IMPROVING DATA AGGREGATION EFFICIENCY USING MULTI-LAYER APPROACH IN IoT

A Vanathi

Research Scholar, Acharya Nagarjuna University & Associate Professor,
Aditya Engineering College Department of CSE, Andhra Pradesh, India
vanathi_a@rediffmail.com

Dr. N. Nagamalleswara Rao

Professor, Department of IT, RVR&JC College of Engineering and Technology,
Guntur, Andhra Pradesh, India
nnmr3654@gmail.com

Abstract - Recently, tremendous growth and interest in the deployment of tiny sensors in the Internet of Things (IoT) for smart applications improves human lives. With the increasing need for energy-efficient mechanisms in IoT communication, the data aggregation technique for reducing data transmissions is considered a significant research problem. The basic idea in most of the aggregation mechanisms is to build the clustering or aggregation tree in an application layer over IoT, resulting in high complexity. To solve such a problem, the proposed MLDA designs an energy-aware aggregation layer that focuses on utilizing the network layer factors in data aggregation by providing transparency of accessing the topology structure from the network layer. Moreover, the proposed work also focuses on the design of the load-balanced topology structure in the network layer for efficient routing and also takes support from such network structure for energy-efficient data aggregation. The proposed Multi-Layer based Data Aggregation approach (MLDA) avoids the hotspot problem and inefficient data aggregation. The MLDA achieves such goals by improving the network layer protocol, RPL activities and designing the aggregation layer to eliminate redundant transmissions. By using an energy-efficient network structure, the impact of redundant data transmissions on network resources and data aggregation efficiency are eliminated. To support SUM, AVG, MAX, and MIN aggregation functions without redundant data transmissions, the Double Hash Bloom Filter (DHBF), observation scheme, and merge sort are used in the developed aggregation layer. Thus, the proposed MLDA improves the data aggregation efficiency in terms of both energy and accuracy.

Keywords: IoT; data Aggregation; RPL; redundant Data Transmission; Multi-layer approach

1. Introduction

There is a growing interest in deploying the Internet of Things (IoT) in various smart applications. With the deployment of a huge number of smart devices and their applications, gathering and analyzing the data is becoming one of the main challenges [Dehkordi, *et al*(2020)]. As the tiny sensor devices are powered by batteries, energy-efficient operations are critical in IoT applications [Karamitsios, and Orphanoudakis, (2017)]. Large scale and dense IoT network creates a tradeoff between the waste of energy due to data redundancy and maintaining the data aggregation efficiency. The data aggregation scheme has to lessen the redundant readings of sensors in the surrounding area [Chandnani, and Khairnar, (2020)][Wala, et al (2020)].

The data aggregation efficiency is largely dependent on the network topology. Thus, the support from lower layers, especially from the network layer is essential to improve the data aggregation efficiency without wasting the energy of sensor devices. It necessitates the importance of a multi-layer data aggregation approach [Guimaraes, *et.al* (2019)]. However, there is a lack of providing load balanced and an energy-efficient topology structure in the network layer, thereby improving the data aggregation efficiency. Mapping the network topology structure to the aggregation activities is crucial. However, it increases the network cost when connecting the non-adjacent layers, such as the network and application layer via messages. Thus, the proposed work plans to integrate the aggregation layer upper to the network layer. Before transferring the received data to the gateway, It is desirable for a parent node in the aggregation layer to eliminate redundancies in the received data from the neighboring nodes and to aggregate the data effectively. Thus, the proposed MLDA plans to take a network layer support in creating the load-balanced DODAG structure in RPL and eliminates the redundant transmissions significantly to improve the data aggregation efficiency without wasting the network resources.

The contributions of the proposed MLDA are as follows.

- To design a data aggregation layer upper to the network layer for avoiding the hotspot problem and data aggregation inefficiency due to data redundancy.
- Design a multi-layer approach that supports the load-balanced topology structure in the network layer for efficient data aggregation in a new layer.
- To mitigate the impact of data redundancy on data aggregation efficiency using the Double Hash Bloom Filtering scheme.
- To support various aggregation functions, such as SUM, AVG, MAX, and MIN in a data aggregation layer using observation scheme and Merge Sort.

1.1 *Paper organization*: Section 2 surveys the existing data aggregation schemes. Section 3 presents the proposed multilayer methodology. The performance of the proposed methodology using simulation model is presented in section 4. Finally, section 5 concludes the multilayer data aggregation paper.

2. Related Works

The conventional data aggregation techniques have been divided into three types, cluster head-based data aggregation, tree-based data aggregation, and centralized data aggregation [Pourghebleh, and Navimipour, (2017)][Guidi, and Ricci,(2019)]. Energy efficiency should be the main consideration for providing a suitable data aggregation technique to resource-constrained IoT networks [Fitzgerald, *et al.*(2018)]. The conventional data aggregation works based on the three techniques above are described as follows.

2.1 Cluster Head-Based Data Aggregation Schemes

The IoT sensor devices build clusters and a cluster head selected by single or multiple factors, and the cluster head is responsible for aggregating the data from the cluster members and forwards the aggregated data to the gateway in the IoT network. However, there is no direct communication in-between the IoT devices and gateway. They make communication via the cluster head. The priority-based channel access and aggregation of the data scheme in [Bhandari, *et al.*(2017)] facilitate the cluster head to reduce the channel access latency. The preemptive queuing model in [Bhandari, et al.(2017)] supports an efficient data aggregation by separating the high and low priority queue before forwarding the aggregated data packets to the gateway. The priority-based two aggregation schemes are suggested in [Ghate, and Vijayakumar, (2018)]. In the first scheme, the machine learning techniques or filtering techniques are used to extract the values when the class label for input data is already known. It assists in monitoring the threshold values and decides the priority levels based on ranking and weights. The cluster head forwards the data, which depends on the priority. The second scheme handles a case of output class labels not known.

A Lightweight Privacy-Preserving Data Aggregation (LPDA) scheme is suggested [Lu, *et al.*(2017)]. The proposed algorithm consists of four parts. They are network initialization, report generation, data aggregation, and report reading and analytics. The Cross-Layer Commit Protocol (CLCP) in [Alkhamisi, *et al.*(2016)] creates a separate cluster topology for performing data aggregation in the IoT application layer. Moreover, the cluster topology creation is depending on the factors of energy and distance. It fails to consider the load-related metrics in network structure creation, resulting in poor data aggregation efficiency. The data packet loss due to network load creates a negative impact on data aggregation. A Recursive Principal Component Analysis (R-PCA)-based data aggregation algorithm is suggested [Amarlingam, *et al.*(2018)]. The R-PCA configuration is depending on cluster-based data aggregation. The data collected from the spatially correlated IoT device data are aggregated by considering the key components. A novel Lightweight Compressed Data Aggregation (LWCDA) algorithm is suggested in [Yu, *et al.*(2017)] for improving the network lifetime. The LWCDA algorithm divides the entire network into the non-overlapping groups for data aggregation. The non-overlapping cluster provides energy efficiency and low complexity to resource-constrained sensor devices. The LWCDA exploits the highly sparse matrix for mitigating the complexity in the network. In a cluster head-based data aggregation scheme, each cluster head is responsible for sending the aggregated data to the gateway. Thus, the probability of packet loss is minimized. However, the selection process for cluster head and lightweight scheme for data redundancy elimination are important for cluster-based data aggregation schemes.

2.2 Tree and Centralized Data Aggregation

Tree-based data aggregation scheme for in IoT network is described in [Preetha, *et al.*(2018)] exploits a reinforcement learning-based fuzzy interference system. Based on the node density in the network, it builds the tree-based cluster. It exploits the neighborhood overlap as the main metric to build the tree structure for data aggregation. The centralized data aggregation schemes send the required information to intermediate nodes and then to the header node. All the intermediate nodes transmit the data packets to the gateway through a single header node. In [Ko, *et al.*(2019)], an energy-efficient sleep scheduling method is proposed along with the data aggregation. It exploits the Markov decision process (MDP) for identifying the best sleep duration of IoT devices as well as data aggregation duration in the IoT gateway device. Some of the works provide privacy to

the aggregated data in IoT [Li, *et al*(2018)]. The Scalable Energy Efficient Scheme (SEES) in [Abdul-Qawy, and Srinivasulu,(2019)] attempts to provide efficient data aggregation in the IoT network. A multi-stage weighted linear combination method is executed based on the election heuristic (MSWE). A Minimum Cost Cross-layer Transmission (MCCT) method is used for data distribution in the whole network. However, a main drawback of the system is the low accessibility of the central node in the network, and the data may be lost if a central node is disconnected.

3.   Overview of the Proposed methodology

Data aggregation in IoT looks at multiple data packets and their contents and combines them into one. The main motivation behind data aggregation is to reduce the number of transmissions by exploiting the redundancy in the sensor readings and providing energy-efficient communication. It necessitates the construction and utilization of efficient data structures to promote in-network data aggregation in IoT. Specifically, creating an energy-efficient data gathering tree, where and how to do the aggregation in this tree are the important questions to be answered. The proposed MLDA solves such problems by proposing the aggregation layer and taking support from the network layer.

Load balanced DODAG Construction at Network Layer: The complexity of connecting the non-adjacent layers for performing multi-layer data aggregation necessitates the development of an aggregation layer adjacent to the network layer. The functionalities of the network layer have an impact on the data aggregation efficiency in the aggregation layer. Thus, the proposed MLDA plans to improve the network layer protocol also. To avoid the hot spot problem and inefficient data aggregation, the network layer protocol, RPL activities are improved. Selecting and applying proper values to the trickle algorithm and multiple parameter based objective function improves the routing and aggregation efficiency. The necessary factors for data aggregation from the network layer are accessed in the adjacent layer of aggregation.

Aggregation Layer: The aggregation layer is built for performing the data aggregation in an energy-efficient manner. For data aggregation, the selected nodes are identified from the adjacent layers from the network layer. The redundant readings among spatially correlated sensors waste communication and storage resources as well as delaying aggregation decision-making. Thus, filtering duplicate data from IoT sensor readings is an important and challenging problem in the aggregation layer. Moreover, the proposed aggregation layer can accommodate different aggregation functions such as SUM, AVG, MAX, and MIN based on the IoT application requirements.

Reading Redundancy Removal using Double Hash Bloom Filter: To solve such an issue, the selected nodes in the aggregation layer executes the filtering approach. The hash-based filtering approach avoids redundancy as well as delayed decision making. Existing Bloom Filter-based approaches for removing redundant readings are complex since they apply multiple hash functions. To avoid such issues, a double hash-based bloom filter is used. For each unique reading, the bloom filter creates a new location, and for the same reading, it maps to the same location. It reduces the redundant readings received from the IoT sensors.

Observe Scheme for SUM and AVG Functions: The child nodes initiate the observe scheme by appending the count of redundant readings in the packet header. The packets are forwarded to the aggregator or parent node. Thus, it supports the SUM and AVG aggregation functions at parent nodes without sending the redundant readings.

MAX-MIN Functions using Merge Sorting: Sorting algorithm is essential for performing both MAX and MIN aggregation functions at parent nodes in the aggregation layer. The double hash-based redundant reading removal reduces the execution delay of MAX-MIN Functions. Indeed, the merge sorting divides the readings into multiple lists to perform sorting and to improve the aggregation efficiency in both time and accuracy. Finally, the aggregated values are accessed in the application layer.
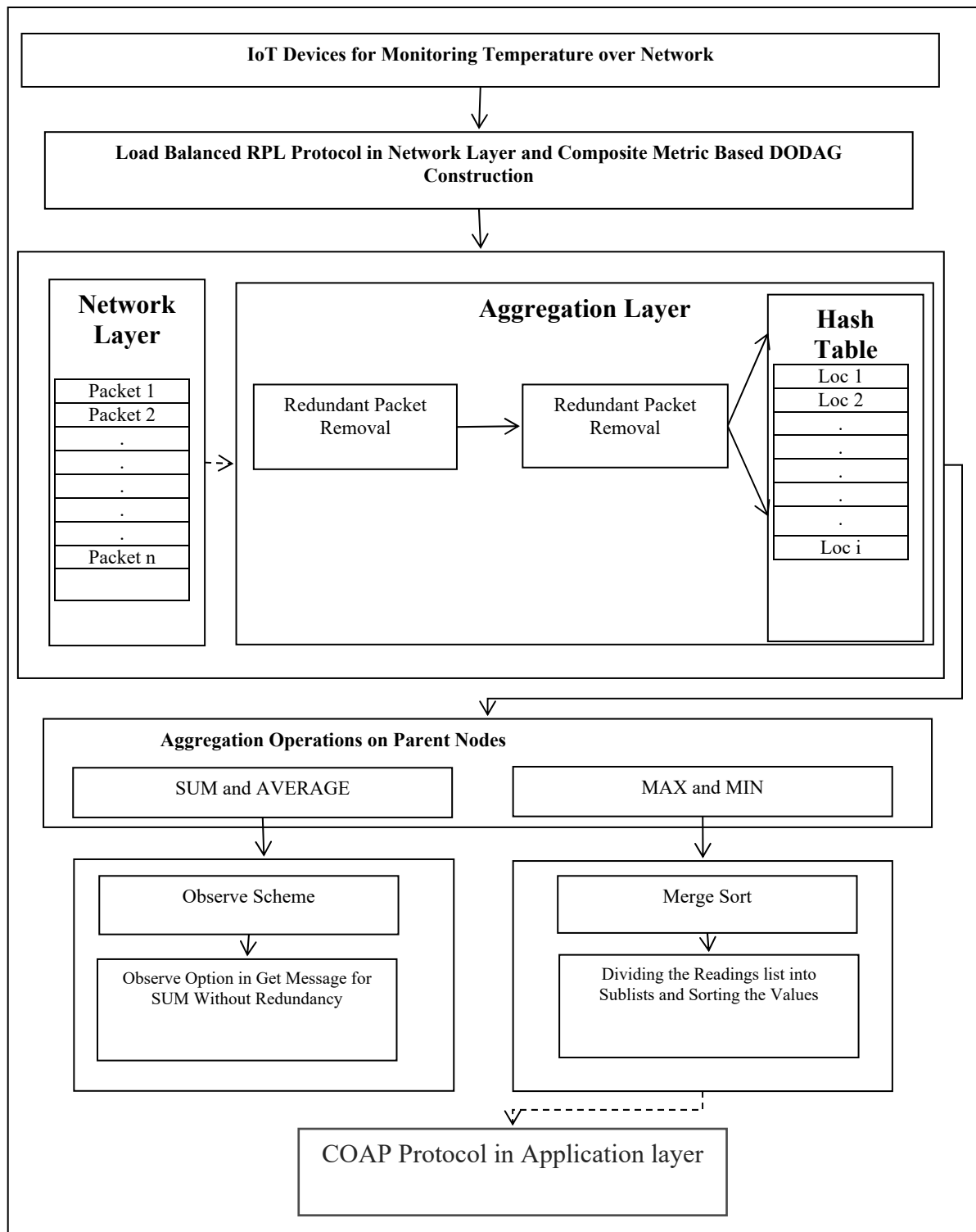
Fig 1: Block Diagram of the Proposed Methodology

3.1 *System Model*

The IoT network G consists of N sensor nodes (sn_1; sn_2; . . ., sn_N). Such a network measures temperature values (T_{sni}) of the particular surrounding area. Every sensor node sn_i is responsible for sensing the temperature at a fixed period. These readings are temperature readings, T. There is a possibility for measuring similar T values among the sensors in the particular surrounding area. To reduce the redundant T_{sni} transmissions, the proposed MLDA designs a multi-layer approach with the existing network and new aggregation layer.

Network Layer: As per the network layer protocol, the nodes $\in$ N select a node with minimum cost value as a parent, whereas cost value measurement depends on multiple routing factors. In the network layer, a full topology constructed as per RPL from each sensor to the root node is called DODAG.

Aggregation Layer: The parent nodes Np in the DODAG structure are informed to the aggregation layer through packet header. The parent nodes execute the SUM, AVG, MAX, and MIN aggregation functions with network-layer support. Finally, the aggregation layer passes the aggregated readings to the application layer via the transport layer.

3.2 *Load balanced DODAG Construction at Network Layer*

The protocol functionalities of the lower layer, specially RPL in the network layer, have an impact on the aggregation efficiency and data communication. A high scale and dense IoT environment induce a heavy load on the sensor nodes nearer to the root node than others, as per RPL functionalities. It induces the hot spot problem and worsens the upper layer performance, especially data aggregation in the application layer. Thus, improving the efficiency of RPL in the network layer assists the IoT network to utilize the advantages of data aggregation technique entirely. Some of the ideas associated with hotspot elimination are as follows.

1.   Controlling the size of the DODAG structure with the number of candidate parents and without losing an energy-efficient DODAG structure.

2.   Trickle algorithm has to avoid an unnecessary DIO message traffic by avoiding unnecessary comparisons between the count of consistent message transmission t and redundancy counter c.

3.   Frequent parent switching in the network affects the DODAG structure, so considering the network stability is important.

Using the above factors, the Basic RPL is modified to avoid the hotspot issue around the root node and supporting an efficient data aggregation at the aggregation layer in the IoT environment. The selected parent nodes in the network layer act as aggregators in the aggregation layer for the readings collected from children's sensors. The nodes' status, either parent or child, is shared from network to aggregation layer to application layer through packet header.

3.2.1 Attaining benefits of Redundant Data without Redundant Transmission

Data redundancy ensures high reliability in data aggregation since bad quality data transmission tends to erroneous decision-making results. In IoT, the sensor nodes are deployed in a region to measure environmental factors. The sensor nodes sense the temperature in the surrounding area and forward it to the gateway. However, temperature sensors sometimes lead to data redundancy at the gateway. The redundant data improves data accuracy, whereas redundant data filtering helps in energy saving since most of the gateway resources get wasted in dealing with the redundant data during data aggregation. Based upon their outcome, it has been concluded that there is a tradeoff towards the use of redundancy to achieve reliable data while minimizing energy consumption. Thus, it is essential to build a data aggregation mechanism that achieves both data reliability and energy savings by utilizing redundancy in IoT. The proposed scheme introduces the double hash bloom filter for partial redundancy removal and observe scheme for attaining the benefits of redundant data without redundant transmission.

3.3 *SUM and AVG Aggregation Functions*

The aggregation measures such as AVG and SUM are basic in temperature measurement in an IoT environment. The transmission of redundant temperature reading for those basic measures is important. However, the redundant transmissions waste the network resources and delay the data aggregation process. Thus, it is important to eliminate redundant transmissions without affecting the AVG and SUM aggregation measures.

3.3.1 Aggregation Layer Observe Scheme

For reducing the redundant reading transmissions over time, the proposed MLDA work introduces the aggregation layer and its communication messages without increasing the complexity. The proposed scheme exploits the aggregation observe scheme for performing the data aggregation functions using the count of redundant readings but without redundant data transmissions. In aggregation GET message, GET/Agg, a count of generated particular readings over a time by a child node are attached in a single observation field, instead of transmitting the packets sequentially in response. In the network layer, a temperature sensor and its children may incur multiple packet transmissions for the same reading over a particular time. It is named redundant transmissions. The redundant transmissions are continued in the aggregation layer also.

The aggregation layer exploits GET request and response messages for getting values from the network layer. If the aggregation layer GET message includes a count for measuring the same temperature reading by the neighboring nodes over time, efficient data aggregation is achieved by limiting the number of redundant transmissions. In the proposed work, a single packet field attached in the GET request and response messages avoids multiple packet transmissions with the same reading for a particular time. The aggregation layer GET message is attached with a count for a temperature reading, which is not modified over time. The steps involved in the observation scheme are explained as follows.

Step 1: If a sender node measures the same temperature reading continuously for more than two times, it stores such temperature reading as a reference reading. Notably, for the first two temperature readings, the data packets are transmitted to the parent node from the child separately.

Step 2: After that, a child node observes the temperature for a particular time T. If it is the same as the reference temperature, the observation field is used to intimate the count of reference temperature measured at a particular child node.

Step 3: After the T period, a child node attaches a count for a particular temperature reading to respond to the GET message instead of sending more packets with the same temperature reading. If the count is too large, too many redundant packets are avoided over time, and it avoids wastage of resources.

Step 4: Moreover, it attaches the currently generated new temperature reading if it is varied from the reference temperature before or after the T period.

Step 5: In contrast, if the count is one, it sends the packet with the actual temperature reading to the receiver.

However, there is a possibility to delay or lose the GET message and to generate the same temperature reading at different child nodes, resulting in an inefficient data aggregation at the aggregation layer. Thus, the proposed MLDA executes the DHBF for removing redundancy at the parent or aggregator node after receiving the redundant transmissions.

3.4 *Reading Redundancy Removal using Double Hash Bloom Filter*

Every sensor in the network is installed with the Double Hash Bloom Filter (DHBF) for storing the unique sensor readings and for reducing the redundant sensor readings significantly. A DHBF does not store the readings themselves, and the main use of a DHBF is to test if reading is present. Thus, the DHBF based approach filter out the duplicate sensor readings and performs aggregation efficiently at parent nodes in the IoT environment. Most of the existing approaches decide an optimal number of hash functions with a single array for removing duplication. However, the number of hash functions improves the accuracy of the filter but increases the time and space complexity.

When the number of hash functions k increases, there is more possibility to determine a zero when examining the Bloom filter, and it reduces a false positive significantly. However, it slows down the IoT performance. In contrast, a small value of k also produces more ones in the Bloom filter, but it increases the chance of a false positive. The tradeoff between the number of hash functions and false-positive needs to yield. It is possible to minimize the number of hash functions less than optimal, increasing the number of arrays. Thus, the proposed approach uses double hash functions with two arrays.

3.4.1 Double Hash Bloom Filter and Redundant Data Removal

A DHBF data structure contains a bit array of size m with double hash functions (k=2). To add a sensor reading, the parent or the aggregator node in the aggregation layer fed it to each of the k hash functions to get k array positions. The proposed MLDA exploits the non-cryptographic hash functions, such as Murmur and FNV [Appleby, (2013)][Fowler, *et al.*(2011)]. Regardless of the type of data, the hash functions treat each sensor reading of a set as a string, and thus they are suitable for the temperature readings. Initially, the DHBF sets all bits in both the arrays to 0 and replaces zero with 1 when the k hash functions have a sensor reading. A sensor reading is verified, whether it is already hashed using the same set of hash functions and a bit of the corresponding index in either array1 or array2 are set to one. An element is considered a member of the particular array if all k indexes corresponding to the element are set to 1. The process of DHBF is given in the following algorithm and also depicted in figure 2.

To query for a sensor reading at parent nodes or test whether it is in the set, the DHBF feeds it to each of the k hash functions and obtain k array positions. If any of the bits at these indexes in both the arrays is 0, the sensor reading is not in the set (Step5). If those bits at the corresponding index are one in any one of the array, then all the bits in another array are set to 1, it is inserted in an array (Step 6 and 7). If all are 1 in both the arrays, then either the sensor reading is already in the set (Step 4). It is considered as duplicate, and for instance, the readings W is considered as duplicate since the bits of its corresponding position 1 and 5 in both the arrays are set as one, and such duplicate reading is represented in dotted lines in figure 2.
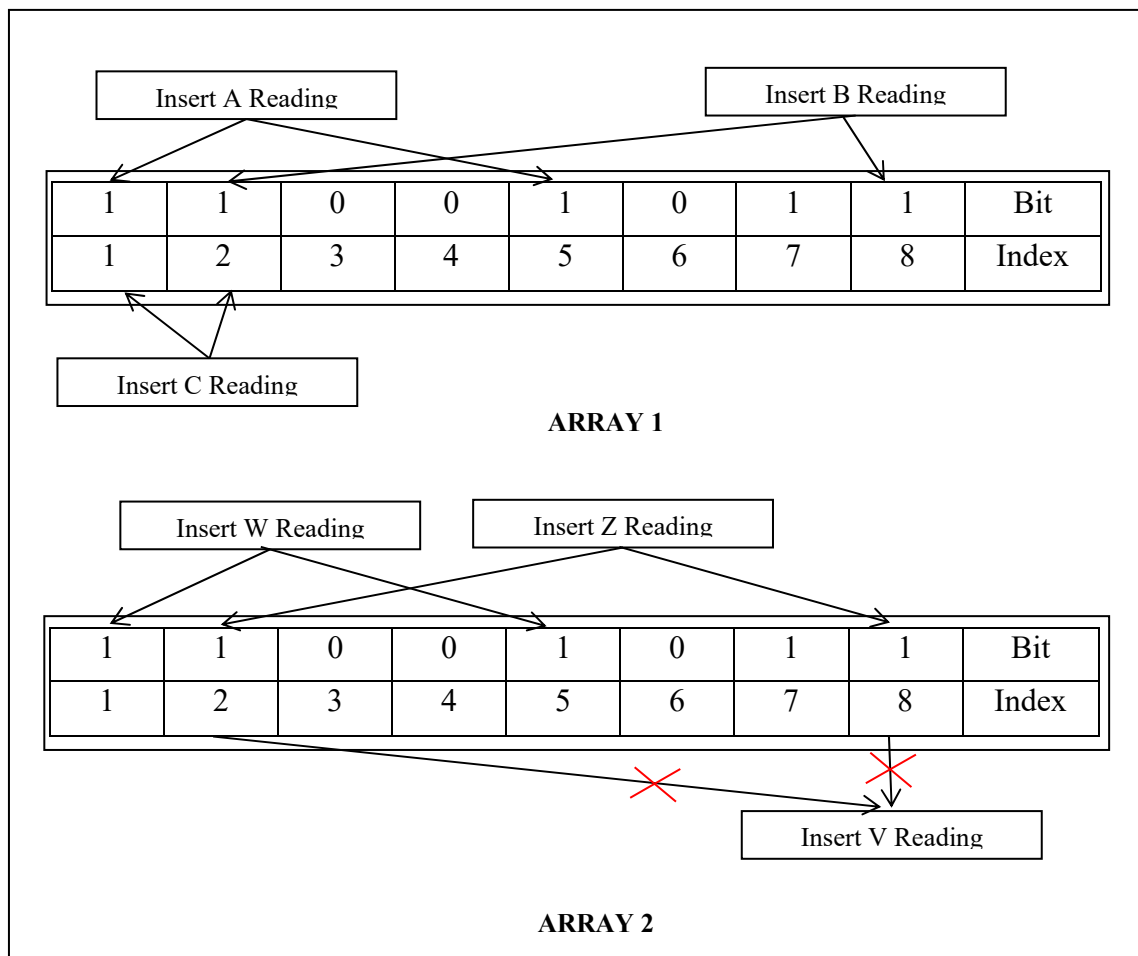
Fig 2: DHBF Process

INPUT: RID
BEGIN
    IF (Time == True) **then**
     array1 = array2 = {0}
    ENDIF
FOReach RID **do**
Pos 1 ← Hash1(RID)
Pos 2 ← Hash2(RID) //get two hash values for RID
    IF ((array1[pos1 & pos 2] == 1) && (array2[pos1& pos 2] == 1)) **then**
    Remove Duplicate reading
    ELSEIF ((array1[pos1 & pos 2] == 0) || (array2[pos1& pos 2] == 0)) **then**
    Array1[Pos1 & Pos2] ← 1 //store in array1
    ELSEIF ((array1[pos1 & pos 2] == 0) || (array2[pos1& pos 2] == 1)) **then**
    Array1[Pos1 & Pos2] ← 1 //store in array1
    ELSE ((array1[pos1 & pos 2] == 1) && (array2[pos1& pos 2] == 0)) **then**
    Array2[Pos1 & Pos2] ← 1 //store in array2
ENDIF
ENDFOR

Algorithm 1: Algorithm for Double Hash Bloom Filter

Instead of using a single array, the proposed MLDA exploits two arrays and minimizes the false positive, but it is not eliminated. There is less possibility for the corresponding bits of one reading to set as 1 when inserting other readings, resulting in a false positive. It is demonstrated in figure 2, and it explains two different cases in the DHBF algorithm.

Case 1: In figure 2, a sensor reading W returns the same positions (1 and 5) of reading B in array 1, and so it is stored in array 2.

Case 2: A sensor reading V also returns the same positions (2 and 8) of reading B in array 1 and reading Z in array 2 using the same set of k hash functions. Thus, the DHBF returns that it is a duplicate reading, but it is a false positive since the reading V is not stored in the parent node already.

If a sensor reading returns the new positions, not stored in both the arrays, it means that the reading is not defined in the array. After that, it is appended in the array. The readings in the arrays are cleared after performing the data aggregation. Using a Bloom filter, there is no way to eliminate the false positive completely, and thus a special technique, such as the Observe scheme assists in addressing and solving this problem effectively.

3.5 *MAX-MIN functions using Merge Sort*

Like the AVG and SUM aggregation measures, the MAX and MIN are also the basic measures for the observed temperature in an IoT environment. Sorting temperature readings in an area at the parent node during MAX and MIN aggregation measures in an energy-efficient way is a challenging task to perform. There are several sorting algorithms developed, but their execution time is still to be optimized. The merge sorting is an efficient technique since it exploits a comparison-based sorting algorithm [Auger, *et al*.(2015)].

For instance, to sort n temperature readings in the array, A[1,....n], the following steps are executed.

Divide: It divides the temperature readings into two lists of n/2 readings. It continues the process until reaching the smallest unit, one reading.

Conquer: It sorts the readings using merge sorting compares each reading with the adjacent list to sort.

Merge: It merges those sublists to generate a sorted list.

Thus, the proposed scheme efficiently supports SUM, AVG, MAX, and MIN aggregation functions in an aggregation layer by reducing the amount of redundant temperature readings transmission and energy consumption of sensor devices. The final readings are accessed in the application layer via the transport layer from the aggregation layer.

4. Performance Evaluation

The proposed protocol is assessed using the Cooja simulator. The simulation settings are utilized to complete an execution investigation of the proposed MLDA to contrast with CLCP [Alkhamisi, *et al*(2016)]. In these simulations, 15, 30, 45, and 60 homogeneous sensor nodes are deployed with equal battery energy, and they are placed in an area of 500×500 m^2. Based on the proposed data aggregation approach, the network performance is simulated in terms of the Aggregation Ratio, Transmission count coefficient, and Energy cost by varying the number of sensors.

1. **Aggregation Ratio:** The ratio of the number of aggregated packets to the number of generated packets.

2. **Transmission Count:** The ratio of the number of aggregated packets compared to the total packets that have been aggregated.

3. **Energy Cost:** The ratio of consumed energy per parent node during data aggregation to the energy of a parent node.
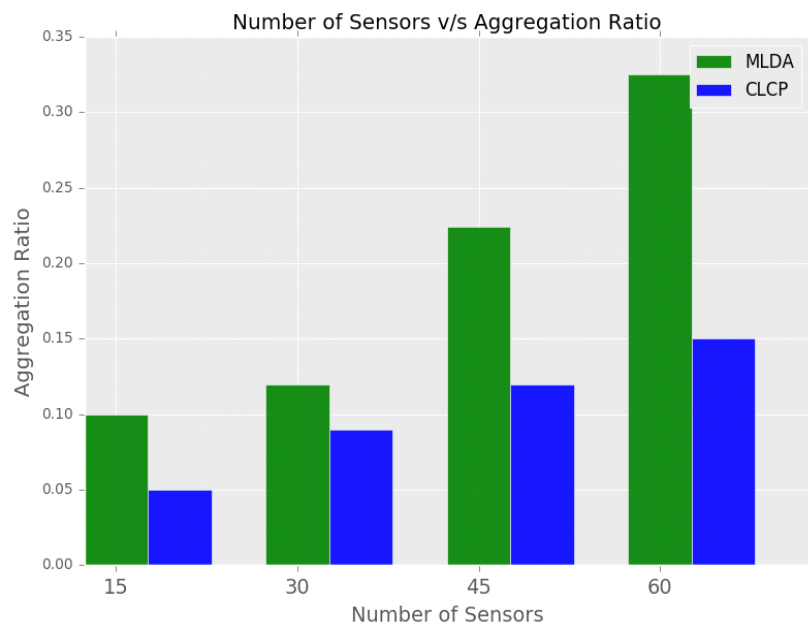
4.1 *Simulation Results*

Fig 3: Number of Sensors vs. Aggregation Ratio

From figure 3, it is observed that the aggregation ratio of MLDA gets incremented when the number of sensors or the node density per particular area increases. Compared to MLDA, the CLCP also increases the aggregation ratio with the number of sensors, but it achieves a poor aggregation ratio than MLDA. In the case of high node density, the CLCP experiences a sudden decrease in aggregation ratio. High network traffic incurs packet loss in CLCP since it builds cluster structure using energy and distance factor only. The MLDA experiences up to 0.325 of aggregation ratio, while that of CLCP is above 0.15. In Figure. 3, the aggregation ratio in the proposed work is higher than 0.1 in all four network traffic scenarios. Because the redundant transmission removal using DHBF and observation scheme in the proposed work avoids unnecessary packet loss and ensures a better aggregation ratio.

Figure 4 demonstrates the result of transmission count for both the proposed MLDA and CLCP. When the network has high traffic, the MLDA attains less transmission count than other scenarios. As the spatially related sensors generate similar temperature readings, MLDA can find redundant readings and remove them during data aggregation.
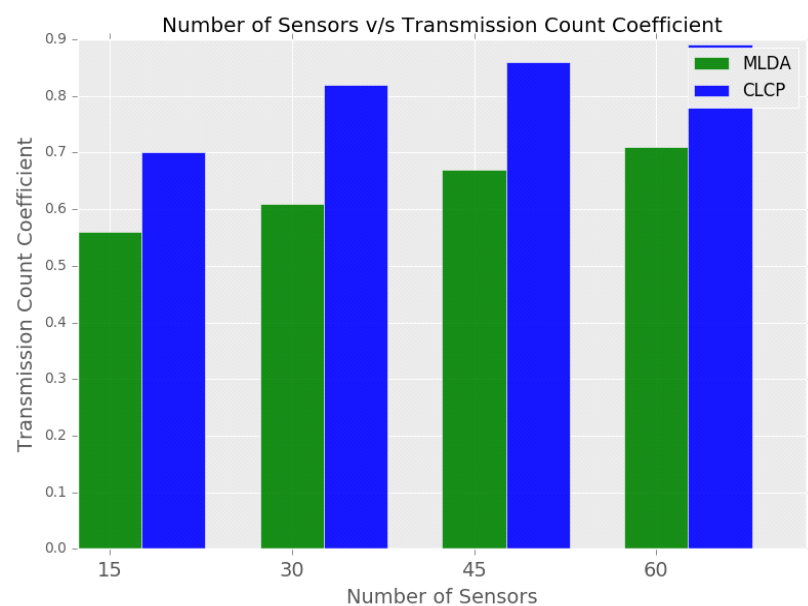

Fig 4: Number of Sensors vs. Transmission Count

However, CLCP work does not handle the data redundancy problem. The CLCP attains poor transmission count since it experiences more packet loss due to the lack of considering expected transmission count and traffic impact on communication. When the network traffic increased, the transmission count of CLCP is increased. Because the CLCP builds a separate cluster structure and simply applies aggregation function on all the received temperature readings. It increases the transmission count as well as the delay of packets. It results in poor communication performance in CLCP than MLDA. For instance, the MLDA attains 0.891 transmission count when the number of sensors is 60, whereas the CLPC attains 0.71 in the same scenario.
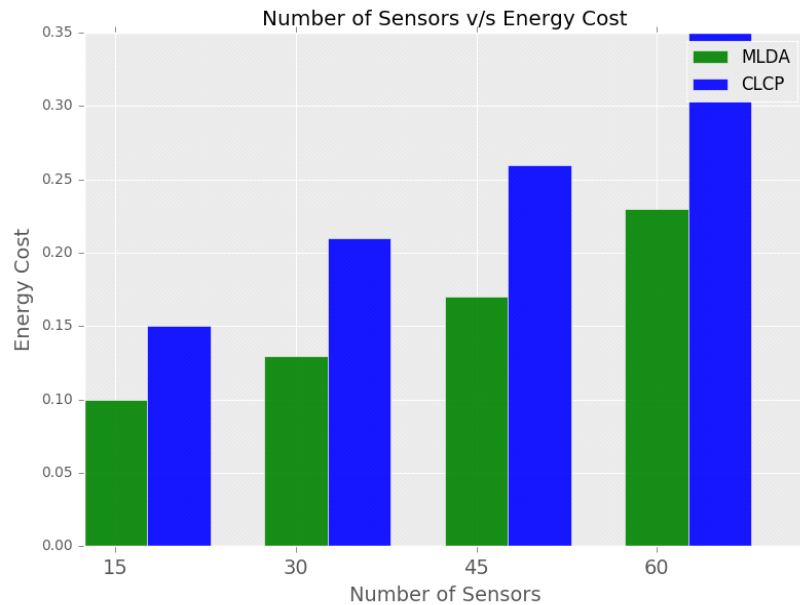

Fig 5: Number of Sensors vs. Energy Cost

Figure 5 shows that the proposed MLDA provides less energy cost due to network topology reuse from the network layer and redundant transmission removal. The MLDA based observation scheme enables the nodes to mostly avoid the transmission of redundant readings more than two times and to execute the aggregation function successfully at parent nodes. Another reason behind that the MLDA provides energy-efficient data aggregation is removing separate topology creation in the application layer. Moreover, with the number of sensors, the energy cost increases in both works. When the number of sensors is 60, the energy cost of MLDA is 0.23, whereas CLCP attains 0.35 energy cost in the same scenario. Moreover, the difference in energy cost between the proposed and existing works is approximately more than 5% in all scenarios.

5. Conclusion

This paper presented an energy-efficient aggregation scheme for IoT with the support of multiple layers, such as network and aggregation layer. To attain energy-efficient IoT communication, the proposed MLDA has incorporated the composite metric based network structure creation and such network structure based data aggregation in the aggregation layer. By applying the double hash filter model and observation field, the impact of redundant data transmissions on network resources and data aggregation efficiency are eliminated. Finally, the proposed scheme performs SUM, AVG, MAX, and MIN aggregation functions without redundant data transmissions, and merge sort is useful in executing the MAX and MIN aggregation functions. From the simulation results, it is concluded that the performance of MLDA is better than the CLCP in all the metrics. The proposed MLDA increases the aggregation ratio from 0.10 to 0.35 when the number of sensors has increased from 15 to 60. However, the existing work varies the aggregation ratio from 0.04 to 0.14.

References

[1] Dehkordi, S.A.; Farajzadeh, K.; Rezazadeh, J.; Farahbakhsh, R.; Sandrasegaran, K.; Dehkordi, M.A. (2020): A survey on data aggregation techniques in IoT sensor networks. Wireless Networks, 26(2), pp.1243-1263.
[2] Karamitsios, K.; Orphanoudakis, T. (2017): Efficient IoT data aggregation for connected health applications. In 2017 IEEE Symposium on Computers and Communications (ISCC) pp. 1182-1185.
[3] Chandnani, N.; Khairnar, C.N. (2020): A comprehensive review and performance evaluation of recent trends for data aggregation and routing techniques in IoT networks. Social Networking and Computational Intelligence, pp.467-484.
[4] Wala, T.; Chand, N.; Sharma, A.K.(2020): Energy efficient data collection in smart cities using iot. In Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's pp. 632-654.
[5] Guimaraes, V.G.; de Moraes, R.M.; Obraczka, K.; Bauchspiess, A.(2019): A novel IoT protocol architecture: Efficiency through data and functionality sharing across layers. In 2019 28th International Conference on Computer Communication and Networks (ICCCN) pp. 1-9.

[6]   Pourghebleh, B.; Navimipour, N.J.(2017): Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research. Journal of Network and Computer Applications, 97, pp.23-34.
[7]   Guidi, B.; Ricci, L.(2019): Aggregation techniques for the internet of things: An overview. The Internet of Things for Smart Urban Ecosystems, pp.151-176.
[8]   Fitzgerald, E.; Pióro, M.; Tomaszwski, A.(2018): Energy-optimal data aggregation and dissemination for the Internet of Things. IEEE Internet of Things Journal, 5(2), pp.955-969.
[9]   Bhandari, S.; Sharma, S.K.; Wang, X.(2017): Latency minimization in wireless IoT using prioritized channel access and data aggregation. In GLOBECOM 2017-2017 IEEE Global Communications Conference,pp. 1-6.
[10] Ghate, V.V.; Vijayakumar, V.(2018): Machine learning for data aggregation in WSN: A survey. International Journal of Pure and Applied Mathematics, 118(24), pp.1-12.
[11] Lu, R., Heung, K.; Lashkari, A.H. ; Ghorbani, A.A.(2017): A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access, 5, pp.3302-3312.
[12] Alkhamisi, A.; Nazmudeen, M.S.H.; Buhari, S.M. (2016): A cross-layer framework for sensor data aggregation for IoT applications in smart cities. In 2016 IEEE International Smart Cities Conference (ISC2) pp. 1-6.
[13] Amarlingam, M.; Mishra, P.K.; Rajalakshmi, P.; Channappayya, S.S.;Sastry, C.S.(2018): Novel light weight compressed data aggregation using sparse measurements for IoT networks. Journal of Network and Computer Applications, 121, pp.119-134.
[14] Yu, T.; Wang, X.; Shami, A. (2017): Recursive principal component analysis-based data outlier detection and sensor data aggregation in IoT systems. IEEE Internet of Things Journal, 4(6), pp.2207-2216.
[15] Preetha, S.S.L.; Dhanalakshmi, R;Kumar, R.(2018): An energy efficient framework for densely distributed WSNs IoT devices based on tree based robust cluster head. Wireless Personal Communications, 103(4), pp.3163-3180.
[16] Ko, H.; Lee, J.;Pack, S.(2019): CG-E2S2: Consistency-guaranteed and energy-efficient sleep scheduling algorithm with data aggregation for IoT. Future Generation Computer Systems, 92, pp.1093-1102.
[17] Li, R.; Sturtivant, C.; Yu, J.;Cheng, X.(2018): A novel secure and efficient data aggregation scheme for IoT. IEEE Internet of Things Journal, 6(2), pp.1551-1560.
[18] Abdul-Qawy, A.S.H.; Srinivasulu, T.(2019): SEES: a scalable and energy-efficient scheme for green IoT-based heterogeneous wireless nodes. Journal of Ambient Intelligence and Humanized Computing, 10(4), pp.1571-1596.
[19] Appleby A.(2013): "Murmurhash 2.0,".
[20] Fowler, G.; Noll, L.C.; Vo, K.P.; Eastlake, D.;Hansen, T., (2011): The FNV non-cryptographic hash algorithm. Ietf-draft.
[21] Auger, N.; Nicaud, C.; Pivoteau, C.(2015); Merge strategies: from merge sort to Timsort. URL https://hal-upec-upem. archives-ouvertes. fr/hal-01212839, working paper or preprint.

Authors Profile

Vanathi Arunachalam, Received her B.E in CSE from Bharadhidasan University, Trichy,T.N, India and M.E., in CSE from Anna University Chennai, TN , India. She is currently pursuing Ph.D in Acharya Nagarjuna University, Guntur. She was a lecturer,Assistant Professor and currently working as an Associate professor and Head Of the Department CSE ,Aditya Engineering College, Surampalem,AP, India.She is qualified in APSET and TNSET. She published papers in international journals and conferences. She guided a Project " Automated Plant Watering" ,which is selected for the First Prize with Cash award of Rs 50,000/- in CSI InApp , a National Level Project Competetition conducted by CSI Trivandrum Chapter. Her research interests include Information Security, Mobile Computing, WSN and IoT. She is a Lifetime Member of CSI and ISTE..

Dr. N. Nagamalleswara Rao, working as Professor, Department of IT, RVR& JC College of Engineering, Guntur (Dt). He Completed his B.Tech(1988) in Bapatla Engineering College, M.Tech (1991) in Motilal Nehru National Institute of Technology, Allahabad and Ph.D (2013) in Acharya Nagarjuna University. He has 30 years teaching experience. His research interests are Computer Algorithms, Compilers, Image Processing. He has published papers in many International conferences, National Conferences and International Journals. He has published many patents. Three Scholars completed their PhD under his guidance. He is a Life member of CSI, ISTE, IEEE.