# Intelligent Data Aggregation Framework for Resource Constrained Remote Internet of Things Applications

Abhijith H V[1]

Department of Information Science and Engineering
Sai Vidya Institute of Technology, Affiliated to Visvesvarya
Technological University, Belagavi, Bangalore, India

Dr. H S Ramesh Babu[2]

Department of Computer Science and Engineering
Sai Vidya Institute of Technology, Affiliated to Visvesvarya
Technological University, Belagavi, Bangalore, India

*Abstract*—**Internet of Things (IoT) is a technology that can connect everything to the Internet. IoT can be used in a wide range of applications which includes remote applications like Underwater networks. Remote applications involve the deployment of several low-power, low-cost interconnected sensor nodes in the specific region. With a massive amount of devices connected to the IoT and the considerable amount of data associated with it, there remain concerns about data management. Also, the amount of data generated in an extensive IoT-based remote sensing network is usually enormous for the servers to process, and many times data generated are redundant. Hence there is a need for designing a framework that addresses both aggregations of data and security-related issues at various aggregation points. In this paper, we are proposing an intelligent data aggregation mechanism for IoT-based remote sensing networks. This method avoids redundant data transmission by adapting spatial aggregation techniques. The proposed method was tested through simulations, and the results prove the efficiency of the proposed work.**

*Keywords—Wireless sensor networks; Internet of Things; intelligent boundary determination; sensor nodes; data aggregation*

## I. Introduction

Internet of Things refers to connecting various things to the internet. IoT can be used in a wide range of applications. IoT applications include remote applications also. Remote IoT applications include military applications such as surveillance, battlefield monitoring, underwater applications such as oil spill detection, analysis of aquatic animals lifecycle, smart farming, forest applications such as fire monitoring, etc. Remote IoT applications use resource-constrained networks. It includes deploying several low-cost and low-power sensor nodes that can sense, process, and communicate the data. Each sensor node has a limited transmission range, limited energy, limited processing capabilities, and limited memory.

As sensor nodes are resource-constrained devices, several sensor nodes will be deployed in a smaller region. This leads to the generation of redundant data. Transmitting redundant data will leads to the wastage of energy and other resources. Redundant data transmission can be controlled by data aggregation techniques. There are different data aggregation techniques such as In-network data aggregation, Tree-based data aggregation, cluster-based data aggregation, Grid-based data aggregation, and hybrid data aggregation. There can be single-level or multiple levels of data aggregation. Data aggregation can be performed in homogeneous networks or in heterogeneous networks. Homogeneous networks include nodes with similar configurations and capabilities in the entire network. Heterogeneous networks include nodes with different configurations and capabilities organized in the multiple levels of hierarchy [15][17].

Data aggregation involves just combining the data from multiple sources into a single packet and forwarding it to the next node or destination. Traditional approaches always aggregate the data based on the count, average, sum, etc. The reduction of redundant data is not that significant through normal data aggregation techniques. Hence there is a need for adding some intelligence to nodes to decide whether to transmit the data or not [16]. In the constrained remote IoT applications where we have limited resources, communication consumes more energy than compared to processing. In this paper, we are proposing an intelligent data aggregation technique, which aggregates the data as well as provides intelligence to nodes to detect whether to send the data or not. Also, our present work adds features to detect the boundary of certain event occurrences in event-driven networks.

The paper is organized as follows: Related work is given in Section 2, Section 3 describes proposed work, Section 4 describes the simulation and results, the conclusion is provided in Section5.

## II. Related Work

There are different traditional data aggregation mechanisms; they are flat-based [1][2][3] and hierarchical-based techniques.

In a flat network, the base station transmits a query message requesting data from the sensor nodes within the network. The nodes which have data relevant to the query sent will respond with the requested data. In this method, the base station performs excessive communications and computations. Because of this, if the base station fails, then the network connection will be lost with the outer world.

Under hierarchical data aggregation, many approaches have been proposed for energy efficiency and scalability. There exist four different types of hierarchical data aggregation they are the cluster-based data aggregation [4][5], tree-based data aggregation [9], grid-based data aggregation [8]., and chain-based data aggregation [6][7].

These traditional data aggregation schemes will just combine the data from multiple sources and forward it to the

next node. In the majority of applications, normal data aggregation is not sufficient. We need to add some intelligence to the data aggregation and data transmission process within the limits of constrained resources.

J. Chen, S. Kher, and A. Somani, et al. proposed a scheme in [9], which involves Majority Voting. This approach is a data outlier detection method based on spatial correlation. Here if a reading of the local sensor node is different from the majority of its neighboring nodes, then it will be classified as abnormal.

Y. Sun, H. Luo, and S. K. Das et al. Proposed a scheme in [10]. Here according to the trustworthiness ranked by comparison with historical data and neighbor data, Weight will be assigned to every sensor data. Then weighted mean value will be calculated at the aggregator. This will be considered as the aggregated data.

S. Din, A. Ahmad, et al. Proposed a scheme in [11], where the nodes closer to the sink node performs direct communication and remains unclustered; the nodes which are one-hop away from the sink node perform multi-hop communication and remains clustered.

Bo Yin et al. in [12] specified a Tree-based scheme that involves the construction of an aggregation tree for complex queries. This ensures minimum communication cost.

Xiong Li et al. proposed a scheme in [13] where there will be three participants, they are edge server, terminal device, and public cloud center. The data from the terminal devices is encrypted and communicated to the edge server, then the edge server performs data aggregation of the data from terminal devices and forwards the aggregated data to the public cloud center. The aggregated plaintext data can be recovered by public cloud center through its private key.

S. Kumar and V. K. Chaurasiya et al. in [14] proposes a scheme, where data mining techniques are used to generate more accurate, consistent, and useful information than that generated by any individual sensor node.

## III. PROPOSED WORK

### A. Assumptions and Architectural Setup

Following are the assumptions made in deployment of nodes in the WSN test bench.

*1) Grid based deployment of nodes:* nodes are deployed on a crisscross grid, to specifically determine location of each individual node. This assumption is both safe and valid as, there is a commercially available underwater robot capable of deploying nodes at specific coordinates of latitude and longitude. Such a deployment is also helpful in making intelligent inferences based on which nodes are communicating sensed data.

*2) Number of nodes per grid is fixed*: this assumption helps us compute statistically significant results after data collection is done and analysis is to be performed.

*3) Node density is uniform:* non-uniform node deployment does not allow implementation of intelligence.

*4) Each node in a grid represents a single unit of area*: data collected is representative of a fixed area under surveillance.

*5) Unit squares (Grids) form a Level -1 cell*: this logical rule imposed on grid helps in aggregation of data across levels of nodal deployment.

*6) Every level-1 cell has a level-1 cell aggregator node*: this is an architectural requirement of WSNs. aggregator nodes act as data aggregator for sensed data before being forwarded to upper layers.

*7) Four level-1 cells combine to form a level 2-cell*: multiple layers in the architecture will always help in implementation of an additional second layer of intelligence and additional aggregation.

*8) One of the unit cell's cell-sink also works as a sink for four adjacent level-2 cells*: this is strictly based on energy levels of nodes, suitable cell-sink is elected as the sink of adjacent cells based solely on this criteria.

*9) All level-2 cells combine to form the region under observation*: this is the highest level of aggregation implemented under our framework.

### B. Node Deployment Phase and Network Establishment

*1)* Four different levels of nodes are deployed.

*a) Ground Level Sensor Nodes*: They are low power sensor nodes with capability to sense and forward the data to next level.

*b) Grid Head*: They are high power nodes. They can aggregate the data and also sense the information.

*c) Level -1 Aggregator:* They act like second level grid head. They gather the data from multiple grids and transfer it to Level-2 Aggregator after processing.

*d) Level-2 Aggregator:* They are at the top level of hierarchy, they can provide minor conclusions about the sensed information.

*2)* These nodes are deployed at the required locations.

*3)* The area of interest is divided into equal sized grids. Ground level nodes and Grid heads are deployed at each grid. High power node will be the grid head of that particular grid.

*4)* Each High power node broadcast a beacon message to the ground level sensor node to indicate its presence in the grid.

*5)* Each ground level nodes respond back to their grid heads by sending the beacon response. Same Procedure is followed between grid heads and anchor nodes, Anchor nodes and surface buoyant nodes.
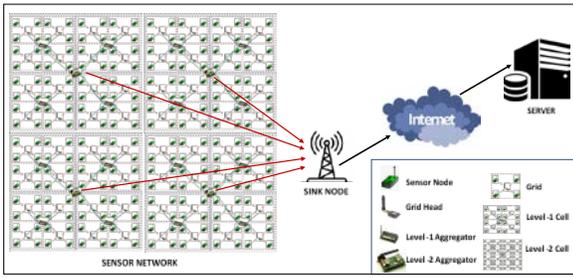
Fig. 1.    Network Architecture.

Fig. 1 shows the proposed network architecture. The network is divided into smaller grid. Grids are combined to form level-1 Cell. Level-1 Cells are combined to form level -2 Cells. There will be sink node at the border of area of interest. Sink node forwards the data to server through internet connection.

### C. Data Aggregation

Tier- 1Local aggregation (level 1):

Spatial aggregation: this primary level of aggregation is based on the percentage of area generating similar readings. Each grid has a designated grid head. The grid head communicates reading only if number of nodes reporting similar readings are greater than threshold percentage.

Spatial Aggregation is adapted in Grid as well as level-1 Cell by Grid head and Level-1 Aggregator.

Tier-2 Aggregation (level 2)

TABLE I.        SYMBOLS USED IN THE ALGORITHM

| Symbol | Description |
|--------|-------------|
| S | Ground Level Sensor Node |
| G | Grid Head |
| L1 | Level 1 Aggregator |
| L2 | Level 2 Aggregator |
| SI | Sink Node |
| Ts | Threshold Percentage used by G for Level 1 Aggregation |
| Tg | Threshold Percentage used by L1 for Level 1 Aggregation |
| R1 | Threshold Radius 1 for level 2 Aggregation |
| R2 | Threshold Radius 2 for level 2 Aggregation |
| $t_i$ | Time interval, i=1,2,3…. |
| Ns | No. of Sensor Nodes which has sent similar readings |
| Ng | No. of Grid Heads which has sent similar readings |
| Ps | Packet sent by sensor node |
| Pg | Packet sent by Grid Head |
| PL1 | Packet sent by Level-1 Aggregator |
| PL2 | Packet sent by Level-2 Aggregator |
| D | Distance between L2 and L1 |
| Dist | List of D computed in 1 time interval |

Boundary value aggregation: region under observation has a Level -2 Aggregators that acts as a data aggregator. Radial survey is made periodically at three lengths of radii- minimum, nominal and maximum. Minimum is close to the Level -2 Aggregators, maximum is distance from Level -2 Aggregators to boundary of region under observation, and nominal is an intermediate distance.

Algorithm 1 shows how sensor senses and sends the data. Algorithm 2 and 3 depicts the level-1 aggregation techniques (Spatial aggregation). Level -2 aggregation is given in Algorithm 4. Table I describes the symbol used in the algorithm.

---

**Algorithm 1:** Sensing at Ground Level

1.  For each ti i=1,2,3….

2.    S senses the data

3.    Prepare the packet Ps

4.    Send Ps to its G

5. End

---

**Algorithm 2:** Level-1 Aggregation at Grid Level

1. For each ti, i=1,2,3…..

2.    G: Compare the Ps received by all S under G and compute Ns

3.    If Ns >= Ts

4.        Aggregate the data

5.        Prepare the packet Pg

6.        Send Pg to L1

7.    Else

8.        Discard the data

9. End

---

**Algorithm 3:** Level-1 Aggregation at Level-1 Cell Level

1. For each ti, i=1,2,3…..

2.    L1: Compare the Pg received by all G under L1 and compute Ng

3.    If Ng >= Tg

4.        Aggregate the data

5.        Prepare the packet PL1

6.        Send PL1 to L2

7.    Else

8.        Discard the data

9. End

---

| **Algorithm 4:** Level 2 Aggregation at Level-2 Cell level |
| --- |

1. Consider coordinates of L2: (XL2 , YL2) as centre point.

2. Let (XL1, YL2) be the coordinates of L1

3. For each L1 which has sent data to L2

4.     Find distance between the L1 and L2

5.     D = Sqrt( (XL2 – XL1 )2 – (YL2 – YL1 )2 )

6.     Add D to list Dist

7. End

8. Find largest D in Dist

9. If largest Di is within radius R1

10.     Radial_survey = minimum: Event spread in shorter region

11.     Prepare PL2 with a field Radial_survey

12.     Send PL2 to Sink

13. Else if Largest Di is between radius R1 and R2

14.     Radial_survey = nominal: Event spread in Moderate region

15.     Prepare PL2 with a field Radial_survey

16.     Send PL2 to Sink

17. Else if Largest Di is greater than R2

18.     Radial_survey = maximum: Event spread in longer region

19.     Prepare PL2 with a field Radial_survey

20.     Send PL2 to Sink

## IV. SIMULATION AND RESULTS

In the proposed research work, we have used MATLAB as the simulation tool to check the efficiency of our proposed work. Initially, we considered the area of the application region with dimensions 800m (length) x 800m (breadth). Then the area under consideration is divided into equal-sized grids of size 100m each at ground level. Table II shows the various parameters used in the simulation. Fig. 2 depicts the topology created through simulation. Here blue color nodes represent sensor nodes deployed at ground level. Red color nodes represent the grid head; the magenta color nodes indicate nodes that are Level-1 Aggregators, Black color represents Level-2 Aggregators, and Green color indicates sink node.

Fig. 3 shows the amount of time required to detect the boundary of the event spread through Level-2 aggregation. In certain remote IoT applications detecting the spreading rate of a certain event is very important to take immediate measures. For Ex: the spread of fire in the forest, oil leakage in the ocean, etc.

Fig. 4 helps us to understand that the number of redundant transmissions in various techniques compared with our work. It is evident that redundant data transmission is significantly dropped compared to other traditional approaches. Our method

requires a lesser number of packet transmissions compared to without aggregation deployment, traditional in-network architecture, and cluster-based approach. It is clear that the proposed work reduces redundant data transmission significantly.

TABLE II. SIMULATION PARAMETERS

| Parameter Type | Parameter Value |
| --- | --- |
| Area of application region | 800 X 800 $m^2$ |
| Grid range | 100 m |
| Number of Ground level node in each grid | 4 |
| Number of High power nodes in each grids | 2 (1Grid Head) |
| Number of Level-1 Aggregators in each group of Level-1 Cell | 1 |
| Number of Level-2 Aggregators in each group of Level-2 Cell | 1 |
| Simulation Time | 150sec |
| Payload length | 512 Bytes |
| **Parameters of Ground level sensor node** | |
| Initial Energy | 4J |
| Transmission range | 20m |
| Data rate | 4kbps |
| **Parameters of High power node/Grid Head** | |
| Initial Energy | 7J |
| Transmission range | 25m |
| Data rate | 6kbps |
| **Parameters of Level-1 Aggregator** | |
| Initial Energy | 10J |
| Transmission range | 30m |
| Data rate | 8kbps |
| **Parameters of Level-2 Aggregator** | |
| Initial Energy | 14J |
| Transmission range | 35m |
| Data rate | 12kbps |
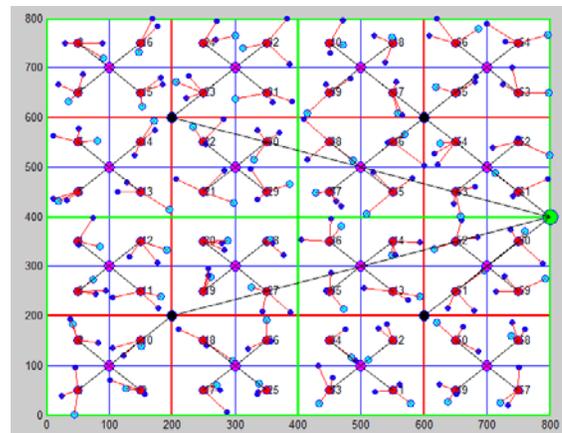


Fig. 2. Topology Setup.

Fig. 5 shows the number of the packet transmitted to sink. In this result also it is clear that our proposed method is better compared to others. The proposed method involves very little data transmission.

Fig. 6 shows the comparison of the proposed method with other traditional methods in terms of energy consumption. Result clearly shows that the proposed method consumes lesser energy compared to other techniques.
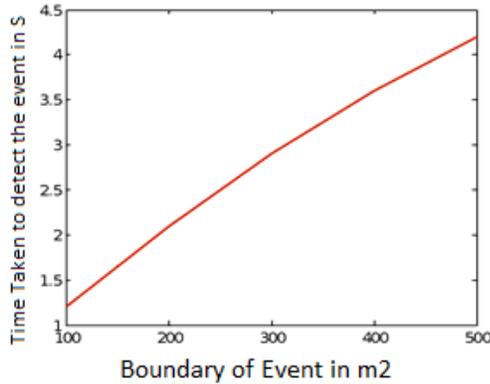


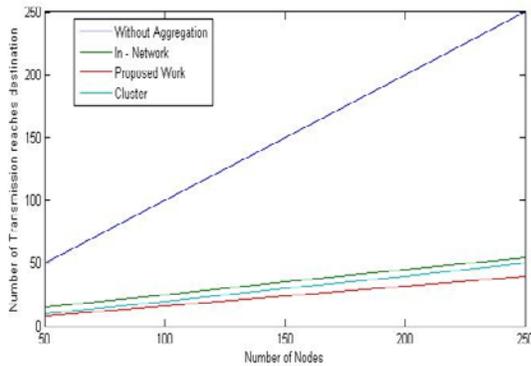Fig. 3.   Time Required for Detecting the Boundary of Event.



Fig. 4.   Number of Redundant Transmission in Various Approaches.
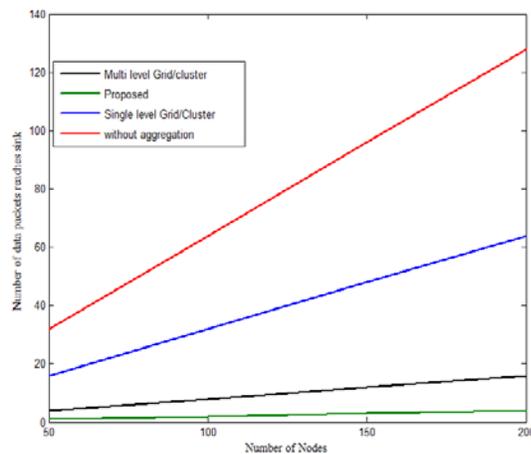


Fig. 5.   Number of the Data Packets Transmitted to the Sink.
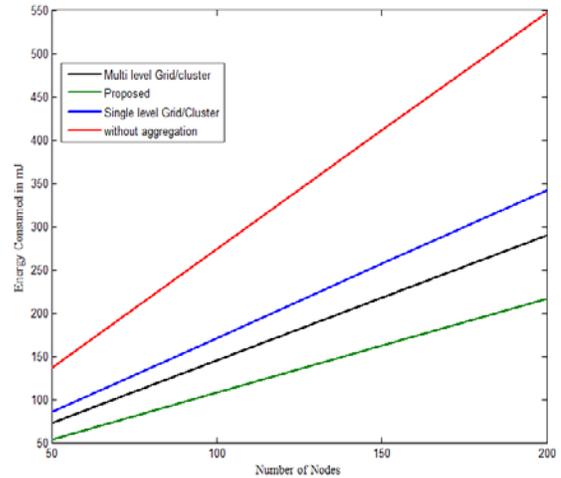


Fig. 6.   Energy Consumed Per Round in Various Approaches.

## V.   CONCLUSIONS

Controlling redundant data transmission is a significant challenge in Constrained IoT Remote applications.  In this proposed research work, Data Aggregation is addressed through spatial and boundary value aggregation, which is the novel scheme. The aggregation technique adapted reduces the number of redundant transmissions significantly. The efficiency of the proposed approach is compared with existing methods through simulation.

REFERENCES

[1]   J.Kulik.W.R.Heinzlman & H.Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks", WirelessNetwork, Vol 8, Mar-2002, pp.169-185.

[2]   C.Intanagonwiwat,R.Govindan and D.Estrin, "Directed Diffusion: A Scalable and robust communication paradigm for sensor networks", 6th annual international conference on mobile-computing and networking, Aug-2000.

[3]   B.Krishnamachari and J.Heidemann, "Application specific modeling of information routing in wireless sensor networks", IEEE international performance, computing, communications conference, Vol-23, pp.717-722, 2004.

[4]   W.R Heinzelman, "Application-specific protocol architectures for wireless networks", PhD-Thesis, MIT, June-2000.

[5]   O.Younis and S.Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks" IEEE transactions on mobile computing, Vol-3, Dec-2004,pp.366-79.

[6]   Lindsey.S, Raghavendra.C, "PEGASIS Power-efficient Gathering in sensor Information system", IEEE Aerospace Conference 2002, pp-1125- 1130.

[7]   Kuong-Ho-Chen, Jyh-Ming-Huang, Chieh-Chuan-Hsiao "CHIRON: An Energy-Efficient Chain-Based Hierarchical Routing Protocol in Wireless Sensor Networks", IEEE. 2009.

[8]   Liyang-Yu, Neng-Wang, Wei-Zhang, Chunlei-Zheng, "GROUP: a Gridclustering Routing Protocol for Wireless Sensor Networks", 2006 IEEE International conference on wireless-communication networking & mobile-computing, china 2006.

[9]   J. Chen, S. Kher, and A. Somani, "Distributed Fault Detection of Wireless Sensor Networks" ACM Workshop on Dependability Issues in Wireless Ad-hoc Sensor Netw. 2017.

[10]  Y. Sun, H. Luo, and S. K. Das, "A Trust-based Framework for FaultTolerant Data Aggregation in Wireless Multimedia Sensor Networks," IEEE Trans. Depend. Sec. Comput., 2017.

[11] S. Din, A. Ahmad, A. Paul, M. M. U. Rathore, and J. Gwanggil, ``A cluster based data fusion technique to analyze big data in wireless multi-sensor system,'' IEEE Access, vol. 5, pp. 50690 5083, 2017.

[12] Bo Yin et. all. "Communication-Efficient Data Aggregation Tree Construction for Complex Queries in IoT Applications", IEEE INTERNET OF THINGS JOURNAL, 2018 .

[13] Xiong Li et. all. "Privacy Preserving Data Aggregation Scheme for Mobile Edge Computing Assisted IoT Applications", IEEE INTERNET OF THINGS JOURNAL, 2018.

[14] S. Kumar and V. K. Chaurasiya, "A Strategy for Elimination of Data Redundancy in Internet of Things (IoT) Based Wireless Sensor Network (WSN)," in IEEE Systems Journal 2018.

[15] Abhijith H V, & Sindhu M P. (2015). Energy efficient multilevel hierarchical data aggregation mechanism for wireless sensor networks.

2015 IEEE International Advance Computing Conference (IACC). doi:10.1109/iadcc.2015.7154688.

[16] H.V., Abhijith and Raj, S. Deepak and Babu, H. S. Ramesh, Intelligent Boundary Determination of Oil Spill Detection Using IOT (April 21, 2018). Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), 2018 held at Malaviya National Institute of Technology, Jaipur (India) on March 26-27, 2018, Available at SSRN: https://ssrn.com/abstract=3167315 or http://dx.doi.org/10.2139/ssrn.3167315.

[17] F. Al-Doghman, Z. Chaczko and J. Jiang, "A Review of Aggregation Algorithms for the Internet of Things," 2017 25th International Conference on Systems Engineering (ICSEng), Las Vegas, NV, 2017, pp. 480-487. doi: 10.1109/ICSEng.2017.43.