protocol for wirelesssensor network. IEEE Trans NuclSci, 61(2), 732–740.
[10] H. Hayouni&M. Hamdi. (2017). A data aggregation security enhancing scheme in WSNs using homomorphic encryption. IntellAutomSoft Comput. 1–9.
[11] M. HosseinAnisi, J, Rezazadeh&M. Dehghan. (2008). FEDA: fault-tolerant energy-efficient data aggregation in wireless sensor networks. In Proceedings of 16th International conference on software, telecommunications and computer networks, Split, (pp 188–192).
[12] R.H. Hwang, C.C. Wang & W.B. Wang. (2017). A distributed scheduling algorithm for ieee 802.15. 4e wireless sensor networks. ComputStand Interfaces. 52, 3–70.
[13] W. Y. Wang, I. H. Li, S. C. Li, M. S. Tsai & S. F. Su, A dynamic hierarchical fuzzy neural network for a general continuous function. In Proceedings of IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence).
[14] V.Geetha, P.V.Kallapur, S. Tellajeera. (2012). Clustering in Wireless Sensor Networks: Performance Comparison of LEACH & LEACH-C Protocols Using NS2. Procedia Technology. 4, 163-170.
[15] A. Ghaheri, S. Shoar, M. Naderanand S. S.Hoseini. (2015). The Applications of Genetic Algorithms in Medicine. Oman Med J. 30(6). 406-416.
[16] J.Barbancho, C. León, F. Javier Molina & A.Barbancho. (2007). A new QoS routing algorithm based on self-organizing maps for wireless sensor networks. Telecommunication Systems. 36, 73-83.
[17] S.Cai, B.Gallina, D.Nyström, C.Seceleanu& A. Larsson. (2018). Tool-supported design of data aggregation processes in cloud monitoring systems. Journal of Ambient Intelligence and Humanized Computing. 10, 2519–2535.
[18] Kumar H, Singh PK (2018) Comparison and analysis on artificial intelligence based data aggregation techniques in wireless sensornetworks. ProcediaComputSci 132:498–506.
[19] V. T. Pham, T. N. Nguyen, B. H. Liu, M. T. Thai, B.Dumba and T. Lin. (2022). Minimizing Latency for Data Aggregation in Wireless Sensor Networks: An Algorithm Approach. ACM Transactions on Sensor Networks. 18(3), 1-21.
[20] P. William, A.Badholia, V.Verma, A. Sharma & A.Verma. (2022).Analysis of Data Aggregation and Clustering Protocol in Wireless Sensor Networks Using Machine Learning. Evolutionary Computing and Mobile Sustainable Networks. 925-939.

