International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 9, No.2 (Mar-2020)
http://dx.doi.org/10.12785/ijcds/090214

Data Aggregation Mechanisms in the Internet of Things: A Study, Qualitative and Quantitative Analysis

Ab Rouf Khan[1] and Mohammad Ahsan Chishti[1]

[1]*Department of Computer Science and Engineering, National Institute of Technology Srinagar, Hazratbal Srinagar, India.*

Received 09 Mar. 2019, Revised 28 Jan. 2020, Accepted 23 Feb. 2020, Published 01 Mar. 2020

Abstract: Internet of Things (IoT), a paradigm added to the ever-growing technological arena in recent times acts like a bridge between the things in the physical world and their representation within the digital world. The basic "things" in the IoT are sensor devices, which gather as well as monitor all types of data on physical machines and human social life. IoT enables data sending and receiving for each "thing" through the communication network. The purpose of Data Aggregation is to decrease the number of communications/transmissions among the objects/things in the Internet of Things framework. The effectiveness of the data aggregation technique employed is a key factor in the success of IoT systems in terms of data freshness and efficiency. Different data aggregation techniques have been proposed in the recent past, which include – Tree-Based, Cluster-Based and Centralized data aggregation techniques. The paper aims at a detailed study and analysis of data aggregation schemes employed in the Internet of Things in terms of working and time complexity. Lowest Common Ancestor (LCA) aided Tree-Based Data Aggregation algorithm is designed. In addition, the Cluster-Based data aggregation algorithm incorporated with the β-dominating set and Centralized Data Aggregation algorithm incorporated with the $SUM()$ aggregation function are proposed. The algorithms are supported by well-formed flowcharts describing the flow and working of the data aggregation mechanisms designed. The results are obtained on a system consisting of 60 nodes with all the three aggregation algorithms being evaluated against each other. The centralized data aggregation algorithm is better when the number of nodes in the network is lesser. However, as the number of nodes increases, the cluster-based and tree-based algorithms produce better results as compared to the centralized data aggregation algorithm.

Keywords: Data Aggregation, Tree-Based, Cluster-Based, Centralized Data Aggregation, Time Complexity, and Internet of Things

1. INTRODUCTION

Internet of things (IoT) is a new computational dimension added to the modern arena of technology. The sensors and objects on the internet of things are integrated in such a way that they communicate with each other without any human intervention. The "things" in the internet of things include physical devices, such as the devices capable of sensing data that gather and monitor different kinds of data related to human social life and the machine data [1]. The IoT enables the sensor devices to visualize, think, hear, and perform tasks by making the devices to "talk" with each other, to coordinate decisions as well as sharing of information. Devices in IoT are transformed from being "traditional" to "smart" by exploiting the underlying technologies – embedded devices, sensor networks, ubiquitous and pervasive computing, communication technologies, internet protocols and applications [2]. As per the survey conducted by GSMA, 1.3 trillion dollar revenue opportunities can be generated for mobile network operators alone – covering the segments such as utilities, automotive, consumer electronics and health [3][4]. A large number of nodes in the internet of things framework collect the data from the smart devices. The data nodes usually vary from each other and are autonomous in nature. Among the major inherent features of the internet of things is heterogeneity and being distributive in nature, which leads to many of the IoT applications of being in need of information that is distributed on multiple nodes of data. The data nodes exchange the information and cooperate with each other to complete the assigned tasks [5].

Data aggregation comprises collecting data from different objects in the Internet of Things framework and representing it in a summarized form. To minimize the bandwidth and energy consumption, the data aggregation mechanisms are used in IoT largely. Furthermore, these aggregation schemes enhance security as well. A rather simpler approach to the data aggregation is that all the source nodes collect the data from different sources, send this data, which is not pre-processed, at any level to a

E-mail: khanrouf_17@nitsri.net, ahsan@nitsri.net

centralized aggregator node, and perform the different data aggregation functions directly on the aggregated data [6][7]. Interactivity, connectivity, and sensing among objects are treated as the primal features of IoT. Despite the heterogeneity of the IoT devices, the data in IoT applications – smart energy, smart home, and smart city services could be combined, merged, compared and correlated easily to meet the needs and requirements of people. Batteries with limited power supply typically power devices in IoT. Thus, the lifetime of the applications, which are of long-term nature for instance environmental monitoring needs to be improved. In addition, the processing power at base stations is not up to the mark of handling the data, which is ever-increasing in size. Thus, we need to focus on making the data aggregation techniques more efficient within the network itself where the nodes are energy as well as resource-constrained. Data aggregation strategies work at the basic/physical level to collect and aggregate the data from the different source nodes in an efficient manner in order to improve the quality of service parameters of the network including – network lifetime, data accuracy, energy consumption, and traffic bottleneck. Data aggregation eliminates the redundancy in the data, which saves energy. The efficiency of any data aggregation strategy depends upon the size of sensing data and the network design. In addition, some malicious attacker may attack the data aggregator nodes during the aggregation process. Thus, the accuracy of the data aggregated cannot be guaranteed by the base station in case the middle node is compromised [8][9]. The general scenario of the Data Aggregation process is demonstrated in Figure 1.

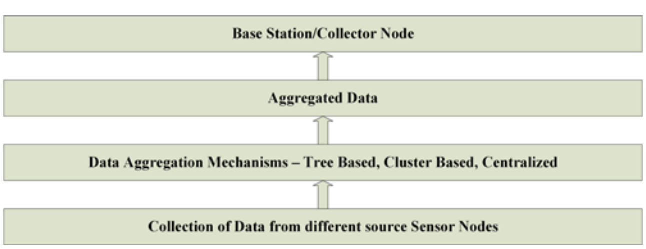

Figure 1. The general scenario of data aggregation process in IoT

2. DATA AGGREGATION IN IOT – BACKGROUND

Researchers in the field of Internet of things in the recent past have proposed new ways and techniques to carry out the process of data aggregation in order to increase its efficiency as well as make the process precise. This section discusses the work carried in the area of data aggregation by prominent experts in the field. The three techniques – tree-based, cluster-based and centralized will be discussed sequentially.

The authors in [10] have proposed a lifetime-balanced data aggregation scheme for IoT. The suggested model while increasing the end-to-end network constraints and the network lifetime guarantees the requisite data delivery delay. From the perspective of device lifetime, the focus is mainly on network lifetime. For lifetime balancing among the devices without the end-to-end delay increase, the delays in aggregation among the devices are adjusted together in a cooperative manner. However, working with multiple sinks does not yield better results.

Hitch hacker – a constituent binding model was proposed by G. S. Ramachandran et al. [11] to handle the aggregation of data and corroborate multi-hop aggregation of data in IoT. This technique uses metadata. The bindings are divided into low or high priority. The component bindings provide meta-data to make a multi-hop data aggregation. A central metadata manager is employed to perform the discovery of the routes on the multi-hop network. However, it does not consider the issues node heterogeneity and accuracy.

The authors in [12] have suggested a mechanism to provide security and efficiency in the data aggregation. In this method, to design an accurate secure data aggregation method the communication and computational limitations are allied to the IoT network while keeping security features in mind. Security is guaranteed by this mechanism, but suffering from high traffic load is the limitation.

A. Koike et al. [13] have suggested a data aggregation method and implemented the same in a WAN to corroborate IoT traffic. Apart from implementation, the requirements for the proposed method in a wide area network are also elicited. From the viewpoint of architecture, the technique creates overlay networks, which creates a logical network and reduces the load of the packet processing at the router. Only the packets are aggregated in this mechanism and no packet information is changed. However, the disadvantage lies in the fact that it experiences high latency.

A data aggregation method, which provides security for IoT data aggregation proposes to encrypt the data from sensors using a seed exchange, based elliptic curve algorithm and data transformation based on a Hilbert curve. It uses Tree-based architecture to perform data aggregation. It makes the tracing of actual value difficult even to the attackers. The proposed data aggregation method is better than many of the methods proposed in the recent past for energy preservation and privacy as per the analysis related to performance, but suffering from high traffic-load is the disadvantage [14].

The authors in [15] have proposed a scalable and secure internet of things stowing system in terms of scalability, security, reliability, and flexibility to accomplish the requisites for data mining and analytics with aggregated data massive in size. To achieve the security of data deprived of complex key management, a revised secret sharing scheme is the basis of design.

CRT – Chinese Remainder Theorem built on a coding algorithm for the collection of data in IoT is proposed. To sense the data, this mechanism is quite suitable. The data, which is massive in size, is sensed in the form of multiple shorter residues. The aggregator node receives all sensing data from the source nodes, performs the data aggregation process and sends the outcome to the base station. However, the final collected data suffers from low accuracy [16].

The authors in [17] have presented a privacy-preserving aggregation method that allows aggregation of multi-attribute nature for clusters. The privacy is preserved using the correlation function. This technique allows data aggregation regarding many attributes of each object in a single operation guaranteeing privacy and authenticity of data. Moreover, the method can deal with situations of large-scale in nature, allowing detection of malicious manipulation of the data aggregated. The proposed mechanism focuses on privacy-preserving, correlative aggregation and collision resistance verifiable aggregation. The proposed method suffers from high latency.

A. R. G. Ramirez et al. [18] have presented a cross-layer structure for data aggregation for the applications in IoT. No assumptions regarding static infrastructure are made while handling situations that are either ad-hoc or mobile in nature. Operating both on network and application level, it safeguards tolerances related to network failures.

The authors in [19] have provided an architecture for aggregation of data in the wireless sensor-based internet of things platform using infrastructures based on traditional ICT technology. In addition, keeping in mind the parameters – availability, confidentiality, and integrity, a comprehensive risk analysis is provided. However, the data aggregator platforms complexity disclosures the suggested architecture to different threats including security.

Many of the IoT applications require information to be collected from a large number of source sensor nodes, an architecture, which is distributive as well as service-oriented in nature, is proposed. The nodes are divided into – data nodes and query nodes. Data nodes collect the data while query nodes take care of the assigned product, thereby making the system scalable. Low availability is the disadvantage of this mechanism [5].

3. DATA AGGREGATION MECHANISMS IN THE INTERNET OF THINGS

Different data aggregation techniques/mechanisms in IoT have been proposed from time to time. However, the main mechanisms that have been found feasible and efficient for IoT include – Tree-Based, Cluster-Based, Centralized, P2P and Distributed Mechanisms. Out of these, the ones mainly focused on include – Tree-Based and Cluster-Based Mechanisms. However, all these mechanisms need to be studied well and acted upon to make them better and efficient. This section discusses the proposed LCA aided Tree-Based data aggregation mechanism and the modified β-dominating set based Cluster-Based and Centralized data aggregation techniques in detail. The architecture, flowchart, and algorithms of each technique are proposed.

A. LCA aided Tree-Based Data Aggregation Mechanism

In the Tree-Based Data Aggregation Mechanism, source sensor nodes collect the data from the different sources and transmit the same to the immediate node based on the distance metric. The intermediate node acts as the aggregator. An intermediate/hierarchal node performs the aggregation process [20][21]. Figure 2 expresses the design of the tree-based data aggregation mechanism. The aggregator nodes forward the data after aggregation from the source nodes to the sink/collector node [22].

LCA aided Tree-based data aggregation technique aims at aggregating the data by building a tree data structure. The source nodes collect the data from various application domains of the internet of things framework. Using the Lowest common ancestor (LCA) algorithm alongside the Breadth-First Algorithm, we can compute the distance between each source node and the aggregator to decide whether to forward the data or not. Depending upon the size of the application, the number of source nodes is fixed. Source nodes are marked in the range of 1 to n. Source nodes after collecting the data send the data to the immediate node in the hierarchy – that is, the nodes in the tree at leaf node level (l) send the data to the nodes present at the next level $(l-1)$. Depending upon the volume of the network and the capacity of aggregator nodes, the number of aggregator nodes is fixed. The aggregator node can be an intermediate node in the tree and in certainty is the parent of one or more nodes in the level immediately below to it. The tree turns out to be an m-ary tree with each parent (aggregator) node having multiple child nodes. There can be multiple numbers of aggregator nodes, which aggregate the data and transmit the aggregated data to the sink/collector node or for that case to the base station directly.

The network can be represented in the form of Graph, $G=(V,E)$ where V – represents the set of sensor nodes, $|V|=n$ – represent the total number of nodes deployed in the region. It can be a real-time environment or any simulation setup. Assume that each node at the leaf node level of the tree has the capability to transmit/forward the data to the next immediate node in the level next to the leaf node level is some cutoff distance d.

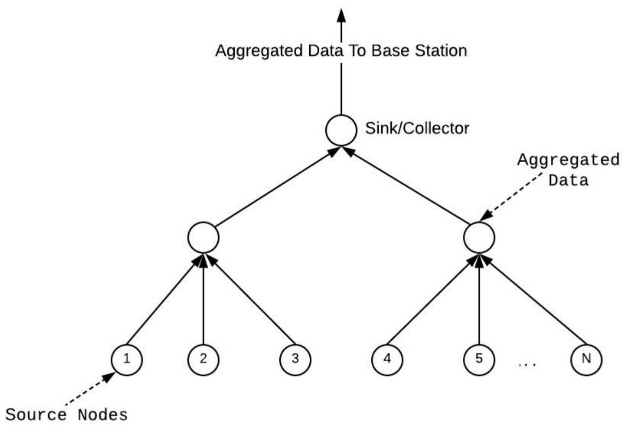

Figure 2. Tree-Based Data Aggregation Mechanism in IoT

The LCA aided Tree-Based Data Aggregation Algorithm is given as below:

Algorithm 1. LCA aided Tree-Based Data Aggregation Mechanism in IoT

Input: Graph G = (V, E), Distance $d\,(i, P_i)$
Output: Data Aggregation Tree $DAT\ (V, E)$

procedure TreeBasedDataAggregation $(G = (V, E))$
Choose any node as Root Node
Apply Breadth-First Traversal Algorithm on the Graph $G = (V, E)$
Using the Lowest Common Ancestor (LCA) algorithm, calculate the distance between each node i_1 & its parent node, $d = |i - P_i| = \sqrt{\sum_{i=1}^{n}(i_1 - P_i)^2}$
for each i in V **do**
  if(d_i <= $CutoffDistance$)
    BroadcastMessage (i, P_i)
  else
  choose other Parent P_i
end for
for each P in P_i **do**
  $Z = \sum_{i=1}^{P}(P_1, P_2, P_3, \ldots, P_i)$
  BroadcastMessage$(Z, SinkNode(C))$
end for
Return Z
end procedure

The nodes that have the distance less than or equal to some cutoff distance d with respect to the aggregator nodes are the only ones actually able to transmit the data to the particular aggregator node, i.e.

$$\forall n \in V\ if\ \{(distance \le d) - Transmit\};$$
$$else\ \{Look\ for\ another\ aggregator\ node\}$$

We will denote the sink/collector node by C and aim at building a data aggregation mechanism based on the m-ary tree. Out of the nodes in V, a set of nodes, that are child nodes of a particular parent (aggregator) node are represented by $S_1, S_2, S_3, \ldots, S_i$. Set S_i can consist of variable a number of nodes depending upon the residual capacity – the deciding factor for distance metric d of the sender nodes. $S_1, S_2, S_3, \ldots, S_i$ correspond to aggregator/ parent nodes $P_1, P_2, P_3, \ldots, P_i$ respectively. The aggregated data is obtained by using the SQL aggregator function $SUM()$, i.e.

$$Aggregated\_Data = SUM(P_1, P_2, P_3, \ldots, P_i)$$

The objective function, Z defining the aggregated data is represented as:

$$Z = \sum_{i=1}^{n} \mathrm{d}(\mathrm{i}, P_i)$$

Where $\mathrm{d}(\mathrm{i}, P_i)$ represents the distance between the node i and its nearest parent node. The flowchart depicting the flow and working of the tree-based data aggregation mechanism in the IoT is presented in Figure 5 (a).

B. Cluster-Based Data Aggregation Mechanism

The cluster-based mechanism works by dividing the entire network into various clusters, with every cluster consisting of a lot of sensor nodes. From every cluster, one node – header/leader node is selected based on an election algorithm, which becomes a cluster-head. Bandwidth overhead reduction as well as transmitting a lesser number of packets is the result of this mechanism [23][24]. Cluster-based data aggregation architecture is shown in Figure 3.

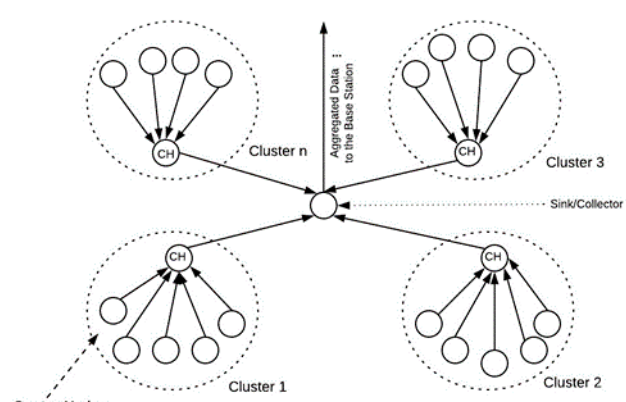

Figure 3. Cluster-Based Data Aggregation Mechanism in IoT

The proposed β-dominating set based Cluster-Based Data Aggregation algorithm is given below.

Algorithm 2. Cluster-Based Data Aggregation Mechanism based on β-dominating set in IoT

Input: Graph G = (V, E), $S = (S_1, S_2, \dots, S_K)$ - set of predefined $K - centroids$, Distance β
Output: Data Aggregation Cluster Formation

procedure ClusterBasedDataAggregation $(G = (V,E)), S = (S_1, S_2, \dots, S_K)$
Compare the value of the distance β, for each sensor node $(n \in V)$ with a predefined centroid of each cluster $(1 \leq m \leq k)$
for each Sensor node-distance β, find the closest centroid S_m and assign the node to cluster m
Set $ClusterID[1] = m$
Set $currentClosestDistance[1] = d(\beta, S_m)$
for each cluster m in K **do**
    Calculate the distance d of a sensor node i from the centroid of the closest cluster
    $if(d_i \leq currentClosestDistance)$
        Sensor node i stays in the current cluster
    else
    for each Centroid m in K **do**
        Re-calculate the distance $d(\beta, S_m)$
    end for
    Assign the Sensor to the cluster with the closest centroid m
    Set $ClusterID[1] = m$
    Set $currentClosestDistance[1] = d(\beta, S_m)$
end for
for each cluster m in K **do**
  $Z = \sum_{i=1}^{K}(K_1, K_2, K_3, \dots, K_m)$
  BroadcastMessage$(Z, SinkNode(C))$
end for
Return Z
end procedure

The cluster-based data aggregation mechanism in IoT can be implemented by pre-defining a k-number of clusters. We propose to obtain the cluster heads of each cluster with the help of defining a β-dominating set. Given an IoT network represented in the form of a Graph $G = (V, E)$, a set C⊆ V is a β-dominating set if every node of G either belongs to C or is within a distance no more than β to one or more nodes of C. The inclusion of β-dominating increases the efficiency of the cluster formation. After obtaining the cluster heads of each cluster, the source nodes belonging to a particular cluster send the data to the respective cluster heads. The aggregated data at each of the cluster heads are sent to the base station using the basic mathematical summation function running from 1 to the number of clusters (k). Depending upon the application, the Sink/Collector can act as the base station itself in certain situations to increase the efficiency of the mechanism.

Cluster-based data aggregation mechanism does scale well even when the network is large because cluster formation takes care of the scalability in the network. Thus, a cluster-based data aggregation mechanism can be a solution to aggregate the data in the IoT based architectures massive in size which deal with big data. The flowchart depicting the working of this aggregation mechanism in IoT is given in Figure 5 (b).

C. Centralized Data Aggregation Mechanism

In the Centralized mechanism, each source/immediate node sends the data to the centralized header node via the minimum cost route [25]. Header Node is responsible for Data aggregation in this technique, which receives the data from the different sensor nodes. The header node is computationally stronger than the other nodes. From the header node, the data aggregate is progressed to the base station. Centralized data aggregation architecture is shown in Figure 4.

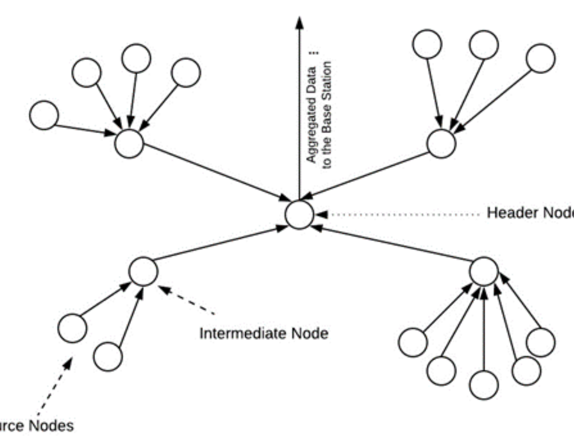

Figure 4. Centralized Data Aggregation Mechanism in IoT

We propose that the Centralized data aggregation mechanism in the IoT can be implemented by choosing the Leader Node randomly initially. The flowchart depicting the working of a centralized data aggregation mechanism is given in Fig. 5 (c). We propose that each source node from multiple directions sends the data to the successor/ intermediate node. Each intermediate node sends the aggregated data to the Leader node via the minimum cost route using the Dijkstra’s algorithm. The data from the aggregators can be summed up using the SQL aggregator function $SUM()$ and the aggregated data is transmitted to the base station. The centralized aggregation suffers from grave scalability concerns due to the storing of large volumes of data at a single node. However, since there is a single Leader node, it is easy to handle the different metrics like route backup, route discovery, etc. of the network.

The proposed Centralized Data Aggregation with aggregate function $SUM()$ Mechanism in IoT is given below.

Algorithm 3. Centralized Data Aggregation with aggregate function $SUM()$ Mechanism in IoT

Input: Graph G = (V, E), Header Node H – Chosen randomly, $S_{IN} = (IN_1, IN_2, \dots, IN_m)$ – set of intermediate nodes

Output: Centralized Data Aggregation

1: **procedure** CentralizedDataAggregation $(G = (V, E)), (H)$
2: Choose the Header Node H randomly
3: $\forall(Source\ Nodes\ S_i)$ from multiple directions, send data to the intermediate nodes IN_i
4: Calculate distance between each source node S_i and its intermediate node IN_i using Dijkstra's shortest path algorithm.
5: Choose the intermediate node IN_i with the least distance $d(IN_i, S_i)$ to the respective source nodes as the aggregator.
6: Transmit the data from all the intermediate nodes IN_i to the Header Node H using the SQL aggregate function $SUM()$
7: $Z = \mathrm{SUM}((IN_1, IN_2, \dots, IN_m))$
8: BroadcastMessage$(Z, SinkNode(C))$
9: Return Z
10: **end procedure**

4. Time Complexity of Data Aggregation Mechanisms in IoT

Time complexity deals with the quantification of the amount of time taken by an algorithm or pseudocode. Since each algorithm is essentially a well-formed combination of steps in a sequence, time complexity essentially deals with finding the order of magnitude of a statement – that is how many numbers of times a statement is executed in a program or an algorithm. The time complexity of an algorithm depends upon the logic of the algorithm rather than the specifications of the machine on which the algorithm is executed.

Given an Internet of Things network characterized in the practice of Graph $G = (V, E)$ – where V represents the total number of nodes in the network, E – the interconnections between the sensors, the complexity of the algorithms is depicted in Table 1.

The time complexity of the Tree-based and centralized data aggregation mechanisms is the same, however, the cluster-based mechanism with m number of clusters and k number of predefined centroids scales well in any network because the tree-based mechanism has a number of sub-modules in terms of functionality. The cluster-based algorithm has time complexity more than tree-based and centralized mechanisms but with the availability of the very powerful processors with multiple-cores, the process can be executed in a parallel fashion to increase its efficiency. The capacity of the links and the propagation delay does not affect the time complexity of the algorithms as such because the channel used to communicate between the nodes is having a constant delay in the worst-case scenario too. Comparing this constant value with respect to either the number of edges E or vertices V of the graph $G = (V, E)$, and using the Big O notation, we can simply ignore the constants in the time complexity.

5. Experimentation and Results

The three algorithms were implemented on a machine running 64-bit Ubuntu 18.04 operating system with configuration – Intel (R) Core(TM) i7-4600M CPU @ 2.90 GHz with 8 GB of RAM. OMNeT++ simulator was used to run the simulations [26]. The nodes were created by sub-classing the base class *cSimpleModule* of OMNeT++ simulator and the module gates *cGate* were used for the incoming and outgoing traffic connections. The channel used to communicate between the nodes is *cDelayChannel* which is the child class of base class *cChannel*. To make the system more realistic a false delay of 0.01 seconds was taken as a parameter in *cDelayChannel*. The maximum number of source nodes chosen is 60. The algorithms were implemented on the number of nodes ranging from 5 to 60 and the corresponding aggregation time was calculated. The results obtained are represented in the form of a bar chart given in Figure 6.

The centralized data aggregation algorithm is better when the number of nodes in the network is lesser. Until the number of nodes in the network reached 35, the centralized algorithm scaled well as compared to the other two algorithms. However, the aggregation time increased drastically as compared to the other algorithms when the number of nodes crosses 40. The reason for the drastic increase in the aggregation time is that the single header node is not able to handle the data from nodes as they increase in the number.

The cluster-based and tree-based algorithm scale almost equally in the experimentation results. As the number of nodes increases, the cluster-based and tree-based algorithms produce better results as compared to the centralized data aggregation algorithm. However, the tree-based algorithm is preferable when the data freshness is important. The cluster-based algorithm scales well as compared to the other two algorithms when the size of the network increases.

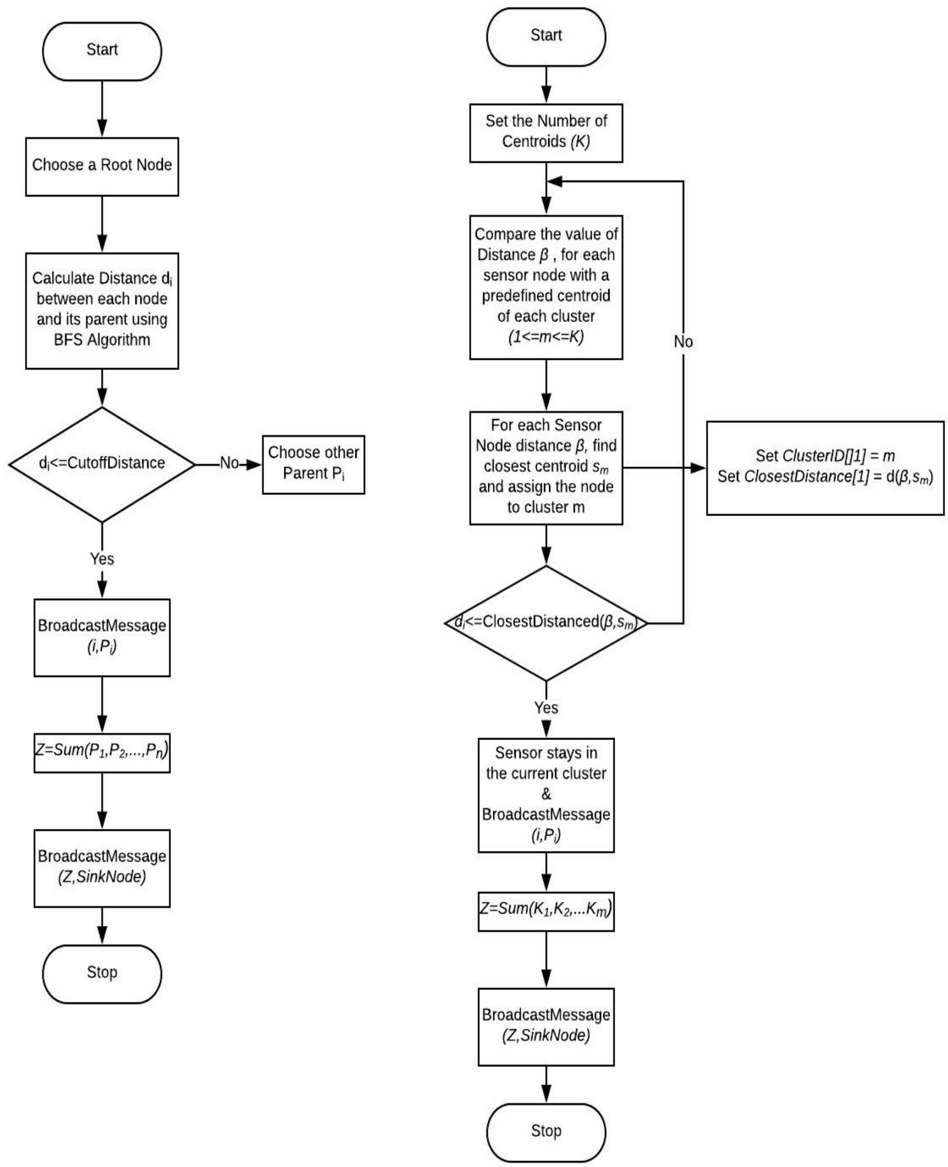

Figure 5. Flowcharts for **(a.)** LCA aided Tree-Based Data Aggregation Mechanism. **(b.)** Cluster-Based Data Aggregation Mechanism. **(c.)** Centralized Data Aggregation Mechanism

TABLE 1. Time Complexity analysis of Data Aggregation Mechanisms in IoT

Tree-Based Data Aggregation Algorithm	Cluster-Based Data Aggregation Algorithm	Centralized Data Aggregation Algorithm
• LCA Algorithm: TC = $O(V)$ • TC To calculate distances between $Node\ i$ and $Parent\ Pi$: $= O(ElogV)$ • To aggregate data from P parent nodes: TC = $O(P)$ • Total TC = $O(V) + O(ElogV) + O(P) = \boldsymbol{O(ElogV)}$	• m: Total no of clusters • k: Total no of predefined centroids • TC= $\boldsymbol{O((Vm) * k)}$	• Dijkstra's Algorithm: TC = $O(ElogV)$ • To aggregate data at the header node from m immediate nodes: TC=$O(m)$ • Total TC = $O(ElogV) + O(m) = \boldsymbol{O(ElogV)}$

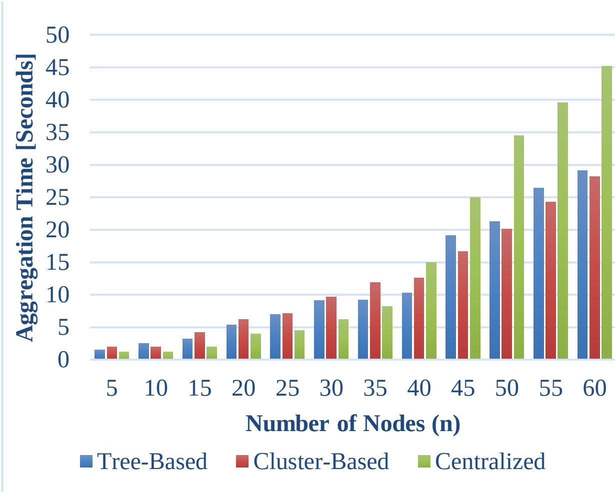

Figure 6. Comparison of Tree-Based, Cluster-Based and Centralized Data Aggregation Algorithms in terms of aggregation time and a number of nodes.

6. CONCLUSION

The Internet of Things has arisen as one of the most prolific paradigms of computing in recent times. Data in IoT is one of the main entities of the system. Data Aggregation in IoT aims at decreasing the number of transmissions among the communicating entities. The effectiveness of the data aggregation technique employed is a key factor in the success of IoT systems in terms of efficiency and data freshness. Three main data aggregation techniques in IoT – 'Tree-based', 'Cluster-based' and 'centralized' mechanisms are modified and evaluated against each other in terms of architecture, algorithms and time complexity. The flowchart and the pseudocode of all three mechanisms are given and evaluated against each other in terms of working and time complexity. Parameter – *data aggregation time* was chosen to validate the algorithms in the OMNeT++ simulator. The results reveal that centralized data aggregation scales well in the case of smaller networks as compared to tree-based and cluster-based data aggregation algorithms. However, for larger networks, the cluster-based algorithm scales well.

REFERENCES

[1] F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of Things security: A survey," *J. Netw. Comput. Appl.*, vol. 88, no. April, pp. 10–28, 2017.

[2] S. Li, L. Da Xu, and S. Zhao, "The internet of things : a survey," vol. 17, no. April 2014, pp. 243–259, 2015.

[3] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Futur. Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[4] H. Elazhary, "Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions," *J. Netw. Comput. Appl.*, vol. 128, no. October 2018, pp. 105–140, 2019.

[5] T. Zhu, S. Dhelim, Z. Zhou, S. Yang, and H. Ning, "An architecture for aggregating information from distributed data nodes for industrial internet of things," *Comput. Electr. Eng.*, vol. 58, pp. 337–349, 2017.

[6] P. Zhang, J. Wang, K. Guo, F. Wu, and G. Min, "Multi-functional secure data aggregation schemes for WSNs," *Ad Hoc Networks*, vol. 69, pp. 86–99, 2018.

[7] S. Goudarzi, N. Kama, M. H. Anisi, S. Zeadally, and S. Mumtaz, "Data collection using unmanned aerial vehicles for Internet of Things platforms," *Comput. Electr. Eng.*, vol. 75, pp. 1–15, 2019.

[8] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *J. Netw. Comput. Appl.*, vol. 97, no. July, pp. 23–34, 2017.

[9] S. Sanyal and P. Zhang, "Improving quality of data: IoT data aggregation using device to device communications," *IEEE Access*, vol. 6, pp. 67830–87840, 2018.

[10] Z. Li, W. Zhang, D. Qiao, and Y. Peng, "Lifetime balanced data aggregation for the internet of things," *Comput. Electr. Eng.*, vol. 58, pp. 244–264, 2017.

[11] G. S. Ramachandran *et al.*, "Hitch Hiker 2.0: a binding model with flexible data aggregation for the Internet-of-Things," *J. Internet Serv. Appl.*, vol. 7, no. 1, 2016.

[12] S. S. Sruthi and G. Geethakumari, "An Efficient Secure Data Aggregation Technique for Internet of Things Network: An Integrated Approach Using DB-MAC and Multi-path Topology," *Proc. - 6th Int. Adv. Comput. Conf. IACC 2016*, pp. 599–603, 2016.

[13] A. Koike, T. Ohba, and R. Ishibashi, "IoT network architecture using packet aggregation and disaggregation," *Proc. - 2016 5th IIAI Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2016*, pp. 1140–1145, 2016.

[14] A. Alghamdi, M. Alshamrani, A. Alqahtani, S. Safar, A. Al Ghamdi, and R. Harrathi, "Secure Data Aggregation Scheme in Wireless Sensor Networks for IoT."

15] H. Jiang, F. Shen, S. Chen, K. C. Li, and Y. S. Jeong, "A secure and scalable storage system for aggregate data in IoT," *Futur. Gener. Comput. Syst.*, vol. 49, pp. 133–141, 2015.

[16] F. Xie, "CaCa: Chinese Remainder Theorem Based Algorithm for Data Aggregation in Internet of Things on Ships," *Appl. Mech. Mater.*, vol. 701–702, pp. 1098–1101, 2014.

[17] L. González-Manzano, J. M. D. Fuentes, S. Pastrana, P. Peris-Lopez, and L. Hernández-Encinas, "PAgIoT - Privacy-preserving Aggregation protocol for Internet of Things," *J. Netw. Comput. Appl.*, vol. 71, pp. 59–71, 2016.

[18] A. R. G. Ramirez *et al.*, "Smart city architecture and its applications based on IoT," *Sensors (Switzerland)*, vol. 16, no. 11, pp. 611–616, 2016.

[19] H. Sandor, B. Genge, and Z. Gal, "Security Assessment of Modern Data Aggregation Platforms in the Internet of Things," *Int. J. Inf. Secur. Sci.*, vol. 4, no. 3, pp. 92–103.

[20] M. Dagar and S. Mahajan, "Data Aggregation in Wireless Sensor Network : A Survey," vol. 3, no. 3, pp. 167–174, 2013.

[21] B. Yin and X. Wei, "Communication-Efficient Data Aggregation Tree Construction for Complex Queries in IoT Applications," *IEEE Internet Things J.*, vol. PP, no. c, p. 1, 2018.

[22] H. Yanhua, X. Z.-I. J. of Online, and undefined 2016, "Aggregation Tree Based Data Aggregation Algorithm in Wireless Sensor Networks," *Online-Journals.Org*, pp. 10–15.

[23] S. Sirsikar and S. Anavatti, "Issues of data aggregation methods in Wireless Sensor Network: A survey," *Procedia Comput. Sci.*, vol. 49, no. 1, pp. 194–201, 2015.

[24] A. K. Idrees, W. L. Al-Yaseen, M. A. Taam, and O. Zahwe, "Distributed Data Aggregation based Modified K-means technique for energy conservation in periodic wireless sensor networks," *2018 IEEE Middle East North Africa Commun. Conf. MENACOMM 2018*, pp. 1–6, 2018.

[25] X. Li, S. Liu, F. Wu, S. Kumari, and J. J. P. C. Rodrigues, "Privacy Preserving Data Aggregation Scheme for Mobile Edge Computing Assisted IoT Applications," *IEEE Internet Things J.*, vol. PP, no. c, p. 1, 2018.

[26] A. Varga, "Simulation Manual," *OMNeT++ Simul. Man.*, pp. 3–6, 2016.

Ab Rouf Khan has done his bachelors in Computer Science and engineering from Islamic University of Science and Technology, Awantipora Pulwama, J and K, India. He has completed his M Tech from prestigious Vellore Institute of Technology University, Vellore Tamil Nadu, India. He has both the industry and academic experience. Presently he is pursuing his Ph.D. in the field of Internet of Things from National Institute of Technology, Srinagar, J and K, India

Mohammad Ahsan Chishti has done his Bachelors in Engineering and M.S. in Computer and Information Engineering from International Islamic University Malaysia with specialization in Computer Networking. He has received his Ph.D. from National Institute of Technology Srinagar, India. Presently he is working as Assistant Professor in the Department of Computer Science & Engineering, National Institute of Technology Srinagar, India. He has more than 50 research publications to his credit and 12 patents with two granted International Patents. He is a Young Scientist awardee from Department of Science & Technology, Government of Jammu and Kashmir for the year 2009-2010 and IEI young engineer awardee 2015. He is Senior Member-Institute of IEEE, Member IEI, Life Member CSI, and Member IETE. He is a certified White belt in Six Sigma by Six Sigma Advantage Inc. of USA (SSAI).