An Efficient Information Aggregation Scheme in Internet of Things: Multi Agent based Approach

Sharanappa P. H., Mahabaleshwar S. Kakkasageri

Abstract: ***The use of wireless sensor technology in various Internet of Things (IoT) applications is growing rapidly. With the exponential increase of smart devices and their applications, collecting and analyzing data is gradually becoming one of the most difficult tasks. As sensor nodes are powered by batteries, energy efficiency is essential. To that intention, before passing the final data to the central station, a sensor node should reduce redundancies in the received data from neighbor nodes. There will be some redundancy in the data because different sensor nodes typically notice the same phenomenon. Data aggregation is one of the most important approaches for eliminating data redundancy and improving energy efficiency, as well as extending the life time of wireless sensor networks. Furthermore, the effective data aggregation technique might help to reduce network traffic. In this paper we have proposed cluster based data aggregation using intelligent agents. The performance of the proposed scheme is compared with Centralized Data Aggregation (CDA) mechanism in IoT.***

Keywords: ***Internet of Things (IoT), Information Aggregation, Multi agents.***

I. INTRODUCTION

The Internet of Things is the latest trend in the information technology industry, following the emergence of the Internet and mobile communication. It is expected to be a key feature of the future Internet due to the foresight and development of applications that are compatible with IoT. Despite the fact that it is a widely used term in today's society, its exact definition remains a mystery because it incorporates a vast range of notions. The IoT can be thought of as a diverse network of interconnected things that are individually uniquely identified and addressable. Using various types of sensors and actuators connected to the Internet via various access networks such as Wireless Sensor Network (WSN) and Wireless Mesh Network (WMN), billions of things can be found in the Internet of Things. Heterogeneous objects used in the IoT can transmit information about their identities, status and whereabouts to humans and other connected devices. IoT allows physical items throughout the world to be accessed from anywhere and at any time owing to the ability to connect devices like sensors and actuators to the Internet. WSN plays an active role in the integration of surrounding data in the Internet of Things. It consists of a large number of small devices that are deployed in a sensing region of particular interest in order to collect necessary information from the environment and to take action using an actuator. WSN devices are typically limited in terms of power backup, computing capability, and memory. Furthermore, applications including structural health monitoring, scientific research, environmental monitoring, health care, industrial monitoring, and military tracking have their own set of criteria and aspects. IoT applications necessitate the transmission of massive volumes of data from one network node to another. Because sensor nodes are energy restrictive, it is ineffective for all nodes to send all of the gathered data to the sink node at the same time. Data from nearby sensor nodes is frequently correlated and redundant. Furthermore, the volume of data collected in large sensor networks is sometimes too much for the sink node to handle. As a result, at the intermediate nodes, a technique for integrating this redundant and associated data into genuine high-quality information is required. In terms of power consumption, bandwidth usage, load balancing, network longevity, and data accuracy, this form of data aggregation has been proven to be beneficial. It includes aggregating data from several sensor nodes and sending it to the sink node via intermediate nodes. A good data aggregation strategy can lower energy consumption and network traffic volume while also extending network lifetime and improving data accuracy. In this article, we have proposed a novel scheme to aggregate data efficiently keeping in mind the above mentioned issues. The simulation results shows that the proposed algorithm performs better in terms of aggregation delay, bandwidth utilization etc., when compared with CDA scheme. The remaining part of the paper is organized as follows: Related works are discussed in Section II; The proposed validation approach is discussed in Section III; Section IV and V includes simulation and result analysis as well as a crucial comparison with existing schemes respectively; Conclusions are drawn in Section VI.

II. RELATED WORKS

A review of different forms of data aggregation algorithms and protocols, taking into account the key issues and characteristics of data aggregation in wireless sensor networks are discussed in [1].

Manuscript received on September 01, 2021.
Revised Manuscript received on September 06, 2021.
Manuscript published on September 30, 2021.
* Correspondence Author

Sharanappa P. H.*, Electronics and Communication Engineering Department, Basaveshwar Engineering College (Autonomous), Bagalkot, India. Email: phsharanu@gmail.com

Mahabaleshwar S. Kakkasageri, Electronics and Communication Engineering Department, Basaveshwar Engineering College (Autonomous), Bagalkot, India. Email: mahabalesh_sk@yahoo.co.in

The ultimate goal of this research is to lay the groundwork for developing more advanced designs based on data integration and clustering techniques that have been offered previously. This paper discusses the major strategies for data integration in wireless sensor networks, including ground, subterranean, and underwater sensor networks, as well as the uses, benefits, and drawbacks of each technique. To analyze raw data, various data aggregation techniques necessitate differing amounts of energy. The data aggregation method chosen is determined by the application requirements as well as the relative energy savings gained. There will be some redundancy in the data because different sensor nodes typically notice the same phenomenon. In the meanwhile, many applications use more sensors than are really necessary to accurately detect the desired phenomenon. The significance of data collection is discussed, and different hierarchical clustering algorithms are evaluated [2].

Adding more sensor nodes to the network reduces resource restrictions while increasing data redundancy rates. Data aggregation techniques in sensor networks overcome this constraint. Cluster head nodes are used in data aggregation protocols to gather, aggregate, and transfer data to the base station. Energy, latency, cluster size, and data rate are the most important factors to consider while designing data aggregation algorithms. A novel way of classifying energy-efficient data aggregation protocols using structure, search, and time-based approaches are provided [3]. A comparison of LEACH and LEACH-C strategies is made using the NS-2 simulation tool for a few key selectable parameters, and the simulation results are compared to selected performance indicators. The primary goal of this study is to present a more current perspective on this topic [4]. A hybrid strategy for protecting data privacy is presented in [5]. The proposed algorithm combines the Message Digest algorithm, Elliptic Curve Cryptography, and Advanced Encryption Standard techniques. The destination IoT node shares its geotag with the source node in the hybrid approach.

A detailed study and review of data aggregation mechanisms are presented in [6]. Where tree-based, cluster-based, and centralized data aggregation processes are the three basic types of data aggregation mechanisms. In addition, the extensive comparison of the important techniques in each class provides encouragement for additional research. The tree-based data aggregation algorithm is built with the help of the Lowest Common Ancestor (LCA). Furthermore, the cluster-based data aggregation algorithm with the-dominating set and the centralized data aggregation technique with the SUM() aggregation function are proposed. The algorithms are backed up by well-designed flowcharts that depict the flow and operation of the data gathering processes. The findings were achieved on a system with 60 nodes, and each of the three aggregation procedures was compared to each other [7].

A framework for wireless sensor network aggregation, based on heuristic tree building and realistic sensing priority and decentralized nature assumptions is presented in [8]. And is referred as Semi Distributed Heuristic Energy efficient Aggregation Tree (SDHEAT) technique and is the source independent aggregation protocol . It gives a detailed analysis of the residual energy, longevity, and end-to-end delay based on simulation data. The tree is generated in a semi-distributed way using the best first search approach over the network field. The tree is constructed once when the sensors are deployed and updated when a node fails or when energy is lost at certain aggregation. The work proposed in [9] integrates the Advanced Encryption Standard, Elliptic Curve Cryptography, and Message-Digest algorithms to create a hybrid confidentiality solution. To ensure that all devices are secure, geo encryption, or location-based encryption, is combined with a hybrid technique. For the Internet of Things, the hybrid algorithm offers strong data transfer secrecy.

The work mentioned in [10] presents a comprehensive overview of numerous data aggregation strategies. It describes the basic operation of the algorithms as well as its distinguishing characteristics. The algorithm's performance is then discussed, as well as comparisons to other similar approaches. The distinct data aggregation protocols are categorized and the network flow is described. It also describes Quality of Service (QoS) aware data aggregation methods and the tradeoffs involved. Protocols which addresses security are also discussed. To set child balance among nodes, a distributed technique is proposed [11]. The height of the network graph was enhanced by constraining the degree in this strategy, and network congestion was reduced as a result. In addition, for the Routing Protocol for Low-Power and Lossy Networks, a dynamic data aggregation solution based on Learning Automata was developed (LA-RPL). In order to accomplish data aggregation and transmissions, each node was outfitted with learning automata. Three different forms of data aggregation techniques are compared in [12]. Coding schemes based on relative difference (CS-RD), adaptive data aggregation method (ADAM), and coding schemes based on accuracy factor are some of the strategies used (CS-PF). The algorithms' performance is compared using 15 different scenarios. The following parameters are used to apply the algorithms separately: (i) Mean; (ii) Median; (iii) Mode; (iv) Geometric mean; (v) Harmonic mean. Experiments are done for each circumstances separately for various sensors which record humidity, temperature and light. Energy usage, mean absolute error, and data compression ratio are the performance indicators investigated. A study that presents a comprehensive overview of IoT architectures, applications, and research challenges is provided in [13]. Items and platforms are obligated to release personal or confidential data in this situation. The question of how to achieve a balance between privacy and security in this scenario remains unsolved. Detecting dangerous things is also important during the communication process. It's possible that detecting a fraudulent object will cause a delay. Overview of different data aggregation methodologies in IoT architecture are provided along with a new type of dependable data aggregation algorithm. In fog computing, this new kind of algorithm combines a consensus-based aggregation with fault tolerance methodology. The new method encourages adaptive behavior and enables for more effective transfer of aggregate results to the ascendant node. The proposed technique is fault tolerant and addresses the issue of node reliability [14]. Existing data aggregation and routing methods save energy by minimizing data redundancy, hence reducing bandwidth and memory use. There has been a tradeoff between data aggregation and routing security and energy usage, as higher levels of security need more demanding computational activities, and vice versa.

Enhanced Modified Power-Efficient Gathering (Aggregation) in Sensor Information System (Modified-PEGASIS) method for efficient data aggregation and routing in IoT WSN is presented [15].

Information regarding a single product can be distributed over several data nodes in many industrial applications, and aggregating the data from these nodes has become a frequent task. A distributed service-oriented architecture is presented for this task [16]. Each manufacturer provides service for their own products under this design, and data nodes store the data they acquire. Semantic technologies are used to address heterogeneity issues and act as the foundation for a variety of applications. It is also shown that how this design may be used to tackle the problem of product tracing as an example. The Internet of Things is a fundamental enabler for transforming present urban landscapes into Smart Cities. Making cities smarter has several purposes, one of which is to create a healthy environment that improves inhabitants' quality of life and welfare. A novel data aggregation mechanism that is geared to the application of large-scale air pollution monitoring using IoT devices. Using the paradigm of compressed sensing with side information the design uses source correlations among air-pollution data.

A hybrid Quality of Service-Aware Data Aggregation (QADA) scheme is discussed in [18]. This technique combines the advantages of cluster and tree-based data aggregation schemes while also addressing some of their major drawbacks. In terms of power consumption, network lifetime, and bearing increased traffic loads, simulation findings reveal that QADA surpasses cluster and tree-based aggregation techniques. For edge computing-enabled IoT, a Blockchain-based Secure Data Aggregation technique, called (BSDA), is presented in [19], where the block header is intergraded with a security label including task security level and task completion requirement in order to prohibit task receivers (i.e., Mobile Data Collectors (MDC)) from searching for and accepting tasks. As a result, new block creation rules are being created in order to enhance system throughput and transaction latency. In addition, BSDA divides sensitive jobs and task receivers into categories to protect privacy. The current drive toward fog computing, which moves control, computation, and storage to nodes at the network edge, necessitates data collection at several sinks rather than the single sink normally considered in WSN aggregation techniques [20]. The authors presented mixed-integer programming formulations and methods for the problem of energy-optimal routing and multiple-sink aggregation of sensor measurement data in IoT edge networks, as well as joint aggregation and dissemination. It optimizes the network for both minimum total energy consumption and min-max per-node energy consumption. In the pure aggregation situation, it also gives a formulation and technique for throughput-optimal transmission scheduling using the physical interference model.

Authors in [21] presented a distributed Cross-Layer Commit Protocol (CLCP) for data aggregations, as well as its support for query-based search in IoT applications. This method automatically identifies the best WSN nodes for data aggregation in order to save energy. By default, cluster head selection in CLCP is based on the CL factor, which takes into account two parameters: residual energy and average distance between cluster members. The cluster head of the related cluster is picked from among the cluster members having the highest CL factor. The cluster members are chosen in this technique based on the query response, which defines the target sources node for a specific query generated by the Extract-Transformation-Loading (ETL) registry.

Recursive principal component analysis (R-PCA) is used to provide a cluster-based data analysis methodology that can consolidate redundant data while also detecting outliers [22]. At a cluster head, spatially linked sensor data obtained from cluster members is aggregated using principle components (PCs), and prospective data outliers are identified using the aberrant squared prediction error (SPE) score, which is defined as the square of residual value after PC extraction. The parameters of the PCA model can be recursively changed using R-PCA to react to changes in IoT systems. The computational and processing costs on sensor nodes are also reduced by using a cluster-based data analysis architecture.

The duty cycling technique effectively extends the life of sensor nodes, but it does so at the expense of increased data aggregation time. The subject of minimum time aggregation scheduling in duty-cycled sensor networks is investigated in [23]. A Collision-Resistant Dynamic (CORD) scheduling strategy is presented with the goal of providing fresh data for growing IoT applications. The suggested method works with any starting routing configuration and dynamically changes the recipient of a message when doing so reduces aggregation time. The choice to switch receivers is made on the fly, avoiding any collisions between the determined signals.

The sensors in the IoT are vulnerable to a variety of attacks due to their resource constraints and the inherent unreliability of wireless transmission. As a result, figuring out how to create a data aggregation technique that is both efficient and secure becomes crucial. A trust based secure data aggregation approach is discussed (TBDSA). TBSDA uses a behavior-detected trust evaluation and data assembling technique to enable safe data aggregation [24]. Based on the concept of so-called abstract sensors, authors suggested a method for leveraging data redundancy [25]. This can compensate for the sensors' inadequacies and allow us to switch between different degrees of precision, dependability, and energy consumption as needed during operation. This framework will involve the discovery and maintenance of redundant sensor cliques. It will provide an API that will allow applications to continuously describe their optimization goals, map these goals to optimal parameter settings, and then perform the system reconfiguration that results. The goal of the research work mentioned in [26] is to create an energy-efficient data gathering environment for large-scale, randomly deployed cluster-based wireless sensor networks by using a virtual grid-based mechanism to localize and stabilize cluster sizes. This was done as a prerequisite for the suggested differential data aggregation scheme for spatially correlated data in a cluster to be implemented. Effective data management during transmission and execution is critical to the Fog architecture's effectiveness. To that aim, this study 0examines the benefits of fog computing over cloud computing in terms of managing energy and time restrictions in IoT applications, as well as the issues of data aggregation in fog computing [27]. This work makes a significant contribution by reviewing data aggregation methodologies and recommending a hybrid data aggregation scheme for Fog-based IoT applications.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

The work mentioned in [28] focuses on spotting outliers in the IoT network. Outlier detection is one of the most important uses of machine learning in IoT, and it is critical for cloud and IoT security because data is expected to have patterns. A threat could be predicted by classifying a specific pattern. Existing outlier identification algorithms for sensing systems could serve as a foundation for IoT outlier detection.

Authors in [29] define a new problem called n × 1-out-of-n oblivious transfer and offer a strategy to solve it effectively using contemporary cryptography and concealed permutation. The permutation's computational complexity is O(lg(n)), but the communication overhead is O(nlg(n)). An anonymous communication system can be implemented using the concealed permutation.

Big data authentication with aggregate signatures is a time-saving method. It has the ability to significantly cut computation and communication costs. Using these properties, authors develop VDAS, a verifiable data aggregation technique based on an enhanced certificateless aggregate signature algorithm for the Internet of Things [30]. The number of IoT terminals has no bearing on the length of the aggregated authentication message in VDAS. It is shown that VDAS is existentially unforgeable against adaptive chosen message attacks. The research work mentioned in [31] describe an intelligent technique for IoT data collection based on an agent-based methodology. The main contribution of the proposed research study is to design and develop an intelligent system for IoT data collection.

A. Proposed Work

The proposed aggregation model uses software agents, some of which are static and some are dynamic, namely: Node Static Agent (NSA), Node Mobile Agent (NMA), Cluster Static Agent (CSA) and Cluster Mobile Agent (CMA). NSA & CSA are static and NMA & CMA are mobile agents. In the proposed aggregation scheme, it deals with Node Agency (NA) and the Cluster Agency (CA) that are discussed in detail in the following sections.

The scheme operates in the following sequence:

i) The values that are sensed periodically from the environment are collected by nodes.

ii) On sensing, the NSA triggers the NMA to carry the data to the Cluster Head (CH).

iii) On receiving the data from different nodes from NMA, the CSA aggregates the data.

This aggregated data is to be sent to the sink node/base station or it may be stored in the knowledge base using the CMA.

B. Our Contributions

Our contributions for the proposed scheme includes the following.

1) Design of intelligent scheme for information aggregation.
2) Employing the agent based approach for aggregation by removing the redundant data.
3) Elimination of storing the invalid redundant data.

III. INTELIGENT INFORMATION AGGREGATION

This section provides the details of the network environment for the proposed aggregation scheme and the intelligent agencies for the proposed work.

A. Network Environment

For the proposed system of aggregation, the environment shown in Fig.1 is considered, where number of IoT devices are considered to be sensor nodes, which constitute the sensing layer. These devices have limited resources for computations, and will be able to communicate with their Cluster Heads (CH) based on their scheduling. The CH have the required computational resources for the aggregation of data. The CH upon receiving the data sensed by the sensor nodes will aggregate the data using the agents involved. This aggregated data can be transmitted to the sink node/ base station or it can be stored in the cloud or database. Each device has its own agent platform and agency that contains software agents. Software agents are defined as self-contained pieces of code or programs that run on their respective host's agent platform. These agents use their Knowledge Base to complete a task without interfering with the host's functionality. Agents might be stationary or move about. Static agents, which may or may not be platform dependent, can be generated, used, and destroyed as needed. Mobile agents, on the other hand, are programs that are intended to execute regardless of the platform on which they are designed to run.

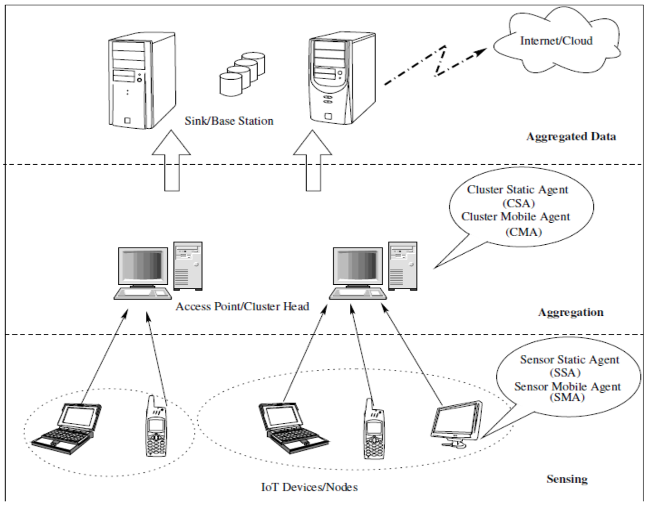
Internet/Cloud
Sink/Base Station
Aggregated Data
Cluster Static Agent (CSA) Cluster Mobile Agent (CMA)
Access Point/Cluster Head
Aggregation
Sensor Static Agent (SSA) Sensor Mobile Agent (SMA)
IoT Devices/Nodes
Sensing

Fig.1 Network Environment

B. Aggregation Mechanisms

From time to time, many data aggregation techniques/mechanisms in IoT have been presented. However, tree-based, cluster-based, centralized, P2P, and distributed mechanisms have all been found to be practical and efficient for IoT. tree-based, cluster-based, and centralized mechanisms are the ones that have received the most attention. To make them better and more efficient, all of these techniques must be thoroughly studied and applied. Source sensor nodes acquire data from various sources and transfer it to the immediate node based on the distance metric in the tree-based data aggregation mechanism. The aggregator is the intermediary node. The aggregation procedure is carried out by an intermediate/hierarchical node. The tree-based data aggregation mechanism is depicted in Fig.2. After aggregation, the aggregator nodes send the data from the source nodes to the sink/collector node.

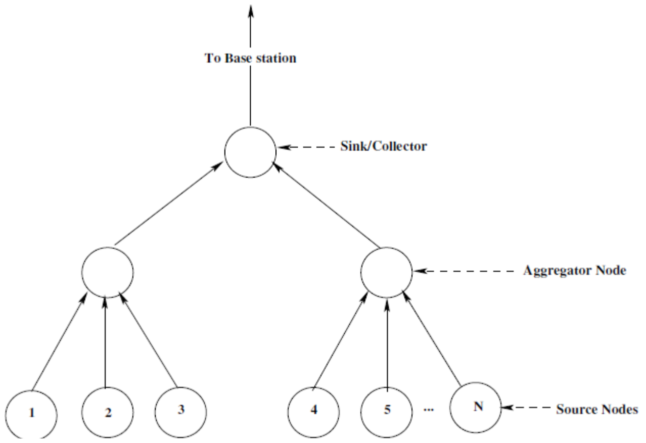

Fig. 2 Tree based aggregation

The cluster-based technique divides the entire network into multiple clusters, each of which has a large number of sensor nodes. Based on an election mechanism, one node – the header/leader node is chosen from each cluster and becomes the cluster-head. This method reduces bandwidth overhead while also allowing for the transmission of fewer packets. Fig. 3 depicts a cluster-based data aggregation architecture.

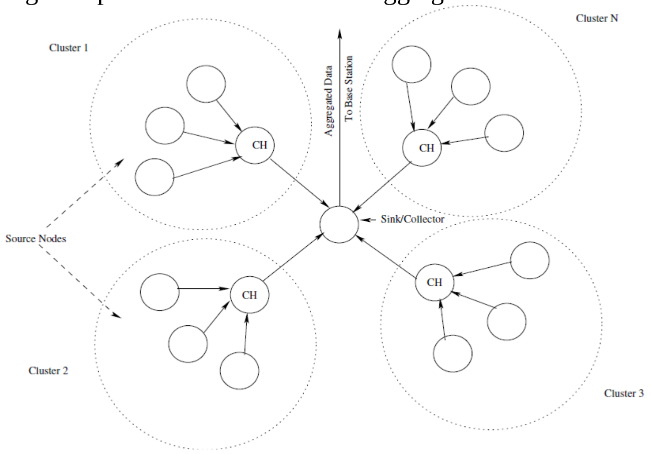

Fig. 3 Cluster based aggregation

Pre-defining a k-number of clusters can be used to build the cluster-based data aggregation process in IoT. After receiving the cluster heads for each cluster, the source nodes for that cluster deliver the data to the cluster heads for that cluster. Each cluster head's aggregated data is relayed to the base station using a simple mathematical summing formula that runs from 1 to the number of clusters (k). In some cases, the Sink/Collector might operate as the base station itself to improve the effectiveness of the mechanism, depending on the application. Even when the network is enormous, the cluster-based data aggregation mechanism scales well since cluster creation ensures network scalability. As a result, a cluster-based data aggregation mechanism could be a solution for aggregating data in large-scale IoT-based architectures that deal with big data.

Each source/immediate node transmits data to the centralized header node through the least expensive path in the centralized mechanism. This technique's data aggregation is handled by the header node, which receives data from several sensor nodes. The header node outperforms the other nodes in terms of processing power. The data aggregate is sent to the base station from the header node. Fig. 4 depicts a centralized data aggregation architecture. The data from the aggregators is combined using the SQL aggregator function SUM(), and the resulting data is sent to the base station. Because enormous volumes of data are stored on a single node, centralised aggregation has serious scalability issues. However, because there is only one Leader node, it is simple to manage the network's many metrics such as route backup, route discovery, and so on.

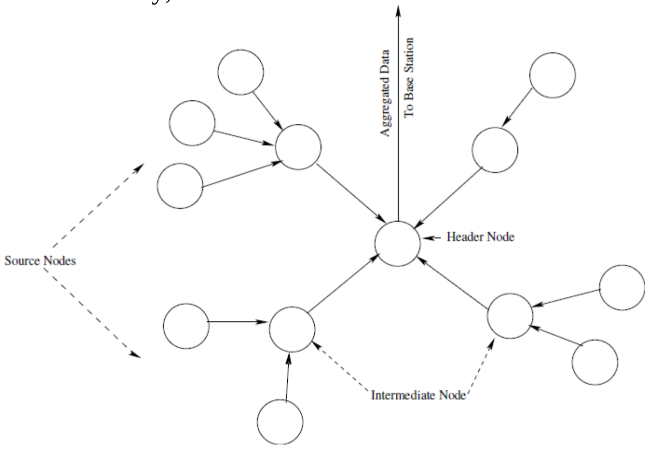

Fig. 4 Centralized aggregation

C. Proposed Aggregation Scheme

For the proposed aggregation scheme, we use the combination of cluster-based model and tree based model. It converts raw data into legitimate, high-quality information, reducing redundancy in data transmission and thus increasing energy efficiency and network lifetime. The proposed scheme works in four stages: (i) Selection of Cluster Head (CH) (ii) formation of cluster (iii) formation of tree and (iv) aggregation of data.

i) Selection of Cluster Head (CH): The sink node coordinates the CH selection process. Each node delivers its residual energy and position information to the sink node at first. The sink node evaluates the average residual energy and chooses a CH node with a residual energy larger than or equal to the average residual energy and approximately more neighbor nodes than the others. To disseminate CH information, the sink node broadcasts an advertisement message (ADV) via the CSMA MAC protocol. This message includes the CH node ID as well as a field that indicates that this is an announcement message. Each sensor node verifies their ID to the one received after getting the ADV message. The node with the same ID will serve as the CH node. For the selection of the CH, the system is modeled as shown below. Let the network contains a set of sensor nodes represented as $S = \{S_1, S_2, S_3, \ldots, S_n\}$, where n is the total number of nodes in the network that sends data to the sink node inside the network. We assume that all of the sensor nodes/IoT devices in the network are static and distributed at random. The sink node receives residual energy and position information from each node. Sink determines average energy (E_{avg})based on this by using (1) as shown below.

$$E_{avg} = \frac{\left(E_{S_1} + E_{S_2} + E_{S_3} + \ldots\ldots + E_{S_n}\right)}{n} \qquad (1)$$

where E_{S1}, E_{S2},...E_{Sn} are the residual energy of the sensor nodes $S_1, S_2, \ldots, S_n$.

Let (x_1, y_1), (x_2, y_2),...(x_n, y_n) be the location coordinates of the nodes *1, 2, ...n*. Then the shortest distance (D) between the two nodes can be determined by the distance formula as shown in (2).

$$D=\sqrt{\left(x_2-x_1\right)^2+\left(y_2-y_1\right)^2} \quad (2)$$

Sink node will select S_1 as CH if the residual energy of node *1* is greater than the calculated average energy. i.e., when $E_{S1} \geq E_{avg}$ and also it should be near to the sink node ($D \leq D_{min}$), where D_{min} is the minimum distance.

ii) Formation of Cluster: Sink node will select S_1 as CH if the residual energy of node *1* is greater than the calculated average energy. i.e., when $E_{S1} \geq E_{avg}$ and also it should be near to the sink node ($D \leq D_{min}$), where D_{min} is the minimum distance.

iii) Formation of tree: The sink node initiates the development of the logical tree after CH selection and cluster formation. The position and current residual energy of CH nodes are used to form this tree. Initially, the sink node sends a control message to all CH nodes. The ID, parent, power, status, and level fields in this control message indicate the CH's ID, parent in the aggregation tree, current residual power, logical tree status (leaf node, relay node, or danger state), and path length (number of hops from the sink), respectively. Assuming the sink node has infinite energy supply and is the root node of the aggregation tree, the control message for it is $ctrlmsg(ID_{sink}, -, \infty, status_{sink}, level_0)$. The parent node with the highest residual power and the shortest path to the sink is recorded by the CH1 (assumed). CH1 now sends out the message $ctrlmsg(ID_{CH1}, parent_{CH1}, power_{CH1}, status_{CH1}, level_{CH1})$, where $level_{CH1}$ == 1 + $level_0$. This process continues until each CH sends out a single control message. The end result is a CH node aggregation tree with a sink at the root node. Sink nodes can regularly re-construct the aggregate tree based on the CH nodes' remaining power. The TDMA schedule for each CH node is also included in this data. The CH nodes that don't have any data to send turn off their radios to save energy

iv) Aggregation of data: According to the suggested structure, sample data is sent to the cluster head via sensor nodes. Because numerous IoT sensors are integrated on a single sensor node, the data matrix provided by a given sensor node *i* is most likely multivariate, as shown in (3).

$$Y_i=\begin{bmatrix} y_{i,1}(1) & y_{i,1}(2) & y_{i,1}(3) & \dots & y_{i,1}(m) \\ y_{i,2}(1) & y_{i,2}(2) & y_{i,2}(3) & \dots & y_{i,2}(m) \\ y_{i,3}(1) & y_{i,3}(2) & y_{i,3}(3) & \dots & y_{i,3}(m) \\ . & . & . & . & . \\ . & . & . & . & . \\ y_{i,l}(1) & y_{i,l}(2) & y_{i,l}(3) & \dots & y_{i,l}(m) \end{bmatrix} \quad (3)$$

where *m* is the number of samples from a particular sensor, *l* is the number of physical attributes, e.g. temperature, humidity, light, etc. The sensor data which gets generated from a node *i* at a specific instant of time t_m is given by the (4) below.

$$Y_i(t_m)=\left[y_{i,1}(t_m), y_{i,2}(t_m), \dots\dots, y_{i,l}(t_m)\right]^T \quad (4)$$

Let the number of nodes in a cluster be *k*, then the data collected at any CH is denoted by $\{Y_1, Y_2, ..., Y_k\}$, because of the high correlation between the neighboring nodes data and their physical property, the sensor data is reorganized by the CH as $\left[\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_l\right]$. The matrix $\hat{Y}_{i,j\in[1,l]}$ gets transformed in to the equation shown in (5)

$$\hat{Y}_j=\begin{bmatrix} \hat{y}_{1,j}(1) & \hat{y}_{1,j}(2) & \hat{y}_{1,j}(3) & \dots & \hat{y}_{1,j}(m) \\ \hat{y}_{2,j}(1) & \hat{y}_{2,j}(2) & \hat{y}_{2,j}(3) & \dots & \hat{y}_{2,j}(m) \\ \hat{y}_{3,j}(1) & \hat{y}_{3,j}(2) & \hat{y}_{3,j}(3) & \dots & \hat{y}_{3,j}(m) \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{y}_{k,j}(1) & \hat{y}_{k,j}(2) & \hat{y}_{k,j}(3) & \dots & \hat{y}_{k,j}(m) \end{bmatrix} \quad (5)$$

D. Agent Model

The agent model proposed here for information aggregation contains two agencies referred as the Node Agency and the Cluster Agency. Both agencies uses mobile and static agents and contains their respective Knowledge Base(KB).

i) Node Agency: The agent model proposed here for information aggregation contains two agencies referred as the Node Agency and the Cluster Agency. Both agencies uses mobile and static agents and contains their respective Knowledge Base(KB).

- *Node Knowledge Base (NKB):* Node Knowledge Base consists of information regarding device or node identity (ID) and IDs of the CH and neighbours, residual energy, bandwidth available for communication, parent and child node information, present and past information gathered, status of the node (live or dead), etc. Knowledge base is updated and read by NSA and NMA.
- *Node Static Agent (NSA):* The NSA is a static agent that runs on the device/source node. As soon as the end nodes senses the data from the physical environment, the NSA invokes the mobile agent NMA.
- *Node Mobile Agent (NMA):* On getting invoked by the NSA, the NMA carry the sensed data to the respective CH in the network. It also updates the Node Knowledge Base.

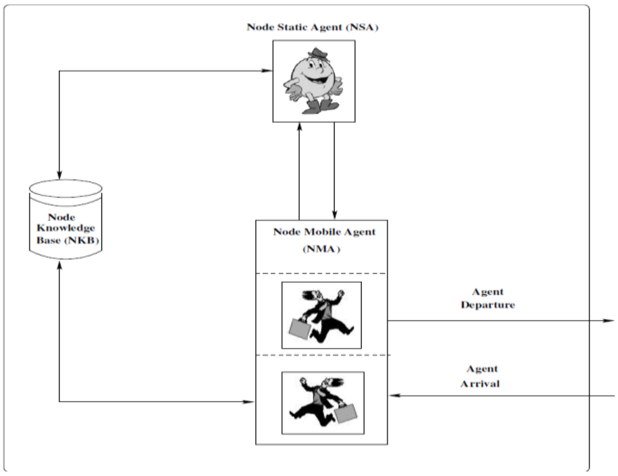

Fig. 5 Node Agency

ii) Cluster Agency: The Cluster Agency includes the Cluster Knowledge Base (CKB) that contains the aggregated data, Cluster Static Agent (CSA) and the Cluster Mobile Agent (CMA). The cluster agency is shown in Fig. (6).

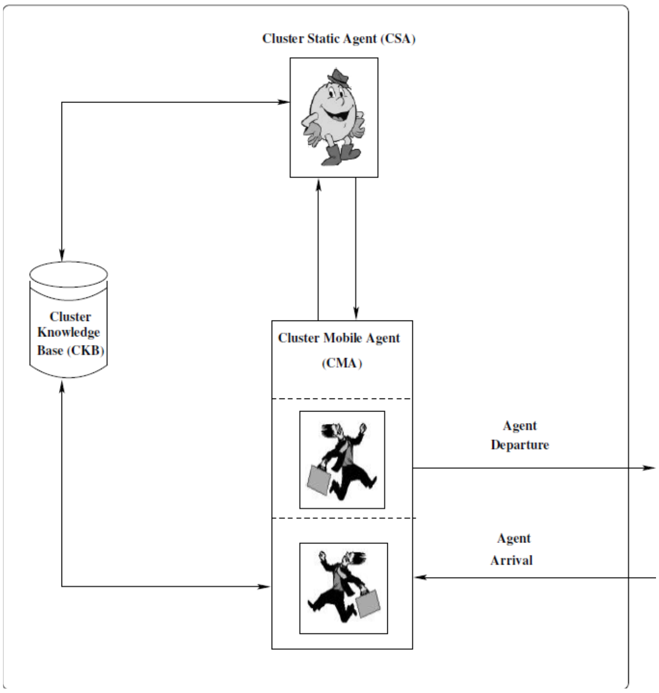
Cluster Static Agent (CSA)
Cluster Knowledge Base (CKB)
Cluster Mobile Agent (CMA)
Agent Departure
Agent Arrival

Fig. 6 Cluster Agency

- *Cluster Knowledge Base (CKB):* Cluster Knowledge Base consists of information regarding device or node identity (ID) and IDs of the CH, parent and member (child) nodes, residual energy, available bandwidth for communication, current and previous information aggregated, status of the node (live or dead), etc. Cluster Knowledge base is updated and read by CSA and CMA.
- *Cluster Static Agent (CSA):* The CSA is a static agent that runs on the header nodes (CH)/ aggregator nodes. It receives the raw data from NMA in the node agency. The job of the CSA is to aggregate the data avoiding the redundant data and triggers the its agent mate CMA.
- *Cluster Mobile Agent (CMA):* On getting triggered by the CSA, the CMA carry the sensed data to the respective sink nodes or base station in the network. It also updates the Cluster Knowledge Base.

IV. SIMULATION AND PERFORMANCE PARAMETERS

The recommended agent based aggregation scheme has been simulated using C++ programming language as discrete event simulator. In this section, simulation inputs and performance parameters are discussed.

A. Simulation Inputs

For the proposed scheme of aggregation, the inputs considered for simulation purpose are shown in Table 1.

Table 1: Simulation Inputs

Sl. No.	Input parameters	Specifications
1	No. of nodes : N	25-150
2	Sensing area: S	500m×500m
3	Sensing range: SR	25m, 50m
4	Data packet size (PS)	48kb

B. Performance Parameters

To evaluate the performance of the proposed research work we have considered the following parameters.

- Information acquisition delay: It's the time that the node mobile agent takes to acquire the information in the network. It is measured in milliseconds (ms).
- Cluster head identification delay: It is the time taken for identifying the cluster head in a given cluster and is expressed in milliseconds (ms).
- Information aggregation delay: It is defined as the time taken by the cluster static agent to aggregate the data avoiding the redundant one, and is expressed in milliseconds (ms).
- Aggregation processing delay: It is the average time the agency takes to aggregate the data. It is measured in milliseconds (ms).
- End to end delay: It is defined as the total time taken for the agency to sense, gather and aggregate the information and is expressed in milliseconds (ms).
- Packet Delivery Ratio (PDR): PDR is defined as the ratio of the number of packets delivered successfully to the number of packets that are actually sent and is expressed in percentage (%).
- Bandwidth utilized: Bandwidth utilization is nothing but the ratio of the bandwidth that is being used for the transmission to the total bandwidth actually available and is expressed in percentage (%).

V. RESULT ANALYSIS

To evaluate and analyse the performance of the scheme proposed, the performance metrics mentioned above are considered and are plotted across the graph for varying number of nodes and Sensing Range(SR). The performance of the proposed scheme is compared with centralized data aggregation scheme (CDA).

Fig. 7 shows that the time taken for acquiring the data increases with increase in number of nodes, however the proposed work takes less time as compared with the existing CDA mechanism.

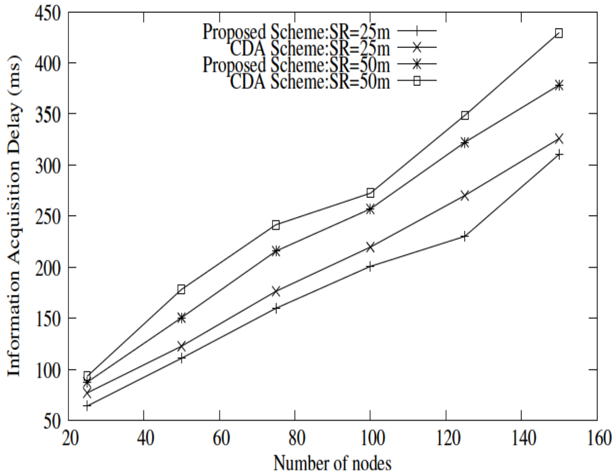

Fig. 7 Information acquisition delay Vs. Number of nodes

Time taken for identifying the cluster head in a given cluster network is plotted against varying number of nodes in figure 8, the proposed work takes less time for the same when compared with the work under comparison.

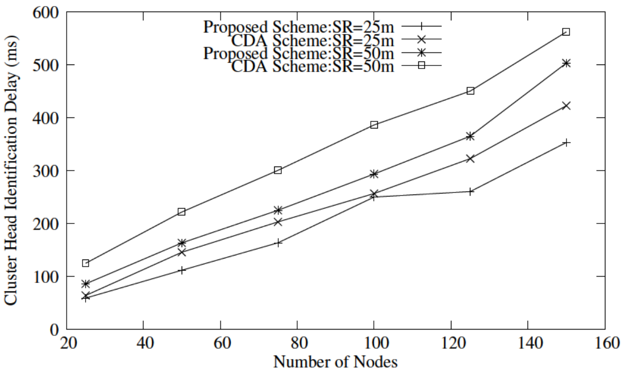

Fig. 8 Cluster head identification delay Vs. Number of nodes

Fig. 9 depicts that the aggregation delay increases with increase in the number of nodes, but still the time taken for the proposed work is lesser than the existing CDA scheme.

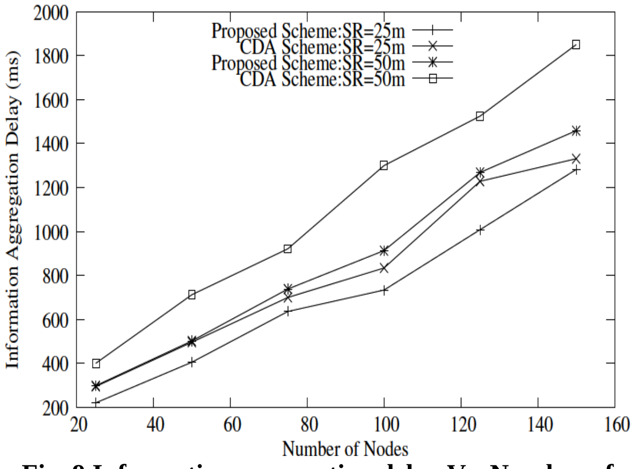

Fig. 9 Information aggregation delay Vs. Number of nodes

The average time taken for the agency to aggregate the data for the proposed scheme is found to be lesser than that of the CDA scheme mentioned as shown in Fig. 10.

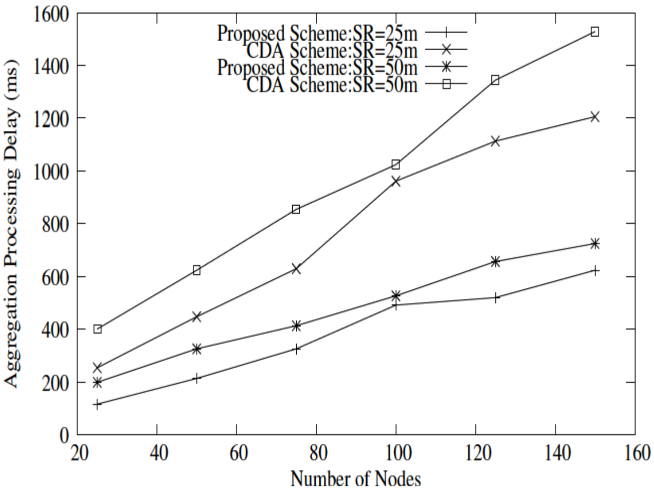

Fig. 10 Aggregation Processing delay vs. Number of nodes

Even though the end to end delay increases with increase in SR and number of nodes but it is evident from the Fig. 11 that the proposed scheme performs better than the work under comparison since it has lesser end to end delay.

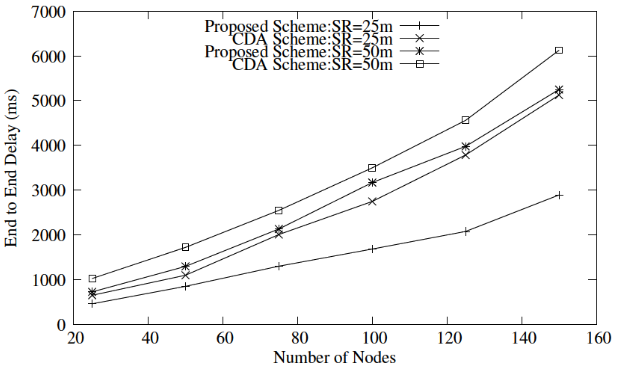

Fig. 11 End to end delay vs. Number of nodes

Delivering the packets successfully is a function of number of nodes. As the number of nodes increases in the network it is difficult for the network deliver all packets that are generated. So, PDR decreases with increase in number of nodes. However the proposed scheme has a better PDR when compared to CDA scheme as shown in Fig. 12.

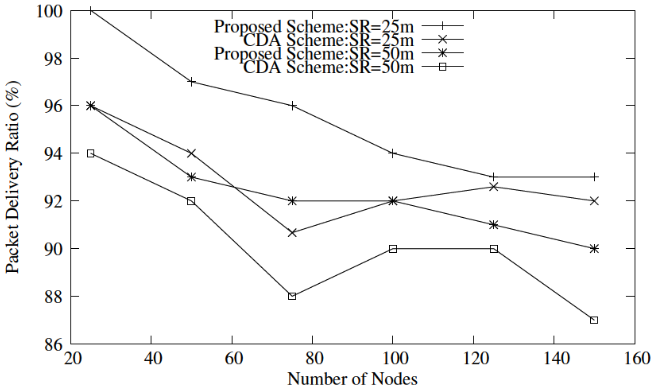

Fig. 12 Packet delivery ratio vs. Number of nodes

The bandwidth requirement will increase if the number of sensor nodes increases. The amount of the bandwidth utilized is depicted in Fig. 13 and it is evident that the proposed work has lesser bandwidth requirement than the work under comparison.

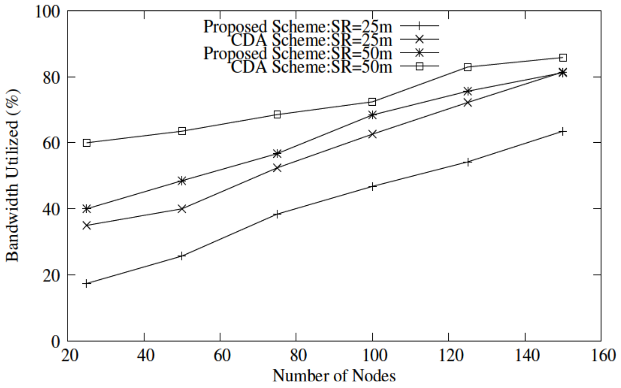

Fig. 13 Bandwidth utilized vs. Number of nodes

VI. CONCLUSION

In this research work, we have proposed an agent based intelligent data aggregation scheme for aggregating the huge data that gets generated across the IoT network.

It appears from the results that the proposed algorithm is more adaptive, can be implemented easily for real-time applications. Further the proposed algorithm is compared with the centralized data aggregation scheme and found to be more efficient in terms of time taken to acquire the data, time required to aggregate the data, bandwidth utilization and end to end delay.

REFERENCES

1. Abbasian Dehkordi S., Farajzadeh K., Rezazadeh J., "A survey on data aggregation techniques in IoT sensor networks". Wireless Networks, Vol. 26, pp. 1243–1263, 2020.
2. Geetika Dhand, S.S. Tyagi, "Data Aggregation Techniques in WSN:Survey", Procedia Computer Science, Vol. 92, pp. 378-384, 2016.
3. M. Bala Krishna and N. Vashishta, "Energy efficient data aggregation techniques in wireless sensor networks," 2013 5th International Conference and Computational Intelligence and Communication Networks, pp. 160-165, 2013.
4. H. Rahman, N. Ahmed and I. Hussain, "Comparison of data aggregation techniques in Internet of Things (IoT)," Proc. International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1296-1300, 2016.
5. Poornima M. Chanal, Mahabaleshwar S. Kakkasageri, "Preserving Data Confidentiality in Internet of Things", Journal of SN Computer Science, Springer, vol. 2, No. 1, pp. 1-12, 2021.
6. Behrouz Pourghebleh, Nima Jafari Navimipour, " 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, Elsevier, Vol. 97 pp. 23–34, 2017.
7. Ab Rouf Khan, Mohammad Ahsan Chishti, "Data Aggregation Mechanisms in the Internet of Things: A Study, Qualitative and Quantitative Analysis", Int. J. Com. Dig. Sys. Vol. 9, No.2, pp. 289-297, 2020.
8. S. M. Al-Tabbakh, "Novel technique for data aggregation in wireless sensor networks", Proc. International Conference on Internet of Things, Embedded Systems and Communications (IINTEC), pp. 1-8, 2017.
9. Poornima M Chanal, Mahabaleshwar S Kakkasageri, "Hybrid Algorithm for Data Confidentiality In Internet of Things", proc. of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IIT Kanpur, 2019, India.
10. Ramesh Rajagopalan, Pramod K. Varshney, "Data aggregation techniques in sensor networks: A survey", pp. 1-30, 2006.
11. Homaei M. H., Salwana E., Shamshirband S., "An Enhanced Distributed Data Aggregation Method in the Internet of Things", Sensors (Basel, Switzerland) Vol. 19, pp. 1-26. 2019.
12. Walaa Hussein, Jiwa Abdullah, N. A. M. Alduais, " Data Aggregation Algorithms with Multiple Sensors in Clustered-Based WSN/IoT", Int. J. Com. Dig. Sys. Vol. 9, No.3, 2020.
13. Poornima M. Chanal, Mahabaleshwar S. Kakkasageri, "Security and Privacy in IoT: A Survey", Journal of Wireless Personal Communication, Springer, Vol. 115, No. 3, pp.1667–1693, 2020.
14. F. Al-Doghman, Z. Chaczko and J. Jiang, "A Review of Aggregation Algorithms for the Internet of Things", 25th International Conference on Systems Engineering (ICSEng), pp. 480-487, 2017.
15. N. Chandnani and C. N. Khairnar, "Efficient Data Aggregation and Routing Algorithm for IoT Wireless Sensor Networks", Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN), pp. 1-7, 2019.
16. Tao Zhu, Sahraoui Dhelim, Zhihao Zhou, Shunkun Yang, Huansheng Ning, "An architecture for aggregating information from distributed data nodes for industrial internet of things", Computers & Electrical Engineering, Vol. 58, pp. 337-349, 2017.
17. E. Zimos, J. F. C. Mota, M. R. D. Rodrigues and N. Deligiannis, "Internet-of-Things data aggregation using compressed sensing with side information", 23rd International Conference on Telecommunications (ICT), pp. 1-5, 2016.
18. H. Rahman, N. Ahmed and M. I. Hussain, "A hybrid data aggregation scheme for provisioning Quality of Service (QoS) in Internet of Things (IoT)", Cloudification of the Internet of Things (CIoT), pp. 1-5, 2016.
19. X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, M. S. Hossain, "A Secure Data Aggregation Strategy in Edge Computing and Blockchain empowered Internet of Things", IEEE Internet of Things Journal, doi: 10.1109/JIOT.2020.3023588.
20. E. Fitzgerald, M. Pioro, A. Tomaszwski, "Energy-Optimal Data Aggregation and Dissemination for the Internet of Things", IEEE Internet of Things Journal, Vol. 5, No. 2, pp. 955-969, 2018.
21. A. Alkhamisi, M. S. H. Nazmudeen, S. M. Buhari, "A cross-layer framework for sensor data aggregation for IoT applications in smart cities", 2016 IEEE International Smart Cities Conference (ISC2), pp. 1-6, 2016.
22. T. Yu, X. Wang and A. Shami, "Recursive Principal Component Analysis-Based Data Outlier Detection and Sensor Data Aggregation in IoT Systems", IEEE Internet of Things Journal, Vol. 4, No. 6, pp. 2207-2216, 2017.
23. T. D. Nguyen, D. T. Le, V. V. Vo, M. Kim, H. Choo, "Fast Sensory Data Aggregation in IoT Networks: Collision-Resistant Dynamic Approach", IEEE Internet of Things Journal, Vol. 8, No. 2, pp. 766-777, 2021.
24. Yanbing Liu, Xuehong Gong and Congcong Xing, "A novel trust-based secure data aggregation for Internet of Things", 9th International Conference on Computer Science & Education, pp. 435-439, 2014.
25. S. Schmeißer and G. Schiele, "Adaptive Aggregation of Redundant Sensor Data in the Internet of Things", International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1387-1388, 2016.
26. R. N. Enam, "Energy efficient differential data aggregation in a dynamic cluster based WSN," 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 580-583, 2013.
27. M. Shahzad, J. Panneerselvam, L. Liu and X. Zhai, "Data Aggregation Challenges in Fog Computing", IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, pp. 1717-1721, 2019.
28. Sampath Kumar Y. R. and H. N. Champa, "IoT Streaming Data Outlier Detection and Sensor Data Aggregation", Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 150-155, 2020.
29. R. Li, C. Sturtivant, J. Yu, X. Cheng, "A Novel Secure and Efficient Data Aggregation Scheme for IoT", IEEE Internet of Things Journal, Vol. 6, No. 2, pp. 1551-1560, 2019.
30. J. Liu, J. Han, L. Wu, R. Sun and X. Du, "VDAS: Verifiable data aggregation scheme for Internet of Things," Proc. IEEE International Conference on Communications (ICC), pp. 1-6, 2017.
31. Sharanappa P. H., Mahabaleshwar S. Kakkasageri, "Intelligent Information Gathering Scheme in Internet of Things (IoT)", Proc. 11th International Conference on Advanced Computing (ICoAC), Department of Computer Technology, Anna University, MIT Campus, Chennai, India, pp.136-140, 2019.

AUTHORS PROFILE

Sharanappa P. H. received his B.E. Degree in Electronics and Communication Engineering, M.Tech Degree in Digital Electronics and Communication from Visvesvaraya Technological University Belgaum, Karnataka, India. He is pursuing his Ph.D Degree in Internet of Things. He has 13 years of experience in teaching. Presently, he is working as Assistant Professor in Department of Electronics and Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India. He has published 3 papers in international conferences and 3 papers in international journals. His areas of interest are wireless networks and Internet of Things. He is a member of IETE India.

Dr. Mahabaleshwar S. Kakkasageri received his B.E. Degree from Karnataka University, M.Tech. degree in Digital Communication and Ph.D. degree from the Visvesvaraya Technological University, Belgaum, Karnataka, India. He has experience of 16 years in teaching. His research interests are Vehicular Ad hoc Networks, Wireless Networks, and Internet of Things. He has published 50 papers in national and international conferences, 25 papers in national and international journals, and 07 books/books chapters. He is a member of IEEE and IETE. He is a reviewer and programme committee member for many journals and international conferences, respectively. He received "Seed Money to Young Scientist for Research" from VGST Karnataka in 2015.

Retrieval Number: 100.1/ijsce A35250911121
DOI: 10.35940/ijsce.A3525.0911121
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)