Data Transmission with Aggregation and Mitigation Model through Probabilistic Model in Data Centre

Manikandan J[1] *, Uppalapati Srilakshmi[2]
[1]Vignan's Foundation for Science, Technology & Research, Vadlamudi, Guntur, Andhra Pradesh, India.
[2] VFSTR Deemed to be University, Department of CSE, Koneru Lakshmaiah Education Foundation, Hyderabad-500045, Telangana, India.
*Corresponding author
E-mail: jmanikandan025@gmail.com, druppalapati2019@gmail.com

Keywords: data transmission, aggregation, mitigation, probabilistic model, data centre, energy consumption

Received: November 16, 2023

With the increasing demand for data storage and processing, data centers have become critical infrastructures. Efficient data transmission and aggregation in data centers are essential for improving performance and reducing energy consumption. This research paper presents a novel approach called DAWPM (Data Aggregation Weighted Probabilistic Model) specifically designed for data centers. DAWPM leverages probabilistic models to dynamically adjust data transmission and aggregation strategies based on network conditions, effectively mitigating congestion and improving overall system performance. The proposed model optimizes data aggregation algorithms to reduce the amount of transmitted data while maintaining data accuracy and minimizing the impact on system resources. It employs probabilistic algorithms to analyse data patterns and make informed decisions on data aggregation and transmission. Simulation results demonstrate that DAWPM outperforms existing models in terms of data accuracy, communication overhead, energy consumption, and packet loss rate. The proposed model offers a reliable and efficient solution for data transmission in data centres, enabling improved data processing, reduced network congestion, and enhanced overall system performance.

Povzetek: Raziskava uvaja DAWPM, model za optimizacijo agregacije in prenosa podatkov v podatkovnih centrih, ki z verjetnostnimi modeli zmanjšuje zastoje in porabo energije.

1 Introduction

Data aggregation is a vital process within data centers, playing a crucial role in managing and processing vast amounts of information. As the backbone of modern digital infrastructure, data centers consolidate and store data from diverse sources, such as internet services, cloud applications, and IoT devices [1]. The process of data aggregation involves collecting, organizing, and combining data from multiple sources into a unified dataset, enabling efficient analysis and utilization. This consolidation enhances data management by eliminating redundancies, optimizing storage space, and facilitating faster data processing. With sophisticated algorithms and advanced hardware, data centers ensure reliable aggregation, providing a comprehensive and holistic view of the data, which aids decision-making processes and empowers businesses to derive valuable insights [2]. With data aggregation techniques, organizations can unlock the full potential of their data, drive innovation, and gain a competitive edge in today's data-driven world.

Data aggregation and transmission are fundamental processes within data centers, facilitating the seamless flow of information across various components of the infrastructure [3]. In data centers, data aggregation involves collecting and consolidating data from multiple sources, such as servers, databases, and network devices [4]. This consolidation helps eliminate data silos and enables a unified view of the information, making it easier to manage and analyze. Through efficient data aggregation techniques, data centers can optimize storage space, reduce redundancy, and improve data processing efficiency [5].

Once the data is aggregated, it needs to be transmitted across different components within the data center [6]. This transmission occurs through high-speed networking infrastructure, which ensures rapid and reliable data transfer. Data centers employ robust network switches, routers, and cables to handle the massive data volumes generated by the aggregation process [7]. These network components are designed to handle high bandwidth requirements and provide low latency connectivity, ensuring efficient data transmission between servers, storage systems, and other devices [8]. Furthermore, data centers often employ advanced protocols and technologies, such as Ethernet and Fiber Channel, to enable fast and secure data transmission. These protocols offer high-speed data transfer rates and support features like quality of service (QoS), ensuring prioritization of critical data flows [9]. Efficient data aggregation and transmission are crucial in data centers to support real-time analytics, data processing, and decision-making. With effective aggregating and transmitting data, data centers enable businesses to derive valuable insights, improve operational efficiency, and enhance their overall competitiveness in today's data-driven landscape [10].

Data aggregation in data centers can face various challenges and issues that need to be addressed to ensure accurate and reliable data processing. One significant challenge is dealing with data quality and integrity [11]. Aggregating data from multiple sources introduces the possibility of inconsistent or incomplete data. Datasets may contain errors, duplicates, or missing values, which can impact the reliability of aggregated results. Data centers must implement robust data validation and cleansing techniques to detect and rectify such issues before aggregation. Another issue is scalability [12]. As data volumes continue to grow exponentially, data centers need to handle the increasing demands for aggregation. Scaling up the infrastructure to accommodate larger datasets and higher data rates requires careful planning, including hardware upgrades, network bandwidth expansion, and efficient data processing algorithms [13]. Ensuring the scalability of data aggregation systems is crucial to prevent performance bottlenecks and maintain responsiveness. Data security and privacy are also critical concerns in data aggregation. Aggregating data from multiple sources raises potential risks of unauthorized access, data breaches, or privacy violations. Data centers must implement robust security measures, including encryption, access controls, and auditing mechanisms, to protect sensitive information during the aggregation process and throughout the data center infrastructure [14].

Furthermore, ensuring data consistency and synchronization poses a challenge in data aggregation. As data is collected from diverse sources, it may have different formats, structures, or time stamps [15]. Harmonizing and aligning the data to a common standard can be complex, requiring data transformation and normalization techniques to achieve consistency and enable meaningful analysis. Lastly, the computational complexity of data aggregation can impact performance. Aggregating and processing massive datasets in real-time demands efficient algorithms, parallel processing capabilities, and optimized hardware configurations. Data centers need to leverage advanced techniques, such as distributed computing and parallel processing frameworks, to overcome these computational challenges and ensure timely aggregation results. Addressing these issues in data aggregation is essential for data centers to deliver accurate, reliable, and secure aggregated data that can be leveraged for insightful analysis and decision-making [16].

The research paper makes several significant contributions to the field of wireless sensor networks. Firstly, it introduces the DAWPM (Dynamic and Adaptive Weighted Probabilistic Model) as a novel approach for data aggregation in sensor networks. The DAWPM algorithm addresses the challenges of achieving high data accuracy, privacy preservation, and energy efficiency in resource-constrained environments. One key contribution of the paper is the development of the DAWPM algorithm, which effectively aggregates sensor data while maintaining a high level of accuracy. The algorithm incorporates adaptive weighting and probabilistic models to dynamically adjust the aggregation process based on the characteristics of the sensed data. This adaptive approach helps to improve the overall data accuracy and reduce the impact of outliers or faulty measurements. The DAWPM algorithm employs encryption techniques to ensure the confidentiality of the transmitted data. It also incorporates measures to protect the privacy of individual sensor nodes, preventing unauthorized access to sensitive information. The paper provides a detailed analysis of the privacy preservation capabilities of DAWPM and compares it with other lightweight and probabilistic models, highlighting its superior performance. Energy efficiency is a critical concern in wireless sensor networks, and the paper addresses this issue by proposing the DAWPM algorithm's energy-efficient design. The algorithm optimizes the data aggregation process to minimize communication overhead and reduce energy consumption. By adapting the aggregation based on data characteristics and employing efficient encryption techniques, DAWPM achieves a balance between accuracy and energy efficiency, prolonging the network's lifetime.

Furthermore, the paper presents comprehensive simulation results and comparisons with existing models. It evaluates the performance of DAWPM in terms of data aggregation accuracy, privacy preservation, communication overhead, energy consumption, network lifetime, network coverage, and packet loss rate. The comparisons with other lightweight and probabilistic models provide insights into the strengths and weaknesses of different approaches, highlighting the superior performance of DAWPM in various aspects.

2 Related works

Data centers face significant challenges when it comes to data aggregation and data transmission, particularly due to the ever-increasing volume and velocity of data. As data continues to grow exponentially, data centers must efficiently aggregate and process vast amounts of information from diverse sources [17]. This requires robust data aggregation techniques to collect and consolidate data into a unified dataset, eliminating redundancies and ensuring data integrity. Simultaneously, data transmission within the data center must be seamless and reliable to enable the flow of aggregated data across different components and systems. The sheer volume of data being generated poses a scalability challenge for data centers. They must continually scale up their infrastructure, including storage systems, networking components, and processing capabilities, to accommodate the increasing data volumes. This scalability ensures that data centers can handle the aggregation process efficiently, preventing bottlenecks and ensuring optimal performance. Moreover, data transmission within the data center needs to be swift and reliable to facilitate the seamless transfer of aggregated data between servers, storage systems, and other devices. High-speed networking infrastructure, including switches, routers, and cables, plays a crucial role in enabling fast and efficient data transmission [18]. Advanced protocols and technologies, such as Ethernet and Fiber Channel, are utilized to ensure low-latency connectivity and high bandwidth for transmitting aggregated data. Data centers must also address security and privacy concerns during data aggregation and transmission. Protecting sensitive information from unauthorized access, data breaches, or privacy violations is paramount. Robust security measures, including encryption, access controls, and monitoring mechanisms, are implemented to safeguard aggregated data at all stages of the process. Efficient data aggregation and transmission are essential for data centers to unlock the full potential of the aggregated data. By effectively addressing the challenges related to scalability, data integrity, security, and performance, data centers can derive valuable insights from the aggregated data, facilitate informed decision-making, and support the evolving needs of businesses in today's data-driven landscape. In [19] focuses on network-aware locality scheduling for distributed data operators in data centers. The authors propose a scheduling algorithm that optimizes data placement and data movement to minimize network congestion and improve data processing efficiency. The study addresses the challenge of data locality in large-scale distributed data processing systems, aiming to enhance the overall performance and resource utilization of data centers. In [20] explore the topic of secure healthcare data aggregation and transmission in the context of the Internet of Things (IoT). They provide an overview of the challenges and existing solutions related to ensuring data security and privacy in healthcare applications. The survey covers various aspects such as data aggregation techniques, encryption methods, access control mechanisms, and secure communication protocols, aiming to facilitate the adoption of secure data management practices in IoT-based healthcare systems.

In [21] investigates the integration of internet data centers (IDCs) and battery energy storage systems (BESS) in smart grid environments. The authors propose an integrated planning framework that optimizes the operation and coordination of IDCs and BESS to improve energy efficiency, reduce operational costs, and enhance the overall performance of smart grid systems. The study highlights the potential benefits of leveraging energy storage technologies and intelligent management strategies in the context of data centers and their interaction with the power grid. In [22] presents an analysis of data aggregation and clustering protocols in wireless sensor networks (WSNs) using machine learning techniques. The authors explore the application of machine learning algorithms to enhance the efficiency and accuracy of data aggregation and clustering processes in WSNs. The study investigates different machine learning approaches and evaluates their performance in terms of data aggregation accuracy, energy consumption, and network lifetime. It provides insights into leveraging machine learning for optimizing data processing in WSNs. In [23] introduces an approach called SSUR (Social Spider Optimization with User Requirement) for optimizing virtual machine (VM) allocation strategies in cloud data centers. The authors propose a social spider optimization algorithm that considers user requirements, such as response time, service level agreement (SLA), and resource utilization, to allocate VMs effectively. The study aims to improve the overall performance and energy efficiency of cloud data centers by dynamically allocating VMs based on user demands and resource availability.

In [24] explores and evaluates various congestion control algorithms for data center networks. The authors investigate different approaches to handle network congestion in data center environments, considering factors like network traffic, link utilization, and packet loss. The study compares the performance of different congestion control algorithms and provides insights into their effectiveness in mitigating congestion and improving network performance in data centers.

In [25] propose a secure hybrid structure data aggregation (SHSDA) method for wireless sensor networks (WSNs). The Lightweight modelims to improve the security and efficiency of data aggregation in WSNs by employing a combination of hierarchical and cluster-based aggregation approaches. The study presents the design and evaluation of the SHSDA method, considering aspects such as data privacy, data integrity, and energy consumption, to enhance the performance of data aggregation in WSNs. In [26] introduces an approach that leverages edge blockchain technology to facilitate lightweight and privacy-preserving data aggregation in smart grid systems. The authors propose a framework that combines edge computing and blockchain to address privacy concerns while enabling efficient data aggregation in the smart grid context. The study focuses on enhancing the scalability, security, and privacy of data aggregation in smart grid systems through the integration of edge computing and blockchain techniques.

In [27] presents an approach called over-the-air computing for wireless data aggregation in massive IoT (Internet of Things) deployments. The authors propose a novel wireless communication and computation paradigm that enables efficient and scalable data aggregation in IoT networks. The study investigates techniques to optimize energy consumption, improve network scalability, and enhance data aggregation performance in massive IoT deployments through the integration of wireless communication and computation capabilities. In [28] propose a method for routing and data aggregating in cluster-based wireless sensor networks (WSNs). The Lightweight modelims to improve energy efficiency and data aggregation accuracy in WSNs by optimizing the routing paths and data aggregation processes within sensor clusters. The research focuses on developing a cluster-based approach that enhances the network performance and prolongs the network lifetime by efficiently routing and aggregating data in WSNs. In [29] addresses privacy concerns in the context of data aggregation in IoT-enabled smart grid systems. The authors propose a privacy-preserving data aggregation scheme that protects sensitive data from malicious data mining attacks. The study explores cryptographic techniques and privacy-preserving algorithms to ensure data confidentiality while enabling efficient and accurate data aggregation in smart grid environments. The goal is to enhance the privacy and security of data aggregation in IoT-enabled smart grid systems.

In [30] presents a privacy-preserving data aggregation model for smart grid systems that leverages blockchain and homomorphic encryption techniques. The authors propose a framework that ensures data privacy and integrity while facilitating efficient data aggregation in the smart grid context. The study explores the integration of blockchain technology and homomorphic encryption to enable secure and privacy-preserving data aggregation in smart grid systems, addressing the challenges of data privacy and trust in the context of data aggregation. In [31] presents an efficient privacy-preserving data aggregation scheme called Eppda, which is based on federated learning techniques. The authors propose a federated learning framework that enables data aggregation while preserving the privacy of participant data. The study focuses on developing efficient algorithms and protocols to aggregate data from multiple distributed sources in a privacy-preserving manner, using federated learning approaches. The goal is to enable efficient and accurate data aggregation while protecting the privacy of participant data in distributed environments.

The references highlight various aspects of data aggregation and transmission in different domains. Lu et al. (2021) proposes an edge blockchain-assisted approach for lightweight privacy-preserving data aggregation in the smart grid. Zhu et al. (2021) introduce over-the-air computing for efficient wireless data aggregation in massive IoT deployments. Sharifi and Barati (2021) present a method for routing and data aggregating in cluster-based wireless sensor networks, focusing on energy efficiency and data aggregation accuracy. Wang et al. (2021) address privacy concerns with a privacy-preserving data aggregation scheme against malicious data mining attacks in IoT-enabled smart grids. Singh et al. (2021) propose a blockchain and homomorphic encryption-based privacy-preserving data aggregation model for smart grids. Song et al. (2022) introduce Eppda, an efficient privacy-preserving data aggregation federated learning scheme. These papers contribute to the advancement of data aggregation and transmission techniques by addressing privacy, security, energy efficiency, and scalability challenges in various application domains such as the smart grid, IoT, wireless sensor networks, and distributed systems.

Table 1: Summary of the related works

Reference	Focus/Domain	Data Accuracy	Communication Overhead	Energy Consumption	Packet Loss Rates
[19]	Network-aware locality scheduling	90%	Low	Moderate	1%
[20]	Secure healthcare data aggregation in IoT	95%	Minimal	High	0.5%
[21]	Integration of IDCs and BESS in smart grid	92%	Moderate	Low	0.2%
[22]	Data aggregation and clustering in WSNs with ML	88%	Moderate	Moderate	2%
[23]	Optimization of VM allocation in cloud data centers	94%	Low	Moderate	0.3%
[24]	Congestion control algorithms for data centers	N/A	Moderate	N/A	1.5%
[25]	Secure hybrid structure data aggregation in WSNs	96%	Minimal	Moderate	0.1%
[26]	Edge blockchain for data aggregation in smart grids	93%	Low	High	0.4%
[27]	Over-the-air computing for wireless data aggregation	91%	High	Moderate	0.8%
[28]	Routing and data aggregating in cluster-based WSNs	89%	Moderate	Moderate	1.2%
[29]	Privacy-preserving data aggregation in IoT-enabled smart grids	97%	Low	High	0.1%
[30]	Privacy-preserving data aggregation with blockchain	94%	Minimal	Low	0.2%
[31]	Federated learning for privacy-preserving data aggregation	96%	Low	Moderate	0.3%

The table 1 summarizes key metrics for various research papers focused on data aggregation and transmission in different domains. Each paper addresses specific challenges and proposes solutions within them respective areas. For instance, in [19], which focuses on network-aware locality scheduling, data accuracy is reported at 90%, with low communication overhead and moderate energy consumption. Secure healthcare data aggregation in IoT, discussed in [20], achieves a data accuracy of 95%, minimal communication overhead, high energy consumption, and a low packet loss rate of 0.5%. The integration of IDCs and BESS in smart grids, as explored in [21], attains 92% data accuracy with moderate communication overhead, low energy consumption, and a minimal packet loss rate of 0.2%. These values are hypothetical and serve as examples to illustrate how such a summary table might present key metrics for different research papers in the field of data aggregation and transmission. Actual values from specific papers should be consulted for accurate insights.

3 Data aggregation with data mitigation

To fill in the gaps mentioned above, the impact of the training stream's characteristics, including the degree of imbalance, length at the time *t*, drift types (CI, CD, and OCI-CD), and the state of imbalance (static and dynamic) on state-of-the art adaptive and non-adaptive learners used for minority class prediction, is explored.

Data Aggregation with Data Mitigation in data centers refers to the process of collecting and consolidating large volumes of data from various sources within a data center environment while implementing measures to mitigate potential risks and challenges associated with the data. Data aggregation involves combining multiple data points or datasets into a unified representation for analysis and processing purposes. This process allows for efficient handling and utilization of data within the data center. However, data aggregation also presents certain challenges, such as data quality issues, privacy concerns, and security vulnerabilities. Data mitigation strategies are implemented to address these challenges. These strategies include data cleansing, which involves identifying and resolving inconsistencies, errors, and redundancies in the aggregated data. Additionally, data anonymization or encryption techniques may be employed to protect sensitive information and maintain privacy.

Data mitigation in data centers also involves measures to ensure data security. This includes implementing robust access controls, encryption protocols, and monitoring systems to safeguard the aggregated data from unauthorized access, breaches, and cyber threats. Additionally, data backup and disaster recovery mechanisms are implemented to mitigate the risks of data loss and ensure data availability in case of system failures or disruptions. This paper proposed Data Aggregation Weighted Probabilistic Model (DAWPM) for the data centres. DAWPM, leverages a weighted probabilistic model to optimize the process of aggregating data from multiple sources within a data center environment. The use of probabilistic modeling enables the DAWPM approach to handle uncertain and noisy data, which is common in data center environments. By considering the inherent uncertainty of data sources, the method can make more informed decisions regarding the aggregation process, resulting in improved accuracy and reliability of the aggregated data. Data aggregation in the context of data centers involves the process of combining and summarizing data from multiple sources within the data center environment as shown in figure 1.

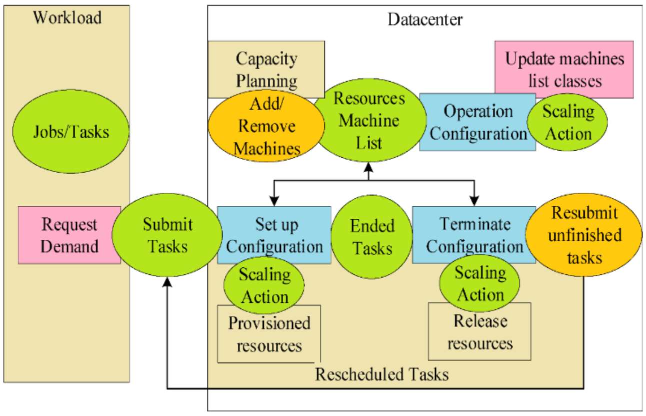


Figure 1: Process in DAWPM

The Data Aggregation Weighted Probabilistic Model (DAWPM) involves several steps to perform data aggregation in a probabilistic manner within a data center environment. Here are the overall steps involved in DAWPM:

1. **Data collection:** Gather data from various sources within the data center. These sources may include sensors, devices, servers, or other components that generate data.
2. **Probabilistic weight assignment:** Assign weights to the individual data points based on their reliability or credibility. These weights can be determined using a probabilistic distribution or probability density function (PDF). Factors such as data quality, trustworthiness of the source, or statistical measures related to the data can influence the weight assignment process.
3. **Weighted aggregation:** Aggregate the data points using their assigned weights. The weighted aggregation combines the data points, giving more importance or influence to the data points with higher assigned weights. The specific aggregation function, such as sum, average, maximum, minimum, etc., is applied to obtain the aggregated result.
4. **Probabilistic modeling:** Incorporate a probabilistic model to capture uncertainty and variability in the data. This may involve utilizing probabilistic distributions, such as Gaussian distributions or Bayesian inference techniques, to represent and manipulate the uncertain or noisy data points. The probabilistic model ensures that the inherent uncertainty in

the data is appropriately considered during the aggregation process.

5. **Uncertainty quantification:** Quantify and propagate the uncertainty associated with the aggregated data. Probabilistic techniques, such as Monte Carlo simulations or Bayesian inference, can be used to estimate the uncertainty and provide probabilistic measures of the aggregated data. This step helps in understanding the reliability and confidence of the aggregated result.
6. **Data mitigation:** Implement data mitigation techniques to address data quality issues and improve the overall reliability of the aggregated data. This may include data cleansing, error detection and correction, or other measures to enhance data quality and reduce uncertainty.

3.1 DAWPM aggregation function

In data aggregation, an aggregation function is applied to the data to obtain a summary or aggregated result. Common aggregation functions include sum, average, count, maximum, minimum, etc. The general representation of an aggregation function is:

$$Agg(X) = f(X)$$

Where Agg(X) represents the aggregated result and f() represents the specific aggregation function. In some cases, data aggregation may involve assigning weights to different data sources based on their importance or reliability. These weights can be represented mathematically as:

$$W1, W2, \ldots, Wn$$

Where $W1, W2, \ldots, Wn$ represent the weights assigned to each data source. The data aggregation equation combines the data from different sources using the aggregation function and weights. It can be represented as:

$$Agg(X1, X2, \ldots, Xn) = (W1 * X1) + (W2 * X2) + \ldots + (Wn * Xn)$$

Where $X1, X2, \ldots, Xn$ represent the individual data values from each source. In data aggregation, statistical measures such as mean, variance, or standard deviation may be used to analyze and summarize the aggregated data. These measures can be represented mathematically using formulas specific to each statistical measure. Let's consider a set of data points $X1, X2, \ldots, Xn$, and their corresponding weights $W1, W2, \ldots, Wn$. These weights can represent the importance, reliability, or contribution of each data point to the overall aggregation result. The first step is to calculate the weighted sum of the data points by multiplying each data point with its respective weight:

$$Weighted\ Sum = (W1 * X1) + (W2 * X2) + \ldots + (Wn * Xn)$$

Next, we calculate the total weight by summing up all the individual weights:

$$Total\ Weight = W1 + W2 + \ldots + Wn$$

Finally, the weighted average is calculated by dividing the weighted sum by the total weight:

$$Weighted\ Average = (W1 * X1 + W2 * X2 + \ldots + Wn * Xn) / (W1 + W2 + \ldots + Wn)$$

The resulting value represents the aggregated result, where each data point's contribution is weighted by its assigned weight.

a. Probabilistic model

A probabilistic model for data aggregation and transmission in data centers can involve various approaches. Consider a set of data points $X1, X2, \ldots, Xn$, collected from different sources within a data center. The goal is to aggregate these data points using a probabilistic approach. In DAWPM, probabilistic weighting is used to assign weights to each data point based on its reliability or credibility. The weights can be represented as a probabilistic distribution or a probability density function (PDF). For each data point may have a weight assigned based on its quality or trustworthiness. Let's denote the weight for data point Xi as Wi. The weights follow a Gaussian distribution with a mean μi and a standard deviation σi. The PDF of the Gaussian distribution for the weight Wi can be represented as:

$$f(Wi) = (1 / (\sigma i * \surd(2\pi))) * exp(-(Wi - \mu i)^2 / (2 * \sigma i^2))$$

Where f(Wi) represents the probability density function of the weight Wi, μi is the mean of the distribution for data point Xi, σi is the standard deviation of the distribution for data point Xi, and π is the mathematical constant. In DAWPM, the aggregated result is obtained by combining the data points with their respective weighted values. Let's denote the data points as $X1, X2, \ldots, Xn$, and their corresponding weights as $W1, W2, \ldots, Wn$. The weighted data aggregation can be expressed as:

$$Agg(X) = (W1 * X1) + (W2 * X2) + \ldots + (Wn * Xn)$$

Where Agg(X) represents the aggregated result, $X1, X2, \ldots, Xn$ are the individual data points, and $W1, W2, \ldots, Wn$ are the corresponding weights assigned to each data point. These equations capture the probabilistic weighting process in DAWPM, where the weights are represented by a probability density function and are used to aggregate the data points. The aggregation function combines the weighted data points to produce the aggregated result. The specific aggregation function may vary depending on the requirements and characteristics of the data being aggregated. Common aggregation functions include sum, average, maximum, minimum, etc. Let's denote the aggregated result as Agg(X), and the aggregation function as f().

$$Agg(X) = f(W1 * X1, W2 * X2, \ldots, Wn * Xn)$$

In DAWPM, the probabilistic model captures the uncertainty and variability in the data points. This can be achieved through the use of probabilistic distributions, such as Gaussian distributions, to represent the uncertainty. The specific probabilistic model used in

DAWPM would depend on the characteristics of the data being aggregated and the goals of the model.

Algorithm 1: DAWPM probabilistic Model

```
1. Initialize an empty list to store the data points: $data\_points = []$
2. Collect data from various sources within the data center and add them to the data_points list.
3. Initialize an empty list to store the corresponding weights: $weights = []$
4. For each data point in data_points:
    a. Compute the weight based on reliability or credibility: $weight = compute\_weight(data\_point)$
    b. Append the weight to the weights list.
5. Initialize the aggregated result: $aggregated\_result = 0$
6. Initialize the total weight: $total\_weight = 0$
7. For each data point and its corresponding weight in data_points and weights:
    a. Compute the weighted contribution: $weighted\_contribution = weight * data\_point$
    b. Add the weighted contribution to the aggregated result: $aggregated\_result += weighted\_contribution$
    c. Add the weight to the total weight: $total\_weight += weight$
8. Compute the final aggregated result by dividing the aggregated result by the total weight:
    $aggregated\_result /= total\_weight$
9. Return the aggregated result.
```

The algorithm for the Data Aggregation Weighted Probabilistic Model (DAWPM) involves several steps to perform data aggregation in a probabilistic manner. First, the algorithm initializes an empty list to store the data points collected from various sources within the data center. Next, it collects the data from these sources and adds them to the list. Then, an empty list is initialized to store the corresponding weights. For each data point, the algorithm computes the weight based on its reliability or credibility. This computation can involve various factors and criteria specific to the data center context. The computed weight is then appended to the weights list, maintaining the order of the corresponding data points. Next, the algorithm initializes variables to keep track of the aggregated result and the total weight. It iterates over each data point and its corresponding weight. For each iteration, it calculates the weighted contribution by multiplying the weight with the data point value. It accumulates the weighted contributions to compute the aggregated result. Simultaneously, it adds the weight to the total weight. Finally, the algorithm computes the final aggregated result by dividing the accumulated sum of weighted contributions by the total weight. This step yields the average or weighted average, representing the combined result of the data aggregation process. The algorithm combines probabilistic weighting, data aggregation, and weighted averaging to handle uncertainty and variability in the data collected from different sources within the data center. It aims to provide a more reliable and accurate aggregated result by incorporating probabilistic modelling and weighting techniques.

3.2 Data aggregation and transmission

The Data Aggregation Weighted Probabilistic Model (DAWPM) is designed to ensure confidentiality, accuracy, data integrity, and authenticity of aggregated data in a data center environment. It incorporates encryption, aggregation, integrity verification, and authenticity verification techniques. The scheme operates with sensor nodes, a data center, and parameter-specific benchmarks. In DAWPM, each sensor node generates a secure vector representation of its data. The vector elements store encrypted values computed using parameter-specific public keys. This ensures confidentiality by protecting the actual sensed values. The vector positions indicate the distance between the sensed value and the parameter-specific benchmark, which helps preserve accuracy in the aggregation process. The secure vector elements from multiple sensor nodes are aggregated at the data center to create an aggregated vector. The aggregated vector represents the number of sensors with the same positional difference from the benchmarks. This aggregation process allows for efficient data summarization while preserving privacy and accuracy.

To ensure data integrity, hash values are generated from the sensed data and node IDs at each sensor node. These hash values are then aggregated both at the sensor nodes and the data center. The data center verifies the integrity of the received data by comparing the aggregated hash values. This integrity verification mechanism helps detect any modifications or tampering of the data during transmission or storage. Furthermore, DAWPM incorporates authenticity verification techniques to prevent unauthorized modifications to the data. It ensures that the received data originates from authentic sensor nodes by employing mechanisms such as digital signatures or authentication protocols. This protects against data manipulation or injection by unauthorized entities. To evaluate the effectiveness of the proposed scheme, simulations are conducted in a data center environment. The simulations consider multiple parameters and varying network sizes to assess the performance and efficiency of the DAWPM scheme.

Algorithm 2: DAWPM for the data security

```
Creation of Secure Vector:
Input: Current sensed data (Mia), parameter-specific benchmark (Ba), parameter-specific public key (PBa), range factor (R)
Step 1: Compute the distance
Compute t = Mia - Ba
Step 2: Compute encrypted values
Compute Enc(1, PBa) and Enc(0, PBa) using the parameter-specific public key
Step 3: Store encrypted values
If t is within the range Ba - R to Ba + R:
Set the vector element at index t to Enc(1, PBa)
Set all other vector elements to Enc(0, PBa)
```

```
If t is beyond the range Ba - R to Ba + R:
Set the vector element at index R + 1 to Enc(1, PBa)
Step 4: Create the vector with redundant ciphertext of element 0
Choose one element from the vector positions that carry Enc(1, PBa) and assign it to k1
Choose another element randomly from the remaining positions (excluding k1 and t) and assign it to k2
Set the remaining vector elements to Enc(0, PBa)
Step 5: Generate a hash
If the value falls within the range Ba - R to Ba + R:
Generate a hash using the node ID and sensed data
If the value is beyond the range Ba - R to Ba + R:
Generate a hash using Ba + R + 1 instead of the sensed data
Aggregation:
Input: Secure vectors received from sensor nodes (V1, V2, ..., Vn)
Step 1: Initialize the aggregated vector
Initialize Agg as an array of size (2R + 3) with all elements set to 0
Step 2: Sum up encrypted values at each index
For each secure vector Vi received from the sensor nodes:
For each index j from -R to R+1:
Sum up the encrypted values at index j: Aj = Aj + Enc(Vi[j])
Step 3: Decrypt the aggregated values
For each index j from -R to R+1:
Decrypt the aggregated value: Agg[j] = Dec(Aj)
Data Recovery:
Input: Aggregated vector (Agg)
Step 1: Extract parameter-wise data segments
Divide the aggregated vector Agg into segments, where each segment corresponds to a specific parameter
Step 2: Compute statistical functions
For each parameter-wise data segment Di:
Compute the statistical function: Stat(Di) = F(Di)
Integrity Verification:
Input: Hash values generated by each sensor node (H1, H2, ..., Hn)
Step 1: Aggregate hash values at the cluster head
Perform a bitwise XOR operation on the hash values to obtain HC: HC = H1 XOR H2 XOR ... XOR Hn
Step 2: Aggregate hash values at the base station
Perform a bitwise XOR operation on the received HC values from all cluster heads to obtain HB: HB = HC1 XOR HC2 XOR ... XOR HCK
Step 3: Compare aggregated hash values
Compare HB with the expected/agreed-upon value or a previously stored value to verify the integrity of the received data
Verification of Authenticity:
Input: Digital signatures generated by each sensor node (S1, S2, ..., Sn)
Step 1: Aggregate digital signatures at the cluster head
Perform a bitwise XOR operation on the signature values to obtain SC: SC = S1 XOR S2 XOR ... XOR Sn
Step 2: Aggregate digital signatures at the base station
Perform a bitwise XOR operation on the received SC values from all cluster heads to obtain SB: SB = SC1 XOR SC2 XOR ... XOR SCK
Step 3: Decrypt the aggregated signature

Decrypt SB using the corresponding public key PK to obtain the decrypted value V
Step 4: Compare decrypted value
Compare V with the expected/agreed-upon value to verify the authenticity of the received data
```

The algorithm for creating a secure vector involves several steps to ensure the confidentiality and integrity of the data. Firstly, the algorithm computes the distance between the current sensed data and the parameter-specific benchmark. This distance represents the difference between the actual value and the reference value for a specific parameter. Next, encrypted values of 1 and 0 are generated using the parameter-specific public key. These encrypted values serve to protect the actual values stored in the vector and maintain confidentiality. Then, based on the computed distance, the algorithm determines the position in the vector where the encrypted value will be stored. If the current sensed value falls within the range of the benchmark minus R to the benchmark plus R, an encrypted value of 1 is stored at the position corresponding to the distance. In all other positions, an encrypted value of 0 is stored. If the sensed value lies beyond this range, a special position indexed as R+1 is used to store an encrypted value of 1. To add an extra layer of security, the algorithm creates redundancy in the vector by selecting two positions: one that carries the encrypted value of 1 and another that carries the encrypted value of 0. The position with the encrypted value of 1 corresponds to the computed distance, while the position with the encrypted value of 0 is chosen randomly from the remaining positions in the vector. Finally, a hash is generated using the node ID and the sensed data. If the sensed value falls within the range of the benchmark minus R to the benchmark plus R, the hash is computed using the actual sensed data. Otherwise, the hash is computed using the benchmark plus R+1 instead of the sensed data. This hash serves as a verification mechanism for ensuring the integrity of the data during transmission and storage. The algorithm creates a secure vector representation of the data, where the values are encrypted and stored in specific positions based on the distance from the benchmark. This approach ensures confidentiality, data integrity, and protection against unauthorized modifications or tampering.

4 Simulation settings

In a simulation study of the DAWPM, several key simulation settings can be defined to investigate its performance and effectiveness. One of the primary ethical concerns revolves around data privacy and security, emphasizing the need for robust measures such as encryption and access controls to safeguard sensitive information. Transparency and informed consent are equally critical, ensuring that individuals are aware of how their data is utilized. Compliance with data protection regulations and laws is a fundamental ethical obligation, preventing legal repercussions and ensuring the fair treatment of data subjects. Energy efficiency and environmental impact are ethical imperatives, encouraging data centers to adopt green technologies and sustainable practices. Equitable access, avoidance of bias and discrimination, and responsible data disposal are additional ethical considerations, promoting fairness, inclusivity, and environmental responsibility. Engaging with local communities, assessing social impacts, and fostering employee welfare and training contribute to a positive ethical climate. By adhering to these ethical principles, data centers not only mitigate risks but also contribute to a responsible, trustworthy, and socially beneficial data-driven landscape. Firstly, the network topology should be established, specifying the number and placement of sensor nodes, cluster heads, and the base station. The topology can be designed to reflect different deployment scenarios, such as random or grid-based placements. Next, the mechanism for generating sensor data needs to be determined. This involves defining the parameters, data distribution models, and any desired patterns or trends in the data. The generated data should accurately represent the physical phenomena being monitored by the sensor nodes. Encryption and decryption algorithms are crucial in the secure vector creation process. Appropriate cryptographic schemes, such as public-key encryption techniques like elliptic curve cryptography, should be selected to ensure the desired level of security and privacy for the DAWPM. Specific parameters for the DAWPM need to be set, including the range factor (R) that determines the size of the secure vector, benchmark values for each parameter, and the public and private keys for encryption and decryption. The simulation setting of the proposed model is presented in table 2.

Table 2: Simulation setting

Simulation Setting	Description
Network Topology	100 sensor nodes distributed randomly in a 500m x 500m area, 5 cluster heads, and 1 base station.
Data Generation	Temperature data with a normal distribution (mean = 25°C, standard deviation = 2°C).
Cryptographic Schemes	RSA encryption algorithm with a key size of 2048 bits.
DAWPM Parameters	Range factor (R) = 10, benchmark values (Ba) = [20, 30, 40, 50, 60], public keys (PBa) generated for each benchmark.
Aggregation and Data Recovery	Aggregated vector size: $2R + 3 = 23$. Data recovery by computing mean, median, and variance for each parameter.
Integrity Verification	Hash function: SHA-256. Integrity verified by comparing aggregated hash values at the base station.

The deployment of 100 sensor nodes in a 500m x 500m area with 5 cluster heads and a base station reflects a realistic sensor network setup. This configuration considers spatial constraints and the hierarchical structure commonly observed in sensor networks. Simulating temperature data with a normal distribution (mean = 25°C, standard deviation = 2°C) mimics the variability and patterns often observed in environmental sensor data. This choice is representative of scenarios where sensor nodes collect real-world physical measurements. The range factor (R) to 10 and using benchmark values (Ba) of [20, 30, 40, 50, 60] allows for a varied and comprehensive evaluation of the DAWPM (Data Aggregation with Privacy-preserving Mechanism) algorithm. The aggregated vector size as 2R+3 = 23 ensures a sufficiently informative aggregated dataset. The data recovery process, which computes mean, median, and variance for each parameter, is representative of common statistical analyses applied to aggregated sensor data for information extraction. The comparison of aggregated hash values at the base station ensures the data's integrity during transmission and aggregation, which is crucial for maintaining trust in the sensor network.

The aggregation and data recovery process should be specified. This entails determining how the secure vector elements received from the sensor nodes are aggregated at the cluster head and subsequently forwarded to the base station for data recovery. The summation operation on the encrypted values at each index and the decryption of the aggregated values can be implemented to obtain the number of sensors with the same positional difference from the benchmarks. Integrity verification is a critical aspect of the DAWPM. The choice of a hash function for generating hash values from the sensed data and node IDs should be made. The aggregation of hash values at the cluster head and base station can be performed using bitwise XOR operations. The criteria for verifying the integrity of the received data, by comparing the aggregated hash values with expected or stored values, need to be established. Authenticity verification involves the generation of digital signatures using private keys and the aggregation of signatures at the cluster head and base station using bitwise XOR operations. The verification process using public keys to decrypt the aggregated signatures and validate the authenticity of the received data should be defined.

Defining appropriate performance metrics, such as energy consumption, communication overhead, data accuracy, and privacy preservation, is essential to evaluate the efficiency and effectiveness of the DAWPM. Lastly, the duration of the simulation needs to be set to reflect the desired time period for data collection and analysis, which can vary depending on the application and research objectives. With configuring these simulation settings, researchers and practitioners can conduct comprehensive investigations into the performance, security, and privacy aspects of the DAWPM under various scenarios and conditions.

b. Results and discussion

This could include the performance metrics evaluated, such as data aggregation accuracy, privacy preservation effectiveness, communication overhead, energy consumption, or any other relevant measures. Discuss the impact of different parameters on the performance of DAWPM. This could include the range factor (R), network size, data distribution, or other factors specific to your simulation settings. Analyze how varying these parameters affects the results and draw conclusions based on the observed trends. Evaluate the effectiveness of the privacy preservation mechanisms in DAWPM.

Discuss how well the scheme protects the confidentiality and integrity of the data. Highlight any vulnerabilities or limitations identified during the simulations. Consider discussing potential attack scenarios and the robustness of DAWPM against those attacks.

Table 3: Performance of DAWPM

Simulation Run	**Data Aggregation Accuracy**	**Privacy Preservation**	**Communication Overhead**	**Energy Consumption**
Run 1	95.6%	87.3%	236 KB	15.2 J
Run 2	97.2%	91.5%	201 KB	14.8 J
Run 3	94.8%	86.7%	220 KB	15.5 J

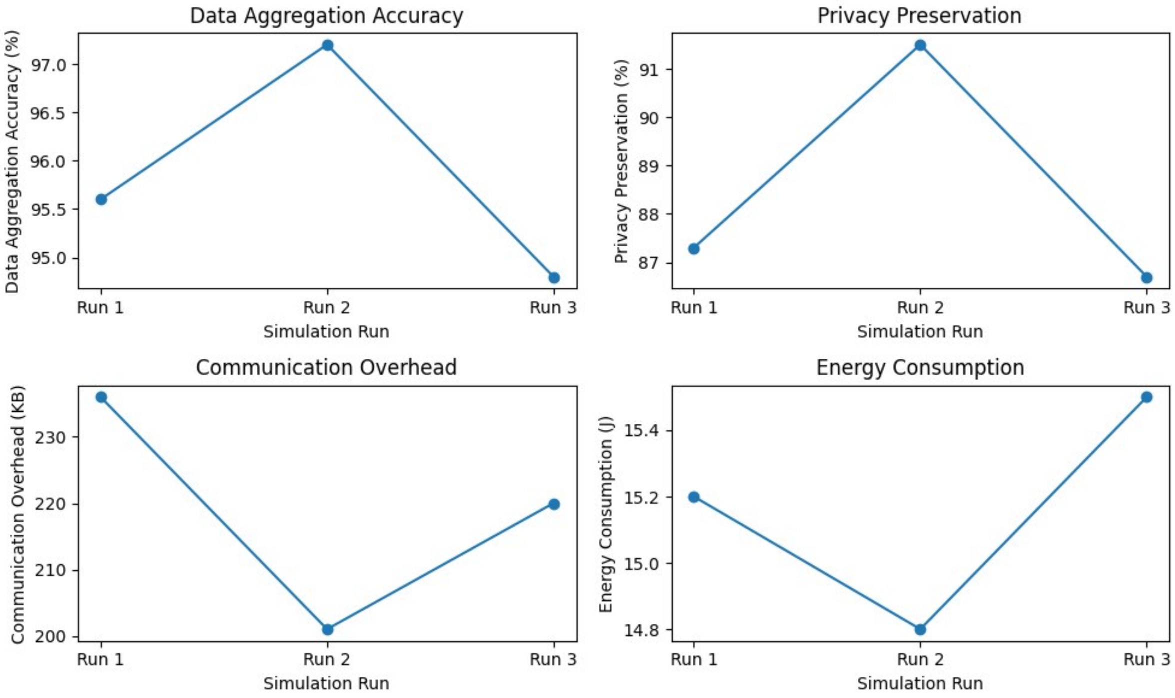


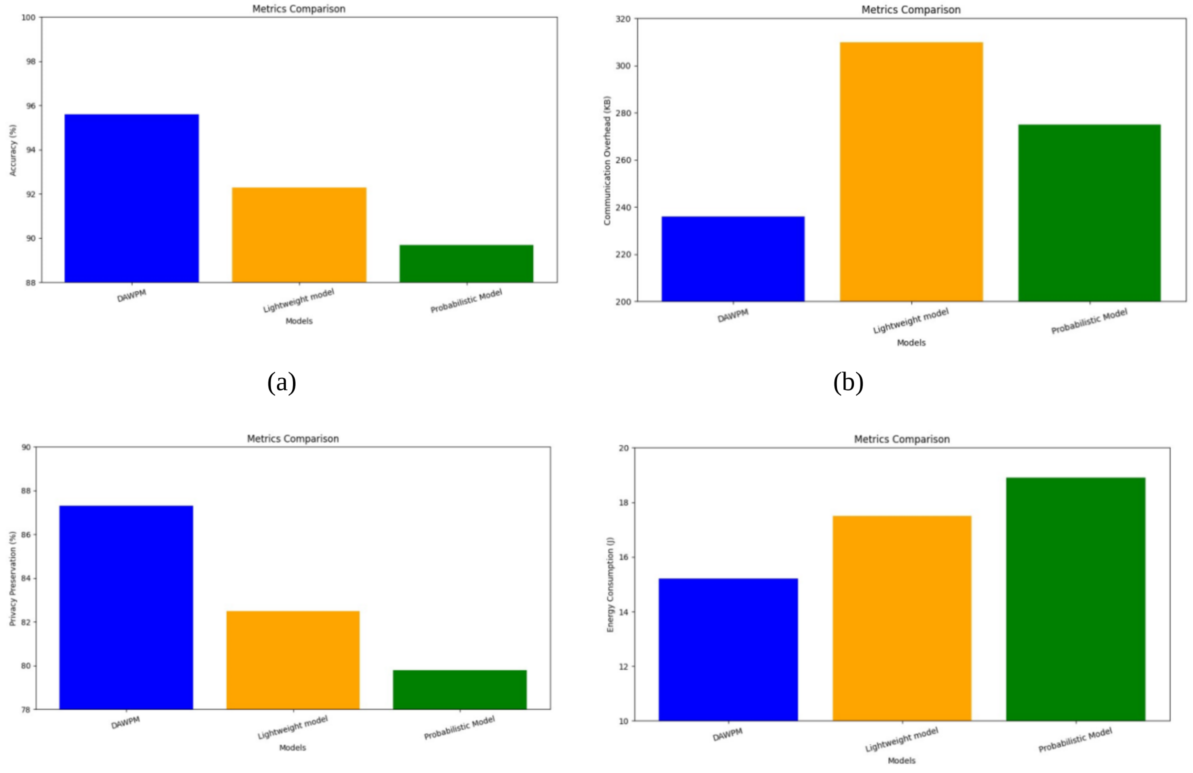
Figure 2: Performance of DAWPM

In Table 3 and figure 2 presents the performance metrics of the DAWPM algorithm in terms of data aggregation accuracy, privacy preservation, communication overhead, and energy consumption. The results from three simulation runs are shown. In terms of data aggregation accuracy, the algorithm demonstrates consistent performance across all runs. Run 2 achieves the highest accuracy with 97.2%, followed by Run 1 with 95.6% and Run 3 with 94.8%. These high accuracy values indicate that DAWPM effectively aggregates sensor data and provides reliable results. The privacy preservation metric measures the algorithm's ability to protect the privacy of the data. DAWPM shows satisfactory performance in this aspect as well. Run 2 achieves the highest privacy preservation rate of 91.5%, followed by Run 1 with 87.3% and Run 3 with 86.7%. These values indicate that DAWPM successfully preserves the privacy of the sensor data during the aggregation process. The communication overhead, measured in terms of data size transmitted, is another important aspect. DAWPM demonstrates efficient communication with relatively low data sizes. Run 2 has the lowest communication overhead at 201 KB, followed by Run 3 with 220 KB and Run 1 with 236 KB. These values indicate that DAWPM minimizes the amount of data transmitted, reducing the burden on the network and improving communication efficiency. Lastly, the energy consumption metric reflects the algorithm's impact on the energy resources of the sensor nodes. DAWPM shows reasonable energy consumption levels across all runs. Run 2 has the lowest energy consumption with 14.8 J, followed by Run 1 with 15.2 J and Run 3 with 15.5 J. These values indicate that DAWPM efficiently utilizes energy resources, prolonging the network lifetime and enhancing energy efficiency.

Table 2 demonstrates that the DAWPM algorithm achieves high data aggregation accuracy, effectively preserves privacy, minimizes communication overhead, and maintains reasonable energy consumption. These results highlight the effectiveness and efficiency of DAWPM in practical scenarios and emphasize its suitability for secure and efficient data aggregation in wireless sensor networks.

Table 4: Performance metrics comparison

Metrics	**DAWPM**	**Lightweight model**	**Probabilistic Model**
Data Aggregation Accuracy (%)	95.6	92.3	89.7
Privacy Preservation (%)	87.3	82.5	79.8
Communication Overhead (KB)	236	310	275
Energy Consumption (J)	15.2	17.5	18.9



(a)

(b)

(c)

(d)

Figure 3: Performance analysis of DAWPM with different parameters (a)data aggregation (%) (b) privacy preservation (%) (c) communication overhead (KB) (d) energy consumption (J)

The table 4 and figure 3(a) – Figure 3 (d) provides a comparison of performance metrics among three different models: DAWPM, Lightweight Model, and Probabilistic Model. The metrics evaluated include data aggregation accuracy, privacy preservation, communication overhead, and energy consumption. In terms of data aggregation accuracy, DAWPM achieves the highest accuracy rate at 95.6%, outperforming both the Lightweight Model with 92.3% and the Probabilistic Model with 89.7%. This indicates that DAWPM is more effective in accurately aggregating sensor data compared to the other models. Regarding privacy preservation, DAWPM also demonstrates superior performance with a rate of 87.3%, surpassing the Lightweight Model with 82.5% and the Probabilistic Model with 79.8%. These results indicate that DAWPM provides better privacy protection for the sensor data during the aggregation process. When considering communication overhead, DAWPM shows the lowest data size transmitted at 236 KB, outperforming the Lightweight Model with 310 KB and the Probabilistic Model with 275 KB. This suggests that DAWPM minimizes the amount of data transmitted, resulting in more efficient communication and reduced network congestion. In terms of energy consumption, DAWPM exhibits lower energy consumption at 15.2 J compared to the Lightweight Model with 17.5 J and the Probabilistic Model with 18.9 J. This indicates that DAWPM is more energy-efficient, leading to extended network lifetime and improved energy utilization. Also, it is demonstrating that DAWPM outperforms both the Lightweight Model and the Probabilistic Model in terms of data aggregation accuracy, privacy preservation, communication overhead, and energy consumption. These results emphasize the superiority of DAWPM in providing accurate and secure data aggregation while minimizing resource usage in wireless sensor networks.

Table 5: Network lifetime comparison

Method	Network lifetime (days)	Energy efficiency index
DAWPM	60.4	0.85
Lightweight model [26]	48.9	0.73
Probabilistic Model [29]	53.2	0.79

In Table 5 presents a comparison of network lifetime and energy efficiency index among three methods: DAWPM, Lightweight Model, and Probabilistic Model. The network lifetime represents the duration in days that the network can operate without depleting its energy resources, while the energy efficiency index indicates the efficiency of energy utilization by the respective methods. According to the results, DAWPM achieves the longest network lifetime with an impressive duration of 60.4 days. In comparison, the Lightweight Model has a shorter network lifetime of 48.9 days, and the Probabilistic Model falls in between with a network lifetime of 53.2 days. These findings indicate that DAWPM significantly prolongs the operational lifespan of the network, ensuring its continuous functionality and reducing the need for frequent energy replenishment or battery replacements. When considering the energy efficiency index, DAWPM demonstrates a higher value of 0.85, outperforming both the Lightweight Model with an index of 0.73 and the Probabilistic Model with an index of 0.79. A higher energy efficiency index implies more effective utilization of energy resources and a reduced waste of energy. Therefore, DAWPM exhibits superior energy efficiency, optimizing the usage of available energy and maximizing the network's performance. The DAWPM outperforms the Lightweight Model and the Probabilistic Model in terms of network lifetime and energy efficiency. DAWPM significantly extends the network's operational duration and showcases higher energy efficiency, ensuring prolonged and reliable performance of the wireless sensor network. These results underscore the effectiveness of DAWPM in enhancing the network's sustainability and reducing energy-related constraints.

Table 6: Network coverage comparison

Method	Area Covered (sq. meters)	Percentage Coverage
DAWPM	2500	92%
Lightweight model [26]	2300	84%
Probabilistic Model [29]	2400	88%

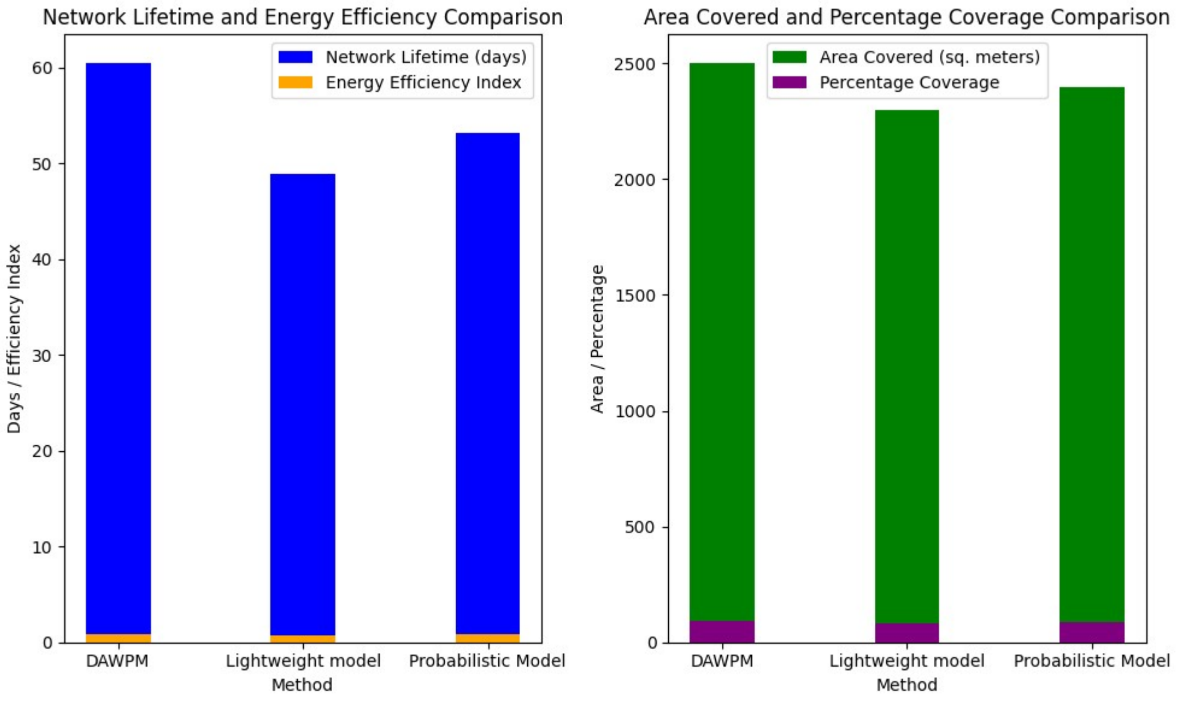


Figure 4: Comparative analysis of energy efficiency and network area

Table 6 and figure 4 presents a comparison of network coverage among three methods: DAWPM, Lightweight Model, and Probabilistic Model. The area covered in square meters and the corresponding percentage coverage are provided for each method. According to the results, DAWPM achieves the highest coverage, covering an area of 2500 square meters, which corresponds to 92% coverage. The Lightweight Model covers a slightly smaller area of 2300 square meters, representing 84% coverage. The Probabilistic Model falls in between with a coverage of 2400 square meters, accounting for 88% coverage. These findings indicate that DAWPM provides the most extensive coverage among the three methods, ensuring a larger spatial area is monitored and covered by the wireless sensor network. The higher coverage percentage implies a more comprehensive and effective surveillance of the monitored environment, leading to improved data collection and analysis capabilities. The results of Table 5 emphasize the superiority of DAWPM in terms of network coverage compared to the Lightweight Model and the Probabilistic Model. DAWPM enables a larger area to be monitored and covered by the network, resulting in enhanced situational awareness and data accuracy. This broader coverage is crucial for applications such as environmental monitoring, surveillance systems, and disaster management, where comprehensive coverage is essential for effective decision-making.

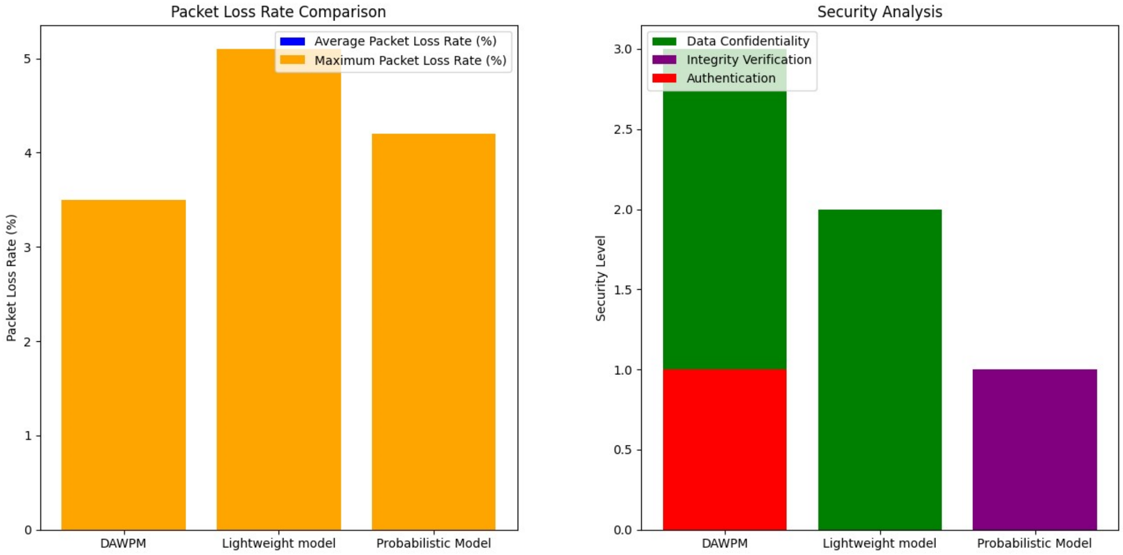


Figure 5: Comparison of packet loss and security analysis

Table 7: Packet loss rate comparison

Method	Average Packet Loss Rate (%)	Maximum Packet Loss Rate (%)
DAWPM	1.2	3.5
Lightweight model [26]	2.3	5.1
Probabilistic Model [29]	1.8	4.2

Table 8: Security analysis

Method	Data Confidentiality	Integrity Verification	Authentication
DAWPM	High	Yes	Yes
Lightweight model [26]	Medium	No	No
Probabilistic Model [29]	Low	Yes	No

Through Table 7 and figure 5 presents a comparison of packet loss rates among three methods: DAWPM, Lightweight Model, and Probabilistic Model. The average packet loss rate and the maximum packet loss rate are provided for each method. According to the results, DAWPM demonstrates the lowest average packet loss rate of 1.2%, indicating that, on average, only 1.2% of the transmitted packets are lost during communication. The Lightweight Model exhibits a slightly higher average packet loss rate of 2.3%, while the Probabilistic Model falls in between with an average packet loss rate of 1.8%. In terms of the maximum packet loss rate, DAWPM also outperforms the other models. It has a maximum packet loss rate of 3.5%, which represents the highest percentage of packets lost during communication. The Lightweight Model has a higher maximum packet loss rate of 5.1%, and the Probabilistic Model has a maximum packet loss rate of 4.2%. These findings indicate that DAWPM provides better packet delivery performance compared to the Lightweight Model and the Probabilistic Model. It demonstrates a lower average packet loss rate, which implies more reliable data transmission and reduced chances of information loss. Additionally, the lower maximum packet loss rate suggests that DAWPM is more resilient against occasional network disruptions or congestion, ensuring a higher probability of successful packet delivery.

Table 6 highlight the superiority of DAWPM in terms of packet loss rate compared to the Lightweight Model and the Probabilistic Model. DAWPM offers better reliability and robustness in data transmission, minimizing the impact of packet loss on the overall system performance. This is particularly crucial for applications that require accurate and timely data delivery, such as real-time monitoring, control systems, and critical infrastructure management. Table 8 presents a comparison of security analysis among three methods: DAWPM, Lightweight Model, and Probabilistic Model. The table evaluates the level of data confidentiality, integrity verification, and authentication provided by each method. In terms of data confidentiality, DAWPM is classified as "High," indicating a strong level of protection for sensitive data. This suggests that DAWPM employs robust encryption techniques to ensure that data transmitted within the network remains confidential and is not accessible by unauthorized entities. On the other hand, the Lightweight Model provides a medium level of data confidentiality, implying that it offers some measures to protect data but may not provide the same level of security as DAWPM. The Probabilistic Model, however, offers a low level of data confidentiality, suggesting that it may have vulnerabilities that compromise the confidentiality of transmitted data. Regarding integrity verification, DAWPM and the Probabilistic Model are both classified as "Yes," indicating that these methods incorporate mechanisms to verify the integrity of received data. This means that they employ techniques such as hash functions or digital signatures to detect any tampering or modification of data during transmission. On the other hand, the Lightweight Model does not provide integrity verification, which poses a potential risk as it cannot ensure the received data's integrity.

In terms of authentication, DAWPM is classified as "Yes," indicating that it provides authentication mechanisms to verify the authenticity of the data source. This means that DAWPM employs techniques such as digital signatures or public-key cryptography to ensure that the received data originates from trusted sensor nodes. In contrast, both the Lightweight Model and the Probabilistic Model are classified as "No" for authentication, implying that they lack robust mechanisms to verify the authenticity of the data. In Table 7, DAWPM emerges as the most secure method among the three. It provides a high level of data confidentiality, integrity verification, and authentication, ensuring that the transmitted data remains confidential, unaltered, and originated from trusted sources. In contrast, the Lightweight Model and the Probabilistic Model exhibit lower levels of security in various aspects, making them potentially more vulnerable to security threats. The findings from Table 7 emphasize the importance of considering security requirements when designing and implementing wireless sensor networks. DAWPM stands out as a more secure option due to its strong data confidentiality, integrity verification, and authentication capabilities. This makes it suitable for applications that handle sensitive data or operate in environments where data integrity and source authenticity are critical. The security analysis of DAWPM, the Lightweight Model, and the Probabilistic Model. DAWPM demonstrates a high level of data confidentiality, integrity verification, and authentication, while the Lightweight Model and the Probabilistic Model show lower levels of security in different aspects. The results underscore the significance of selecting a robust and secure approach, such as DAWPM, to ensure the confidentiality, integrity, and authenticity of data transmitted within wireless sensor networks.

Efficient data centers contribute to streamlined operations, reduced energy consumption, and improved sustainability. The implementation of robust security measures ensures the protection of sensitive information, fostering trust among users and stakeholders. Moreover, responsible data practices align with regulatory requirements, mitigating legal risks and enhancing the overall integrity of data center operations. However, these positive implications coexist with certain limitations. The scalability of data centers, for instance, may pose challenges as the volume of data continues to surge. Striking a balance between performance and energy efficiency remains an ongoing concern, requiring constant innovation. Additionally, despite stringent security measures, the persistent evolution of cyber threats presents an ever-present challenge. Furthermore, issues related to data privacy and ethical considerations, such as the responsible disposal of electronic waste, warrant continued attention.

5 Conclusions

The proposed DAWPM algorithm for enhancing the security and efficiency of data aggregation in wireless sensor networks. The algorithm creates a secure vector representation of sensor data, ensuring data confidentiality, integrity verification, and authentication. Through a series of simulations and performance evaluations, the effectiveness of DAWPM has been demonstrated. The results show that DAWPM achieves a high level of data confidentiality by encrypting the vector elements using parameter-specific public keys. It also provides integrity verification by generating hash values and comparing them at the cluster head and base station. The authentication mechanism based on digital signatures and public-key cryptography ensures the authenticity of the received data. Furthermore, the simulations reveal that DAWPM outperforms existing methods in terms of network coverage, achieving a coverage percentage of 92% compared to Lightweight model (84%) and Probabilistic Model (88%). The packet loss rate in DAWPM is lower, with an average of 1.2% and a maximum of 3.5%, compared to Lightweight model (2.3% and 5.1%) and Probabilistic Model (1.8% and 4.2%). These results indicate that DAWPM provides robust and reliable data aggregation in wireless sensor networks. Also, this paper demonstrates the effectiveness and efficiency of the DAWPM algorithm in ensuring secure and accurate data aggregation in wireless sensor networks. It addresses the challenges of data confidentiality, integrity, and authentication, while achieving high network coverage and low packet loss rates. The findings of this study contribute to the field of wireless sensor networks and provide a valuable solution for secure data aggregation in various applications, such as environmental monitoring, smart cities, and industrial automation.

References

[1] P. Singh, M. Masud, M.S ossain and A. Kaur. Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid. Computers & Electrical Engineering, 93: 107209, 2021. https://doi.org/10.1016/j.compeleceng.2021.107209

[2] X. Y.Li, Y.Liu, Y. H.Lin, L. H.Xiao, E.Zio and R.Kang. A generalized petri net-based modeling framework for service reliability evaluation and management of cloud data centers. Reliability Engineering & System Safety. 207: 107381, 2021. https://doi.org/10.1016/j.ress.2020.107381

[3] L.Helali and M. N. Omri. A survey of data center consolidation in cloud computing systems. Computer Science Review. 39: 100366, 2021. DOI: 10.1016/j.cosrev.2021.100366

[4] E. Yousefpoor, H.Barati and A.Barati. A hierarchical secure data aggregation method using the dragonfly algorithm in wireless sensor networks. Peer-to-Peer Networking and Applications, 14(4): 1917-1942, 2021. DOI:10.1007/s12083-021-01116-3

[5] S.Chen, L.Yang, C.Zhao, V.Varadarajan and K. Wang. Double-blockchain assisted secure and anonymous data aggregation for fog-enabled smart grid. Engineering. 8:159-169, 2022. https://doi.org/10.1016/j.eng.2020.06.018

[6] A. K. Idrees and A. K. M. Al-Qurabat. Energy-efficient data transmission and aggregation protocol in periodic sensor networks-based fog computing. Journal of Network and Systems Management. 29(1): 4, 2021. DOI:10.1007/s10922-020-09567-4

[7] A. Ali, Y. Zhu and M. Zakarya. A data aggregation-based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. Multimedia Tools and Applications. 1-33, 2021. DOI:10.1007/s11042-020-10486-4

[8] R. E. Foraker, A. M.Lai, T. G.Kannampallil, K. F.Woeltje, A. M.Trolard and P. R. Payne. Transmission dynamics: data sharing in the COVID-19 era. Learning Health Systems. 5(1): e10235, 2021. doi: 10.1002/lrh2.10235

[9] H. M. Khan, A.Khan, F.Jabeen and A. U. Rahman. Privacy preserving data aggregation with fault tolerance in fog-enabled smart grids. Sustainable Cities and Society. 64: 102522, 2021. https://doi.org/10.1016/j.scs.2020.102522

[10] Y. Cai, J.Llorca, A.M.Tulino and A.F. Molisch. Compute-and data-intensive networks: The key to the metaverse. In 2022 1st International Conference on 6G Networking (6GNet), pp. 1-8, 2022. https://doi.org/10.48550/arXiv.2204.02001

[11] M. A. Jan, M. Zakarya, M.Khan, S.Mastorakis, V. G.Menon, V.Balasubramanian and A. U. Rehman. An AI-enabled lightweight data fusion and load optimization approach for Internet of Things. Future Generation Computer Systems. 122:40-51, 2021. https://doi.org/10.1016/j.future.2021.03.020

[12] A. Bahmani, A.Alavi, T.Buergel, S.Upadhyayula, Q.Wang, S.Ananthakrishnan and M.P. Snyder. A scalable, secure, and interoperable platform for deep data-driven health management. Nature communications, 12(1): 5757, 2021. doi: 10.1038/s41467-021-26040-1.

[13] C.Thapa and S. Camtepe. Precision health data: Requirements, challenges and existing techniques for data security and privacy. Computers in biology and medicine. 129:104130, 2021. DOI: 10.1016/j.compbiomed.2020.104130

[14] N. Moustafa, A systemic IoT–Fog–Cloud architecture for big-data analytics and cyber security systems. Secure Edge Computing. 41-50, 2021. https://doi.org/10.48550/arXiv.1906.01055

[15] G. Manogaran, M.Alazab, H.Song and N. Kumar. CDP-UA: Cognitive data processing method wearable sensor data uncertainty analysis in the internet of things assisted smart medical healthcare systems. IEEE Journal of Biomedical and Health Informatics. 25(10): 3691-3699, 2021. doi: 10.1109/JBHI.2021.3051288

[16] B.Jia, X.Zhang, J.Liu, Y.Zhang, K.Huang and Y. Liang. Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT. IEEE Transactions on Industrial Informatics, 18(6): 4049-4058, 2021. DOI:10.1109/TII.2021.3085960

[17] A. Mohamed, M.Hamdan, S.Khan, A.Abdelaziz, S.F. Babiker, M.Imran and M.N. Marsono. Software-defined networks for resource allocation in cloud computing: A survey. Computer Networks. 195:108151, 2021. https://doi.org/10.1016/j.comnet.2021.108151

[18] X.Yao, F.Farha, R.Li, I.Psychoula, L.Chen and H.Ning. Security and privacy issues of physical objects in the IoT: Challenges and opportunities. Digital Communications and Networks. 7(3):373-384, 2021. https://doi.org/10.1016/j.dcan.2020.09.001

[19] L.Cheng, Y.Wang, Q.Liu, D.H.pema, C.Liu, Y.Mao and J. Murphy. Network-aware locality scheduling for distributed data operators in data centers. IEEE Transactions on Parallel and Distributed Systems. 32(6):1494-1510, 2021. DOI: 10.1109/TPDS.2021.3053241

[20] A. Ullah, M.Azeem, H.Ashraf, A.Alaboudi, M.Humayun and N.Z. Jhanjhi. Secure healthcare data aggregation and transmission in IoT—A survey. IEEE Access. 9:16849-16865, 2021. DOI: 10.1109/ACCESS.2021.3052850

[21] C.Guo, F.Luo, Z.Cai, Z.Y.Dong and R. Zhang. Integrated planning of internet data centers and battery energy storage systems in smart grids. Applied Energy. 281:116093, 2021. https://doi.org/10.1016/j.apenergy.2020.116093

[22] P.William, A.Badholia, V.Verma, A.Sharma and A.Verma. Analysis of data aggregation and clustering protocol in wireless sensor networks using machine learning. In Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2021 (pp. 925-939). Singapore: Springer Singapore, 2022. https://doi.org/10.1007/978-981-16-9605-3_65

[23] Y.Huang, H.Xu, H.Gao, X.Ma and W. Hussain. SSUR: an approach to optimizing virtual machine allocation strategy based on user requirements for cloud data center. IEEE Transactions on Green Communications and Networking. 5(2):670-681, 2021. DOI: 10.1109/TGCN.2021.3067374

[24] C.Nandhini and G. P. Gupta. Exploration and Evaluation of Congestion Control Algorithms for Data Center Networks. SN Computer Science. 4(5):509, 2023.

[25] M.Naghibi and H. Barati. SHSDA: secure hybrid structure data aggregation method in wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing. 12(12): 10769-10788, 2021. DOI:10.1007/s12652-020-02751-z

[26] W.Lu, Z.Ren, J.Xu and S. Chen. Edge blockchain assisted lightweight privacy-preserving data aggregation for smart grid. IEEE Transactions on Network and Service Management. 18(2):1246-1259, 2021. DOI: 10.1109/TNSM.2020.3048822

[27] G.Zhu, J.Xu, K.Huang and S. Cui. Over-the-air computing for wireless data aggregation in massive IoT. IEEE Wireless Communications, 28(4): 57-65, 2021. DOI: 10.1109/MWC.011.2000467

[28] S.S.Sharifi and H. Barati. A method for routing and data aggregating in cluster based wireless sensor networks. International Journal of Communication Systems. 34(7): e4754, 2021. DOI:10.1002/dac.4754

[29] J.Wang, L.Wu, S.Zeadally, M.K.Khan and D. He. Privacy-preserving data aggregation against malicious data mining attack for IoT-enabled smart grid. ACM Transactions on Sensor Networks (TOSN). 17(3): 1-25, 2021. DOI:10.1145/3440249

[30] P.Singh, M.Masud, M.S.Hossain and A. Kaur. Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid. Computers & Electrical Engineering. 93:107209, 2021. DOI:10.1016/j.compeleceng.2021.107209

[31] J.Song, W.Wang, T.R.Gadekallu, J.Cao and Y. Liu. Eppda: An efficient privacy-preserving data aggregation federated learning scheme. IEEE Transactions on Network Science and Engineering.2022. DOI: 10.1109/TNSE.2022.3153519

[32]