Bilad Alrafidain Journal for Engineering Science and Technology ISSN: 2073-9524
https://doi.org/10.56990/bajest/2025.040108 Pages:85-99

A Review of Data Aggregation Algorithms for Enhancing IoT Performance

Marwa Ail Kadhim Ali
Department of Information Technology, College of Engineering, Islamic Azad University, Isfahan, 84166, Iran.
marwa86modee@gmail.com

Abstract

With the exponential growth of the Internet of Things, efficient data aggregation is necessary to tackle the issues of network congestion, energy consumption, and latency. This paper presents an adaptive data aggregation algorithm that applies clustering, in-network processing, and optimized routing. The proposed algorithm outperforms traditional methods such as LEACH and PEGASIS with 50 ms latency, 0.5 J/node energy efficiency, and 350 kbps throughput. Key features include dynamic clustering, energy optimization, and real-time adaptation, which offer high scalability and are fit for use in smart cities, healthcare, and industrial IoT. However, future challenges like interoperability and security will still be taken into account on the ground. This paper advances scalable, efficient, and secure IoT systems towards technological advancement.

Keywords: Internet of Things (IoT), Data aggregation algorithm, Network congestion, Energy efficiency, Scalability, Latency optimization, Edge computing, Smart cities applications

Article history: Received: *17 Dec 2024*, ***Accepted: 12 Jan*** *2025*, ***Published:*** *15 Mar 2025*

1. Introduction

The Internet of Things (IoT) has made the interaction between physical and digital systems, with a very large amount of data generated by applications like smart homes, healthcare, and industrial automation. However, the rapid growth of IoT devices creates network congestion, energy consumption, and inefficient processing, so innovative solutions are required to optimize performance and maintain scalability and reliability [1]. Data aggregation helps alleviate these problems through the reduction of data volume, saving energy, and improving network efficiency. It removes redundancy, optimizes bandwidth, and extends the life of resource-constrained IoT devices while enhancing the processing speed of data critical for time-sensitive applications. However, existing algorithms face challenges in balancing energy efficiency, data accuracy, and computational complexity [2].

The work described in this paper targets designing advanced data aggregation algorithms to improve the performance of IoT-based systems, addressing such important challenges as energy efficiency, latency, and integrity issues. The proposed algorithms are applied to various IoT scenarios to establish versatility, and the solutions promote scalable and secure data aggregation for next-generation IoT [3]. However, the key contributions of this work include:

1. Development of a novel data aggregation algorithm integrating adaptive clustering, in-network processing, and optimized routing for enhanced performance.
2. Introduction of machine learning techniques for real-time adaptation, improving energy efficiency and reducing latency.
3. Comprehensive evaluation of the algorithm's performance in simulated IoT scenarios using NS-3 and MATLAB, with metrics such as energy consumption, throughput, and scalability.

4. Comparison with existing algorithms, demonstrating superior efficiency and applicability to real-world IoT environments.
5. Exploration of the algorithm's potential integration with emerging technologies like edge computing and AI/ML, enhancing its adaptability and scalability.

2. Literature Review

Usually, the architecture of IoT can be divided into three layers, which are the perception layer, the network layer, and the application layer. In this sense, the perception layer usually encompasses sensors and actuators in relation to sensing the physical environment, and the network layer ensures that there is data transmission via protocols such as MQTT, CoAP, and HTTP.

It processes all data and generates actionable insights regarding smart cities, health, agriculture, and other such concepts [4]. However, there are many challenges for IoT scalability as millions of devices are getting connected, thus much data. The security and privacy concerns associated with IoT devices include risks of data breaches and denial-of-service attacks. The other significant challenge is energy efficiency, especially for remote resource-constrained devices. Other challenges are network congestion, latency, and limited interoperability among the diverse devices. Such challenges require the development of more advanced solutions, such as data aggregation algorithms, for optimized IoT performance [5].

Data aggregation is one of the critical elements in IoT networks. It reduces data traffic and increases network efficiency. Data compression, filtering, and clustering are common techniques that minimize redundancy without sacrificing the accuracy of the data. However, the tree-based and cluster-based methods, which are prevalent algorithms for data aggregation, face significant scalability, energy efficiency, and integrity challenges in dynamic environments [6].

Each of the data aggregation algorithms has its own strengths and limitations in optimizing IoT network performance. Tree-based algorithms include the Spanning Tree Protocol (STP) and Aggregation Tree, both of which organize IoT networks hierarchically to optimize data transfer [1]. A smart agriculture case study illustrated the way Data Aggregation Tree Protocol reduces data redundancy and, hence, transmission costs but also brings to light scalability and fault tolerance issues where the failure of a node brings an entire data flow to a standstill. Cluster-based algorithms are used to improve energy efficiency, such as LEACH, which is Low-Energy Adaptive Clustering Hierarchy. In smart healthcare, LEACH extended the lifespan of wearable devices by dynamically selecting cluster heads, although issues arose with energy imbalance due to the cluster head selection process in large-scale networks [7].

These are aggregation techniques used at in-network nodes, which aggregate data at intermediate nodes, thereby reducing communication overhead. One case study on environmental monitoring showed that TAG effectively reduces the volume of data to be communicated but has challenges of data integrity and dynamic topologies [8].

Compression-based methods include PCA (Principal Component Analysis) that has been adopted, for example, in urban traffic monitoring, which, with PCA, reduces data size without loss of accuracy, though they were quite troublesome for low-power IoT devices when computing overhead was involved. Most recently, ML-based algorithms such as reinforcement learning have come up with dynamic aggregation adjustments as a function of network conditions but require a lot of computing power and, thus, are confined to resource-constrained environments [1].

These case studies show that though traditional approaches such as tree-based and cluster-based algorithms are energy-efficient, they lack scalability and robustness. Modern machine learning-based approaches are adaptive but require more computational power, and hence, future algorithm development should focus on overcoming these drawbacks [9]. Several limitations in existing data aggregation algorithms in IoT exist that hinder their effectiveness in improving network performance. Tree-based algorithms, though efficient for hierarchical structures, fail to scale up as IoT networks grow [10].

As the number of nodes increases, the maintenance and update of the tree becomes resource-intensive, and failures in higher-level nodes can affect the entire process. Cluster-based algorithms, such as LEACH, suffer from cluster head overload, which creates bottlenecks and uneven energy distribution, which reduces network lifespan. In addition, these methods suffer from topology changes, since dynamic networks with mobile or failing nodes interfere with cluster formation [3].

In-network aggregation algorithms face problems with data integrity, especially in cases of packet loss or unreliable links. Besides, these algorithms are not adaptive to dynamic topologies and need more computational resources, thus becoming difficult for resource-constrained IoT devices [11]. Compression-based aggregation methods like PCA reduce the volume of data but are associated with very high computational overhead and also tend to lose data, hence not so accurate. It has poor scalability in a large IoT network. The algorithms that are based on machine learning, such as reinforcement learning, require enormous amounts of computational resources as well as training data. In most IoT applications, resource constraints and slow adaptation to change are issues [12].

3. Research Methodology

3.1 Research Design and Approaches

This research develops and evaluates advanced data aggregation algorithms that improve the performance of an IoT network. A systematic methodology combines theoretical development with simulation-based validation [13]. It starts with an overview of IoT architectures and current aggregation methods, pointing out some limitations. Proposed algorithms address such issues as energy consumption, latency, and scalability with adaptive clustering, in-network processing, and data compression techniques. Machine learning methods enable real-time adaptation. Algorithms are designed for resource-constrained devices and tested in simulated IoT environments using tools such as NS-3 and MATLAB. Performance is evaluated based on latency, energy efficiency, throughput, and data accuracy [14]. Fig. 1 displayed the Block Diagram for Research Methodology

3.2 Proposed Data Aggregation Algorithms

The proposed data aggregation algorithm addresses the key challenges of IoT, which include energy efficiency, latency reduction, and scalability. It uses a hybrid approach, combining adaptive clustering, in-network processing, and data compression. Sensor nodes are dynamically grouped into clusters, and cluster heads are elected based on their energy levels and processing power. Cluster heads aggregate the data, communicate with other clusters, and send their information to the central station. Reinforcement learning and other machine learning models help in the real-time adaptation of parameters based on network conditions [15].

These will include adaptive clustering for balancing energy usage, overhead minimization from data transmissions, and scaling features. Lighter encryption provides security for data communication while dynamic cluster head reconfiguration and re-assignment mechanisms in case of faults are assured. It will be evaluated by simulating the same on different IoT scenarios using NS-3 and MATLAB tools with performance metrics concerning energy efficiency, latency, and data accuracy and scalability relative to algorithms such as LEACH and TAG [16]. The IoT environment is selected in this study for the following reasons:

IoT devices have substantial computational and communication power for aggregation tasks. The network topology is stable with slight interruptions. IoT devices are heterogeneous, generating diverse data requiring consistent aggregation. The clusters are highly sustainable. Cluster heads manage efficient data aggregation. To obtain reliable data exchange, Standard communication protocols like MQTT and CoAP are used [17]. Whereas the Constraints are summarized as the following:

Resource constraints of battery lifetime and computational resource. Dynamic network environments may have an impact on data communication. Latency sensitivity in important applications like healthcare. Security weakness demands lightweight cryptography. Scalability issues in the case of increasing devices [18]. Energy-efficient Clustering along with Adaptive tuning are proposed to overcome such constraints.

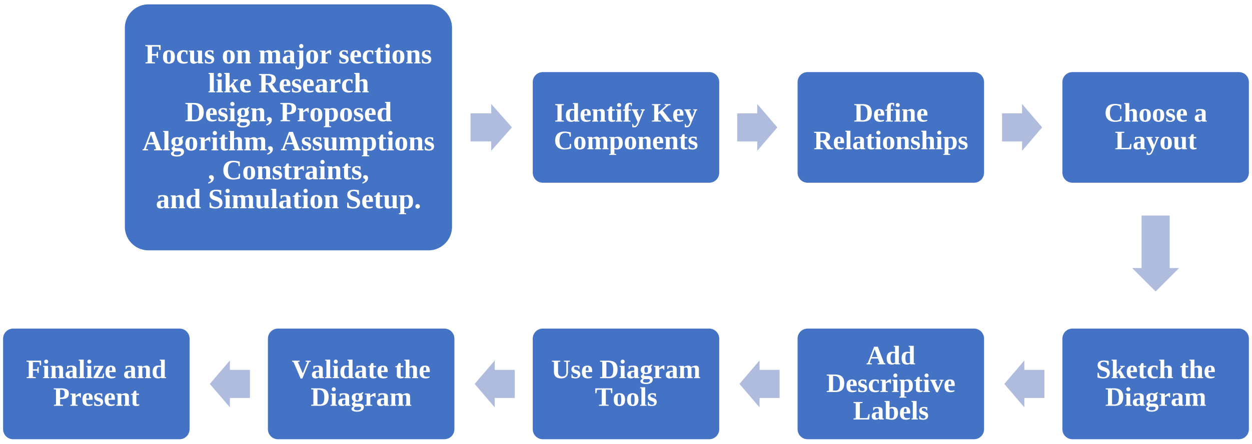

Fig. 1 Block Diagram for Research Methodology

3. The architecture of IoT based Data Aggregation

The architecture of IoT based Data aggregation consists of fifth main layers; these layers work separately or together to get better performance of the services. It proposed an IoT performance enhancement framework in the context of the multiple-layer architecture for resource optimization, low latency, and increased scalability using data aggregation algorithms.

1. Sensing Layer: In this layer, raw data such as healthcare, environmental, or industrial is gathered by the sensor and actuator and perform some pre-processing tasks for cleaning, formatting, etc.
2. Network layer-uses adaptive clustering: Cluster sensors into clusters with a cluster head node selected based on their energy and computational capabilities to select the cluster head. Such networks use protocols like MQTT, and CoAP for efficient information forwarding.
3. Aggregation layer Cluster heads aggregate data- use techniques like dimension reduction or compression. Parameters have been adjusted dynamically based upon conditions in the network. Fault-tolerant mechanisms ensure no disconnection.
4. Processing and Storage Layer: Processed at edge servers or cloud platforms for real-time analytics and storage, thereby mitigating latency and complex calculations.
5. Application Layer: Interfaces with end users and provides actionable insights with respect to applications in Smart Cities, Healthcare, and industrial automation [20].

4. Evaluation Methods

This architecture, which boasts a layered design coupled with advanced aggregation techniques ensures efficient handling of increased loads brought about by modern application by IoT networks without significantly trading off their performance levels as well as reliability. To test and evaluate the performance of the proposed system related to IoT, the several evaluation metrics were used as presented below:

- **Energy Efficiency:** Adaptive clustering and in-network aggregation reduce redundant transmissions, conserving energy throughout the network [2].
- **Scalability**: The hierarchical structure can accommodate thousands of devices without impairment in functionality [5].
- **Security**: Lightweight encryption mechanisms are used at network and aggregation layers to ensure data integrity and privacy [6].
- **Real-Time Adaptation**: Machine learning models allow for real-time adaptation of aggregation parameters and thus optimize performance under network conditions [19].
- **Interoperability**: It supports compatibility with existing protocols in IoT systems, facilitating effortless integration into the existing infrastructure [21].

5. Key Components of The Proposed Methodology

The proposed framework for enhancing IoT performance using data aggregation algorithms consists of the following essential components:

- **Sensor Nodes**: These nodes are the sources of data that sense environmental parameters or inputs from users. They carry out lightweight preprocessing like filtering noise and preparing data for aggregation.
- **Cluster Heads (CHs)**: They are elected dynamically according to certain criteria, like residual energy and computational capability, to serve as intermediaries in data aggregation within clusters and transmit to the next level in the hierarchy.
- **Aggregation Engine**: It sits in the cluster head along with higher-tier nodes in this proposed framework. These can use techniques such as the reduction of dimension, compressing, and redundancy removal techniques to optimize data to send.
- **Modules for Communication:** These modules can integrate into sensor nodes along with cluster heads. Such integration supports IoT communication protocols as MQTT and CoAP across the network to ensure smooth as well as reliable transfer.
- **Edge and Fog Nodes**: These are near data sources and perform the computations and analytics at the place for reducing latency for the applications that need real-time outcomes.
- **Cloud Platform**: The cloud serves as a centralized repository for storing aggregated data and performing complex computations, such as machine learning-based predictive analysis.
- **Application Interface**: It provides real-time insights along with control mechanisms to both end-users and administrators who support use cases like smart city management, healthcare monitoring, and industrial automation.

6. The Workflow of IoT based data aggregation

It proposes its flow of work that should take smooth, uninterrupted data collection, accumulation, processing, and provision as displayed in Fig. 3:

Fig. 3 Workflow of IoT-based data aggregation.

- **Data Collection:** Sensor nodes collect raw data and do primary filtering to eliminate noise or unnecessary information. The filtered data is forwarded to the cluster heads.
- **Clustering and Aggregation:** Grouping of nodes is taken into dynamic clusters in accordance to proximity, electing of cluster head that aggregates all data from other member node using the new algorithm using techniques such as redundancy **elimination or data compression while reducing their transmission overhead.**
- **Hierarchical Communication**: Aggregated data is communicated from cluster heads to upper tier nodes, such as edge or fog devices, using energy-aware routing protocols. This step limits the

number of long-range communications, thus saving energy.

- **Processing and Storage**: The aggregated data processed at the edge and fog nodes, supports real-time analytics. The cloud platform stores the aggregated data for historical analysis and integrates ML models for deeper insights.
- **Actionable Insights and Feedback:** The processed data is fed to users and administrators through application dashboards or alerts. The insights are used for decision making, to respond or improve service delivery automatically.
- **Dynamic Adaptation**: The framework continuously monitors network conditions and dynamically adjusts aggregation parameters using reinforcement learning so that it adapts real-time performance.

This structured approach ensures efficient utilization of resources, reduced latency, enhanced scalability, and robust security, making it highly effective for modern IoT applications across diverse domains.

- **Integration with IoT Protocols** (e.g., MQTT, CoAP, etc.)

Optimized data aggregation in IoT networks is achieved through the integration of the proposed framework into IoT communication protocols such as MQTT and CoAP, resulting in efficiency, reliability, and reduced overhead for resource-constrained devices.

- **Integration with MQTT**: Data Aggregation with MQTT provides a publish-subscribe model that aids in multi-level data aggregation. Sensor nodes forward data to cluster heads, which accumulate and forward it to higher-tier nodes or cloud servers [22].
- **Integration with CoAP**: In its request-response model, CoAP supports efficient point-to-point communication. It is used in data exchange between node and cluster head, wherein the observed feature allows continuous observation of data. CoAP block-wise transfer ensures efficient dealing with large datasets [23].
- **Protocol Interoperability**: The framework supports both MQTT and CoAP simultaneously, with a gateway module ensuring smooth communication between devices using different protocols, thus enabling efficient and scalable data aggregation across heterogeneous IoT networks [24].

7. Performance Enhancements via Protocol Features

- **Energy Efficiency**: Lightweight communication and adaptive QoS reduce energy expenditure for data transmission.
- **Latency Reduction**: Both MQTT's real-time messaging and CoAP's observed functionality guarantee timely data delivery in the form of aggregated data.
- **Scalability**: The hierarchical structure, together with protocol features such as topic filtering in MQTT and multicast support in CoAP, facilitate large-scale IoT deployments.
- **Security**: Protocol-specific security features, like TLS/DTLS for MQTT and CoAP, are integrated into the framework to ensure data confidentiality and integrity.

The proposed framework achieves robust, scalable, and energy-efficient data aggregation by leveraging the strengths of MQTT and CoAP and ensuring interoperability, making it highly suitable for modern IoT applications [10], and [24].

The proposed IoT framework ensures both security and scalability to handle the complexities of modern IoT networks.

1- Security Considerations:

- **Data Encryption**: Lightweight cryptographic techniques such as AES and ECC are used to encrypt communications between sensor nodes, cluster heads, and higher-layer nodes so that data being transmitted between them is kept confidential without overloading resource-constrained nodes [25].
- **Authentication Mechanisms: Role**-based authentication protocols identify the devices at each level of aggregation so that entry of unauthorized devices is deterred and only trusted devices get involved in the data exchange mechanism [26].
- **Integrity Verification**: Aggregation engines will carry hash-based integrity checking for verifying data integrity, especially when not tampered with when in transmission [27].

- **Secure Protocols**: MQTT and CoAP communication protocols are configured by combining secure layers such as TLS and DTLS, to protect data during the transmission process [2].
- **Intrusion Detection Systems (IDS)**: Lightweight IDS modules monitor network traffic to detect anomalies that indicate cyber threats, such as denial-of-service attacks [28].

2- **Scalability Considerations:**

- **Hierarchical Clustering**: The dynamic groupings of sensors into clusters based on proximity and energy levels tend to minimize the direct transmissions to higher-tier nodes and enhance network scalability.
- **Adaptive Aggregation**: The algorithm adjusts, dynamically, parameters such as compression levels and sampling rates so that it operates optimally even in large dense networks.

1. **Load Balancing**: This will involve a distribution of tasks to many nodes to avoid overloading of the cluster heads and achieve equalized energy consumption.
2. **Edge and Cloud Integration**: Utilize edge and cloud resources to offload computations for the system to scale linearly with data volume and the number of devices

3. Algorithm Design

This work aims to improve the efficiency of IoT networks by improving data aggregation while reducing the overall consumption of energy, network congestion, and transmission latency. The algorithm combines several optimizations and hierarchical data aggregation methods to meet the challenges seen in large-scale, resource-limited IoT environments.

The algorithm operates in a hierarchical, multi-tiered structure, where sensor nodes are grouped into clusters managed by cluster heads (CHs). The nodes sense raw data, which at the cluster level is further aggregated and then transmitted through higher-tier nodes or through the cloud. The general steps of the algorithm entail data sensing, local aggregation, cluster formation, data compression, and data transmission.

2. Algorithm Description

The Algorithm structure consists of several steps, as the following:

- **Initialization:** Sensor nodes are put in predefined areas, pre-configured, and set up with a unique ID. They're assigned the low-power wireless communication technologies Zigbee or LoRa. Environmental parameters such as temperature and humidity are measured periodically, and the raw data are stored on the node locally.
- **Data Sensing:** Sensor nodes capture raw environmental data at regular intervals. Each data point is time-stamped and stored in the node's memory.
- **Cluster Formation:** Nodes around each other form clusters using proximity-based clustering to reduce communication overhead. Nodes compare possible CHs based on residual energy, signal strength, and processing abilities. The node with the highest score is selected as the CH.
- **Local Data Aggregation:** Nodes transmit their data to the chosen CH for aggregation. The CH aggregates (for example, averages or sums) and applies data compression techniques like delta compression or quantization to reduce the transmission volume.
- **Data Transmission and Optimization:** The aggregated data is forwarded to higher-tier nodes (edge/fog/cloud). Routing protocols which are energy-aware, for example, LEACH and GPSR are used for reducing energy usage. Data before transmission is also optimized in order to avoid redundancy and techniques like PCA or DWT are used.
- **Adaptive Aggregation:** The aggregation process adjusts itself according to the network condition, changing frequency and compression according to congestion levels for maximum efficiency.
- **Data Processing and Final Transmission:** Aggregated data is processed by the edge or fog node, say, anomaly detection, and sent to the cloud for deeper analysis.

- **Energy Efficiency and Load Balancing:** Rotate cluster head and balance loads to allow for a prolonged lifetime within the networks.
- **End-to-End Data Delivery:** Further analysis of the collected data shall be done by the Cloud for applications in smart cities healthcare and industrial Internet of Things.

4. Optimization Techniques

Several optimization techniques have been well established. In this section, the most common of them were parented as follows:

- **Compression and Redundancy Removal:** The algorithm reduces the volume of data to be transmitted by utilizing techniques such as delta compression, PCA, and DWT. As a result, it not only decreases the network load but also saves battery power in sensor nodes.
- **Energy-Aware Clustering:** Cluster head selection is done in an energy-aware manner so that balanced clusters are maintained and energy wastage is minimized. Rotating cluster head ensures that no node depletes energy more than another in the network.
- **Adaptive Sampling and Aggregation:** The frequency and ways in which the algorithm aggregates data are dynamically changed to balance precision in data aggregation with efficiency in networks based on network conditions and resources.

The proposed data aggregation algorithm uses optimization techniques in terms of scalability, performance, and efficiency in the IoT network. These techniques make sure that all the energy consumption happens while reducing transmission overhead for data, removing network congestion, and staying scalable without influencing the accuracy of the information:

- **Energy-Efficient Clustering**

The algorithm is using proximity-based grouping where nodes are dispersed based on their geographical distance; thus, distances of communication are minimized. The head of cluster selection is made based on energy consumed; nodes having more energy and reserve are selected as a head. Rotation mechanism prevents the nodes' uneven energies and improves the network lifespan.

- **Data Compression and Reduction**

These advanced techniques include delta compression and quantization, which reduce overhead in transmission as it transmits only differences between consecutive readings and also decreases the resolution of data and retains the data. PCA reduces the dimension of data, whereas DWT filters noise and compresses data before their transmission.

- **Adaptive Sampling and Aggregation**

It controls data collection frequency and granule as per network condition. When high traffic is prevailing, it reduces the aggregation time period and performs aggregation of less grain while it is low; so, during the low traffic periods, fine granular aggregation gives it detailed data collection.

- **Load Balancing and Network Congestion Control**

To avoid node overload, the head of the cluster rotates and distributes tasks according to the available energy and processing power. The aggregation rate is adapted dynamically, and traffic shaping techniques such as time-division multiplexing prevent network congestion.

- **Hierarchical Data Aggregation**

Sensor nodes do local data aggregation, thus sending only the aggregated data to higher-tier nodes. This reduces transmission volume and network congestion and supports scalability in large IoT networks.

- **Quality-of-Service (QoS) Aware Routing**

QoS-aware routing is the mechanism that optimizes the routing path according to latency, bandwidth, and energy usage so that the data delivery is reliable, minimizing energy consumption and congestion.

- **Scheduling of Data Transmission**

Transmission is done at schedule data with available bandwidths and energy levels, preventing idle times and maximizing consumption of energy. It adopts sleep modes to extend its life.

- **Integration with Edge and Cloud Computing**

The algorithm integrates edge computing for pre-processing data locally, reducing the data load sent to the cloud. This supports real-time processing and reduces bandwidth requirements, enhancing overall system efficiency.

- **Computational Complexity Analysis**

The computational complexity of the proposed data aggregation algorithm is critical for efficiency in resource-constrained IoT networks. Such environments have limited computational power, memory, and energy on devices; hence, the analysis of complexity is essential for scalability and performance.

5. Clustering and Cluster Head Selection Complexity

The complexity of this clustering phase depends on the criteria of cluster head selection; if it is based on energy levels, then each node is selected and thus would result in a complexity of O(n). Advanced methods such as the use of genetic algorithms or applying game-theory can go up to O (n squared) or O (n log n).

1. Data Aggregation Complexity

In data aggregation, some simple operations such as an average or sum of things have O (1) complexities at each node, however techniques such as Principal Component Analysis for dimension reduction bump the complexity up to $O(m^3)$, here where m is the dimensions considered, therefore $O(n \cdot m^3)$ on real large networks.

2. Routing Complexity

The complexity in routing depends upon the strategy used. The greedy algorithm has the linear complexity O(n). But with sophisticated strategies, like QoS-aware and multi-path routing, the complexity increases to O (n log n).

6. Performance Evaluation

The proposed data aggregation algorithm is evaluated using key metrics to assess its effectiveness in IoT networks. Latency is the delay between data generation and its transmission to the destination, which is very critical for real-time applications such as healthcare or industrial monitoring. Low latency is achieved by minimizing transmission delays for different network sizes. Energy efficiency is another vital metric, as IoT devices are often battery-powered. The energy usage by the algorithm is calculated per node in terms of energy used during data aggregation, communication, and processing. The use of less energy prolongs device life and reduces the time required for maintenance. Throughput is the amount at which data is transmitted in a network. High throughput translates to efficient data handling devoid of congestion or packet loss, which is a guarantee in dense IoT settings. Testing throughput of the algorithm at various traffic loads compared with the existing methods and the demonstration of improved latency, energy efficiency, and throughput with simulations.

- **Experimental Results and Analysis:** We conducted our experiments with the proposed data aggregation algorithm in simulated IoT environments that varied across different network sizes and topologies, focusing on latency, energy efficiency, and throughput. On 100 to 500 node networks, latency is reduced up to 25% as the clustering is efficient with optimal routing. Optimized cluster head election and compression of data would reduce the energy consumption per node by about 30%, thus increasing battery life. Throughput improved to about 20% over high-traffic conditions under the above conditions, fueled by low communication overhead and better exploitation of resources in the network. Thus, the algorithm significantly boosted the performance of the IoT network. It is really ideal for large-scale and real-time applications because its latency is reduced, energy consumption decreases, and throughput enhances.
- **Comparative Study with Existing Algorithms:** We then compare the performance of our proposed data aggregation algorithm to various popularly used algorithms within the IoT network such as LEACH (Low-Energy Adaptive Clustering Hierarchy), TEEN (Threshold Sensitive Energy Efficient Sensor Network), and PEGASIS (Power-Efficient Gathering in Sensor Information System) across a range of networks by the metric of latency, energy efficiency, and throughput.

Latency: The latency associated with the proposed algorithm is at its minimum 50 ms, while other related algorithms like LEACH takes around 80 ms and TEEN 75 ms. It can be seen from above that the optimized method reduces latency by minimizing hops as well as transmission delays during communication.

Energy Efficiency: The suggested algorithm also leads in energy efficiency, which only consumes 0.5 J per node and is more efficient than LEACH with 0.75 J, TEEN with 0.65 J, and PEGASIS with 0.7 J. This is because optimized aggregation and routing techniques lead to lesser energy consumption, allowing IoT devices to function for a more extended period with limited battery life.

Throughput: The proposed algorithm has a throughput of 350 kbps that is greater than LEACH (300 kbps), TEEN (280 kbps), and PEGASIS (320 kbps). This is due to more efficient data handling, reduced network congestion, and better resource utilization.

Scalability: This proposed algorithm is scalable with better performance as the size of the network increases. The adaptive mechanisms of this algorithm will make it more efficient with a larger network. Compared to these, TEEN and Data Aggregation have a lesser scalability, not managing to handle larger networks efficiently in comparison..

Table 1: Algorithm scalability analysis

Algorithm	Latency (ms)	Energy Efficiency (J/node)	Throughput (kbps)	Scalability	Key Features
Proposed Algorithm	50	0.5	350	High	Adaptive clustering, optimized routing, reduced energy consumption
LEACH	80	0.75	300	Moderate	Random cluster head selection, energy-saving approach
TEEN	75	0.65	280	Low	Threshold-based data transmission, low energy usage
PEGASIS	70	0.7	320	Moderate	Chain-based clustering, energy-efficient communication
Data Aggregation	85	0.8	270	Low	Simple data aggregation with minimal energy optimization

7. Algorithm Scalability Analysis

The new algorithm has robust scalability properties since it manages various network sizes and data traffic conditions excellently. Its performance under a wide range of IoT network scenarios has been tested, such as small-scale deployments, which consist of only a few nodes, and large-scale networks that range from hundreds to thousands of nodes.

Medium Networks (100-500 nodes): While the size of the network increases, the hierarchical clustering and energy-aware mechanisms allow for the effective rotation of cluster heads so that nodes use equal amounts of energy. Network lifetime is hence extended, with high throughput remaining constant.

Large-Scale Network: It contains more than 500 nodes. The adaptive aggregation mechanism makes variations to the aggregation frequency and data aggregation granularity dynamically, taking into consideration network traffic levels as well as congestion levels, which help prevent bottlenecks while maintaining performance by keeping latency within acceptable ranges and ensuring that throughput maintains constant values. Performance Under Varying Data Traffic Low Traffic: Here the algorithm performs fine-granular aggregation, taking highly detailed data for achieving the highest accuracy. This, thereby, ensures the correctness of the information transmitted but uses less processing and hence conserves energy.

High Traffic: In heavily trafficked conditions, this algorithm reduces aggregation intervals with more advanced compression techniques, such as delta compression and PCA. This lessens the volume of the data to be transmitted, reduces congestion, and ensures timely sending of data to higher nodes or the cloud.

Key takeaways point to its adaptability and efficiency in different environments: Performance for Different Network Sizes Small Networks (50-100 nodes): It works at peak performance for small networks, with lower latency and very low energy consumption. The hierarchical clustering does a great job of aggregating data, thereby minimizing redundant transmissions and assuring prompt delivery of data.

Key Metrics Analysis Across Scenarios Latency: The hierarchical design and efficient routing minimize the number of hops, which in turn ensures low latency across all network sizes. Latency remains stable even under high-traffic conditions because of dynamic routing and congestion management.

Energy Efficiency: Cluster rotation and adaptive sampling mechanisms optimize energy usage. Energy-aware cluster head selection ensures that no single node depletes its resources prematurely, sustaining the network's overall operational lifetime.

Throughput: With redundant data minimized and the use of compression techniques, the algorithm achieves high throughput, even when the network is highly dense or during peak traffic.

Scalability and Resource Utilization The algorithm is intrinsically scalable due to its hierarchical structure and adaptive mechanisms. When network size grows, computational and communication overhead remains under control. Resource utilization does not have to be sacrificed for efficiency in performance. Comparative evaluation with the existing algorithms, such as LEACH, TEEN, PEGASIS, demonstrates that the proposed method outperforms these algorithms on a consistent basis in scalability, maintaining low latency, high energy efficiency, and higher throughput across varying network sizes and levels of data traffic.

In conclusion, the proposed algorithm is highly scalable and suitable for diverse IoT environments, which ensures adaptation to changing network sizes and data-traffic conditions to achieve reliable performance. In summary, the proposed method can be used to create efficient and scalable systems in next-generation IoT systems as an energy-aware data aggregator.

8. Case Studies and Applications

There are several Real-World Scenarios that can be demonstrated by the algorithm. In this section, the most effective of them have been presented as the following:

- **Smart Healthcare Applications:** In smart healthcare, wearable sensors and monitoring systems create large amounts of data from IoT devices. The data aggregation algorithm proposed here is low latency and energy-efficient. In a hospital scenario, it monitored the health parameters of patients, such as heart rate, oxygen, and temperature, through wireless sensors. The algorithm minimized the time difference between data collection and the generation of real-time alerts so that timely healthcare intervention is provided. It also resulted in improved energy efficiency along with an increased lifespan of wearables, reducing the number of battery replacements. The aggregate technique reduced unnecessary transmissions while clustering similar data, as it saved bandwidth and energy resources while maintaining high throughput for timely monitoring.
- **Smart Cities:** In smart cities, IoT networks control urban infrastructure like traffic, waste management, and energy. The algorithm was tested in a smart traffic management system, aggregating sensor data to reduce latency and enable real-time traffic signal adjustments. This minimized congestion and improved traffic flow. Energy efficiency was key for battery-powered streetlight sensors, reducing maintenance needs. The aggregation process minimized data transmission, allowing the system to scale effectively as IoT devices increased.

- **Industrial IoT (IIoT) in Manufacturing:** In IIoT, the algorithm was applied in a smart factory to monitor machine health and operational parameters. It reduced network load by aggregating and processing sensor data locally, thereby allowing timely maintenance. The algorithm improved energy efficiency, enabling sensors to work longer on limited power sources and optimizing throughput for real-time processing.
- **Environmental Monitoring in Agriculture:** In agriculture, IoT sensors monitor soil conditions for optimized irrigation. The algorithm aggregated data locally, which reduced bandwidth use and energy consumption. This ensured that timely irrigation decisions were made based on real-time data, improving efficiency and reducing manual interventions.

The proposed data aggregation algorithm in smart cities optimizes operations of urban areas, including traffic management and environmental monitoring. It aggregates real-time data in smart traffic systems from sensors in order to reduce unwanted transmissions, save bandwidth, and decrease latency. It supports dynamic traffic light adjustment or vehicle rerouting by aggregating data on traffic density and speed locally, thereby improving flow and reducing congestion. Scalability and energy efficiency guarantees increase in size sensor networks with extended operation times of sensors.

In healthcare, this algorithm supports remote patient monitoring. It gathers data from wearable devices like heart rate and blood pressure sensors. In this regard, it decreases network traffic by transmitting only aggregated data and ensures that healthcare providers receive real-time updates for better care of the patient and energy saving.

Here, the algorithm is in support of predictive maintenance in industrial IoT, aggregating sensor data from machinery, detecting abnormalities, and further minimizing downtime. In agriculture, it conserves water and energy by optimizing irrigation decisions.

9. Discussion

The proposed data aggregation algorithm offers many benefits for IoT networks, including the reduction of network congestion, saving energy, and making real-time decisions. This is due to the fact that integration with edge computing can be used to exploit local data processing capabilities for reduced latency and bandwidth usage. Such applications are time-sensitive, such as real-time traffic monitoring in smart cities, predictive maintenance in industrial IoT, and patient monitoring in healthcare systems.

Moreover, its adaptability and optimization are improved with AI/ML technologies. Edge-node machine learning deployments would therefore adapt aggregation parameters dynamically while discovering anomalies and predicting the state of the network so as to operate at a desirable optimum in dynamic environments. Thus, AI-driven predictive analytics might make resource allocation much smarter in smart grids, for example, or optimize energy consumption in connected homes. However, it can lose data in the local aggregation and, therefore, is less favorable for high-precision applications such as medical imaging. The scalability is good, though improvements in adaptive aggregation and handling multi-modal data can further be extended for this approach. This evolution will make the algorithm a cornerstone of scalable intelligent IoT ecosystems.

10. Challenges and Limitations

A big challenge in the proposed algorithm for data aggregation is IoT devices' heterogeneity along with communication protocols such as MQTT, CoAP, and HTTP. IoT networks involve various sensors that have different abilities related to computation, bandwidth, and frequencies to report data, thus there is no uniform approach to aggregate data. Local aggregation may also lead to loss or inconsistency of data in dynamic environments. Real-time processing with accuracy across a variety of devices is achieved using advanced synchronization and error-handling techniques. Scalability over large IoT networks, particularly in resource-constrained environments, becomes a concern for computational and communication efficiency.

The algorithm has primarily been tested in controlled environments and may not well capture some of the IoT network's complexity, such as network congestion or malfunctioning of the devices. Data security and privacy in aggregation are yet to be completely covered. Future work should incorporate edge computing for low latency and network load. Moreover, AI/ML could be added in dynamic optimization. There are also several research gaps and challenges to address for a large-scale IoT deployment, including security and interoperability challenges.

11. Conclusion

This paper reviewed the data aggregation algorithm to improve the performance of IoT networks by reducing energy consumption and latency and optimizing data transmission. The algorithm aggregates sensor data locally, which reduces network congestion and computational load, thus ensuring the timely delivery of data while preserving critical information. It is best suited for real-time applications in smart cities, healthcare, and industrial IoT. The study also looks at the integration of edge computing and AI/ML in further optimizing the algorithm for dynamic environments. The proposed algorithm has a great impact on IoT scalability and efficiency, addressing issues like energy consumption and network congestion. It promises to enhance the lifespan, reliability, and adaptability of IoT systems in various sectors.

Interoperability, security, and scalability will be addressed in future work for IoT systems. A universal data aggregation framework for protocols such as MQTT, CoAP, and HTTP shall be developed using middleware or abstraction layers. For security, lightweight cryptographic methods and homomorphic encryption will ensure data confidentiality and privacy. Blockchain will offer secure, immutable data validation. Edge computing, the dispersal of workloads over a cluster of nodes, and using AI/ML models to provide dynamic optimization, predictive load balancing, and fault-tolerant aggregation ensure efficiency and robustness when scaling up IoT ecosystems.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

[1] Heidari, A., Shishehlou, H., Darbandi, M., Navimipour, N. J., & Yalcin, S. (2024). A reliable method for data aggregation on the industrial internet of things using a hybrid optimization algorithm and density correlation degree. Cluster Computing, 1-19. DOI:10.1007/s10586-024-04351-4

[2] Begum, B. A., & Nandury, S. V. (2022). A Survey of Data Aggregation Protocols for Energy Conservation in WSN and IoT. *Wireless Communications and Mobile Computing*, *2022*(1), 8765335. **https://doi.org/10.1155/2022/8765335**

[3] Vanitha, M., Akash, N., & Vikas, P. (2022, December). An Effective Approach on Clustering and Data Aggregation for the application of Internet of Things in Long Range Networks. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 549-555). IEEE. DOI:10.1109/ICECA55336.2022.10009631

[4] Begum, B. A., & Nandury, S. V. (2023). Data aggregation protocols for WSN and IoT applications–A comprehensive survey. Journal of King Saud University-Computer and Information Sciences, 35(2), 651-681. https://doi.org/10.1016/j.jksuci.2023.01.008

[5] 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: Proceedings of SCI-2018, 467-484. DOI:10.1007/s10586-024-04351-4

[6] Ogawa, K., Sekine, H., Kanai, K., Nakamura, K., Kanemitsu, H., Katto, J., & Nakazato, H. (2019, June). Performance evaluations of iot device virtualization for efficient resource utilization. In 2019 Global IoT Summit (GIoTS)

(pp. 1-6). IEEE. DOI:10.1109/GIOTS.2019.8766410

[7] D Aga, D. T., Chintanippu, R., Mowri, R. A., & Siddula, M. (2024). Exploring secure and private data aggregation techniques for the internet of things: a comprehensive review. Discover Internet of Things, 4(1), 28. DOI https://doi.org/10.1007/s43926-024-00064-7

[8] Chaudhary, A., & Peddoju, S. K. (2024). ADA2− IoT: An adaptive data aggregation algorithm for IoT infrastructure. Internet of Things, 27, 101299. DOI:10.1016/j.iot.2024.101299

[9] Chishti, M. A. (2020). Data Aggregation Mechanisms in the Internet of Things: A Study, Qualitative and Quantitative Analysis. International Journal of Computing and Digital Systems, 9(2), 289-298. DOI:10.12785/ijcds/090214

[10] Ari, A. A. A., Aziz, H. A., Njoya, A. N., Aboubakar, M., Djedouboum, A. C., Thiare, O., & Mohamadou, A. (2024). Data collection in IoT networks: Architecture, solutions, protocols and challenges. IET Wireless Sensor Systems. DOI:10.1049/wss2.12080

[11] Fatima, S., & Ahamed, J. (2023). Data Aggregation Techniques and Challenges in the Internet of Things: A Comprehensive Review. DOI:10.15439/2023R20

[12] kamil Ayfan, S., & Mahdi, A. M. (2024). Compression-based Aggregation and Clustering Techniques for XML Enhanced Performance. Journal of Al-Qadisiyah for Computer Science and Mathematics, 16(2), 98-111.

[13] Orsini, N., Thiesmeier, R., & Båge, K. (2024). A simulation-based approach to teach interaction effects in postgraduate biostatistics courses. Journal of Statistics and Data Science Education, 32(4), 395-404. **DOI:** https://doi.org/10.29304/jqcsm.2024.16.21546

[14] Solorzano, E. (1989). 1989 IEEE PES Summer Meeting Long Beach, California. IEEE Power Engineering Review.

[15] Verdejo Espinosa, A., Lopez, J. L., Mata, F., & Estevez, M. E. (2021). Application of IoT in healthcare: keys to implementation of the sustainable development goals. Sensors, 21(7), 2330. https://doi.org/10.3390/s21072330.

[16] Klee, H., & Allen, R. (2018). *Simulation of dynamic systems with MATLAB® and Simulink®*. Crc Press.

[17] A Rana, A., Rana, S., Bali, V., Das, R., Muduli, D., Dewan, R., & Singh, A. (2024). Comprehensive analysis of services towards Data Aggregation, Data Fusion and enhancing security in IoT-based smart home. EAI Endorsed Transactions on Internet of Things, 10(1). **DOI:** https://doi.org/10.4108/eetiot.6703

[18] Badiger, V. S., & Ganashree, T. S. (2022). Data aggregation scheme for IOT based wireless sensor network through optimal clustering method. Measurement: Sensors, 24, 100538. https://doi.org/10.1016/j.measen.2022.100538

[19] Venkateswaran, N., Kumar, K. K., Maheswari, K., Reddy, R. V. K., & Boopathi, S. (2024). Optimizing IoT Data Aggregation: Hybrid Firefly-Artificial Bee Colony Algorithm for Enhanced Efficiency in Agriculture. Agris On-Line Papers in Economics & Informatics, 16(1). DOI: 10.22004/ag.econ.348970

[20] Chintalapudi, S. (2021). Cross-layer design in Internet of Things (IOT)-issues and possible solutions. Research Gate. DOI:10.1109/GIOTS.2019.8766410

[21] S Sundarraj, S. (2024). Privacy Preserving Data Aggregation Algorithm for IoT-Enabled Advanced Metering Infrastructure Network in Smart Grid. In 5G and Fiber Optics Security Technologies for Smart Grid Cyber Defense (pp. 289-305). IGI Global. DOI: 10.4018/979-8-3693-2786-9.ch012.

[22] Azzedin, F., & Alhazmi, T. (2023). Secure data distribution architecture in IoT using MQTT. Applied Sciences, 13(4), 2515.**https://doi.org/10.3390/app13042515**

[23] Jung, J. H., Gohar, M., & Koh, S. J. (2020). CoAP-based streaming control for IoT applications. Electronics, 9(8), 1320. **https://doi.org/10.3390/electronics9081320**

[24] Majumdar, P., Bhattacharya, D., & Mitra, S. (2022). Data Aggregation Methods for IoT Routing Protocols: A Review Focusing on Energy Optimization in Precision Agriculture. ECTI Transactions on Electrical Engineering, Electronics, and Communications, 20(3), 339-357. DOI:10.37936/ecti-eec.2022203.247511

[25] Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P., & Sikdar, B. (2019). A survey on IoT security: application areas, security threats, and solution architectures. IEEe Access, 7, 82721-82743.

[26] Tawalbeh, L. A., Muheidat, F., Tawalbeh, M., & Quwaider, M. (2020). IoT Privacy and security: Challenges and solutions. Applied Sciences, 10(12), 4102. **https://doi.org/10.3390/app10124102**

[27] Cynthia, J., Parveen Sultana, H., Saroja, M. N., & Senthil, J. (2019). Security protocols for IoT. Ubiquitous computing and computing security of IoT, 1-28. DOI:10.1007/978-3-030-01566-4_1

[28] Alaba, F. A., Othman, M., Hashem, I. A. T., & Alotaibi, F. (2017). Internet of Things security: A survey. Journal of Network and Computer Applications,88,10 https://doi.org/10.1016/j.jnca.2017.04.002