International Journal of Communication Networks and Information Security
2024, 16(1), 6263
ISSN: 2073-607X,2076-0930
https://ijcnis.org/
Research Article

Hybrid Compressed Sensing and Secure Fault Tolerant Data Aggregation in Wireless Sensor Networks

Shwetha G R [1], Murthy S V N [2*]

[1,2] Department of Computer Science and Engineering, SJC Institute of Technology, Chickballapur, Visvesvaraya Technological University, Belagavi, India
***Corresponding Author:** dr.svnmurthy@gmail.com

Citation: S. G. R and M. S. V. N, "Hybrid Compressed Sensing and Secure Fault Tolerant Data Aggregation in Wireless Sensor Networks," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 16, no. 1, pp. 211-227, Apr.2024.

ARTICLE INFO

Received: 15 November 2023
Accepted: 6 April 2024

ABSTRACT

Wireless Sensor Networks (WSNs) commonly comprise numerous low-cost sensor nodes that possess limited sensing, computation, and communication capabilities. Given the constrained resources of these sensor nodes, it becomes crucial to minimize data transmission to enhance both the average sensor lifetime and overall bandwidth utilization. Data aggregation serves as a process of summarizing and merging sensor data to reduce the volume of data transmitted within the network. Since wireless sensor networks are typically deployed in remote and challenging environments for transmitting sensitive information, sensor nodes are vulnerable to node compromise attacks. Consequently, security issues such as data confidentiality and integrity assume paramount importance. Therefore, when designing wireless sensor network protocols, such as data aggregation protocols, it is imperative to prioritize security and energy efficiency. In this work, we focus on these issues and develop a novel data aggregation approach by using a compressed sensing mechanism. The proposed approach is Hybrid Compressed sensing Secure Fault Tolerant Data Aggregation (HCSFTDA). Moreover, we focus on incorporating security therefore we present a novel mechanism for key distribution and data integrity verification. The performance of the HCSFTDA approach is measured in terms of packet delivery rate, average energy consumption and overhead and compared with existing approaches. The comparative analysis shows that the HCSFTDA achieved better performance. The experimental analysis shows that the proposed model reported average energy consumption as 0.0667, packet delivery as 98% and reduced communication overhead as 400 Kbps.

Keywords: Wireless Sensor Networks, Data Aggregation, Compressed Sensing, Cluster Head, Base Station.

INTRODUCTION

Typically, a wireless sensor network is composed of a large number of inexpensive and low-powered sensing devices with resource-limited memory, computational capability, and communication resources [1], [2]. Such networks provide cost-effective solutions for military and civilian applications like surveillance, target tracking, health monitoring, and traffic management. Being a low-cost network, the hardware of the sensor nodes is simple, and they experience resource constraints, which in turn makes it a challenging task to devise an effective data-gathering solution. A survey of existing network protocols for wireless sensor networks reveals that the most critical design factor is the availability of "battery power." In the literature, several means have been proposed to reduce the consumption of power in such networks, including scheduling, topology management, packet routing, and Data Aggregation (DA) [2], [3]. Data aggregation strategies focus on merging and summarizing data packets collected from different sensor nodes, thus reducing the overall data to be transmitted. Figure 1: Example DA scheme. In Figure 1, a group of sensor nodes senses information from a region. Instead of sending data from each sensor node to the Base Station (BS) individually, information is collected by a particular sensor node, which is called the data aggregator, from its surrounding nodes (neighbours). The data aggregator aggregates data by, e.g., calculating the average, and transmits that data to the BS based on a multi-hop path. Thus, such an operation provides a way to reduce the size of the data sent across the network considerably and, as such, its power resources. In this way, the network attains a fully efficient operation with no additional overhead burdening the resource-constrained sensor nodes.

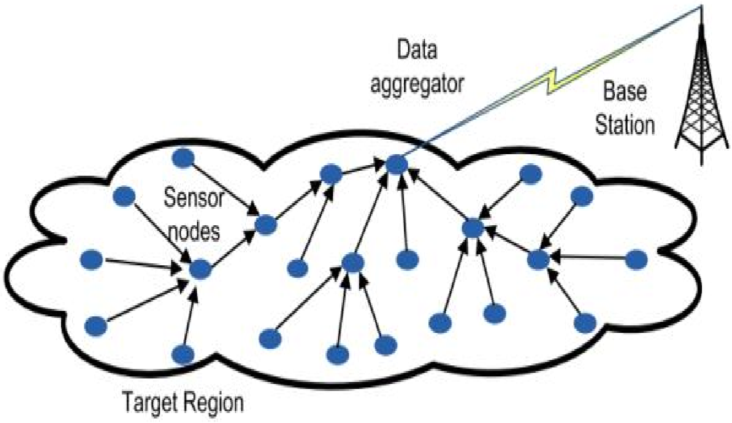

Figure 1. Data Aggregation

Data Aggregation in Wireless Sensory Networks is the process of collecting, summarizing, and transmitting sensory data from various nodes in the network to a central or base station. WSNs consist of a huge number of distributed sensor nodes that facilitate the gathering and monitoring of data from the adjacent environment. Data aggregation is really vital in WSNs, as it ensures less energy consumption, less network traffic, and more network lifetime. In place of raw data being transmitted from each sensory node straight to the base station, it is aggregated in the intermediate node and then transmitted to the base station. It significantly reduces the no. of transmissions and conserves energy. Numerous techniques are there for data aggregation in WSNs, including, (a) Spatial Aggregation: In this technique, neighbouring sensor nodes gather the data from their surrounding area and aggregate it into a single packet before relaying it to the next hop or base station. This decreases the redundancy and eliminates the need for transmitting similar data multiple times. (b) Temporal Aggregation: It involves collection of data over a specific time interval and aggregating it before transmission. Rather than transmitting every sensor reading individually, nodes can aggregate data over a period of time, like by averaging the temperature readings over a minute, and send the aggregated value. (c) Aggregation Tree: It is a hierarchical structure that is designed in the network, where the nearest nodes to the base station act as aggregation points to collect data from their child nodes. Every intermediate node aggregates the data it receives and forwards the aggregated result towards the base station. This significantly reduces the no. of transmissions required and preserves energy. (d) In-network Processing: It performs computations or data aggregation chores directly within the sensor nodes themselves. The base station can thereby receive lesser data if nodes are able to locally process, filter or aggregate that information being sent.

Data aggregation in WSNs should consider network topology, energy efficiency, data accuracy and the specific application requirements. The methods of aggregation should strike a balance between the reduction of communication overhead, conserving energy and maintaining the integrity of information.

Fault Tolerant Data Aggregation

The utilization of sensor networks in fault-tolerant data aggregation is required by a multitude of applications. This is done to guarantee that data is gathered by many nodes or sensors from a variety of sources, even in the event that faults and errors take place. In order to accomplish this, it may be necessary to collect data from a number of different nodes that are part of a network and work through any difficulties that may crop up throughout the process of data collection. The following are the components that comprise fault-tolerant data aggregation:

To make sure that there is redundancy in the data, a high number of nodes or sensors detecting the same information are normally used. This is one way of doing redundant data collection. As a result of the existence of these multiple sources of information, even the malfunction of the individual sensors will not hinder the system from functioning well. Replication, in which several nodes collect data that is the same, or diversity in which different but very similar data is collected, constitutes ways through which redundancy comes about.

Consensus algorithms: In the area of fault-tolerant data aggregation, distributed consensus methods, such as Paxos or Raft, are normally implemented. These algorithms enable multiple nodes to agree on value or result conformity of the collected data, even if failures or network disruptions take place; they guarantee data consistency in the aggregated sets of data across all nodes participating in the aggregation process.

Failure detection and recovery: Generally speaking, fault-tolerant data aggregation systems integrate mechanisms for detecting node failures or errors for continued reliability in the aggregation process. This can be effected using heartbeat messages or some other form of health monitoring technique to ensure the resilience of the fault-tolerant data aggregation systems. A failure can be mitigated by triggering a recovery procedure, which can be done by changing the responsibilities of a node that has failed and collecting data from other sites.

As the number of nodes or sensors in a network increases, having protocols and techniques that support scalability is increasingly important. Load balancing is also a major factor. Methods of load balancing are employed to guarantee no single node becomes too heavily burdened with process work. This is achieved by evenly distributing data-collecting workloads among different nodes. This is to say that a system for fault-tolerant data aggregation, such as this, is able to support colossally vast amounts of data and is able to scale up to further nodes or sensors whenever it needs to.

Validation of collected data and determining the presence of outliers: At times, there is a need to ascertain data that has been collected as a way of realizing any misreadings, or identifying the outliers. Various statistical methodologies, data cleansing algorithms, and machine learning methods exist that can help one identify those points of data that are highly distinct from what was anticipated prior to the point of aggregation.

From a fundamental standpoint, the objective of these methods is to collect information that is accurate and trustworthy from a variety of sources, even in the event that errors or failures occur. These techniques incorporate redundancy, error detection and correction, consensus algorithms, failure detection and recovery, scalability, and data validation in order to provide distributed systems with data aggregation that is both robust and resilient.

Energy Aware Data Aggregation

As indicated before, sensor networks have limited resources that when exceeded can cause the network’s life to be shortened through energy. Hence, Wireless Sensor Network (WSN) data aggregation is another technique for saving more energy during data aggregation. The aim here is to reduce individual sensors’ power consumption while maintaining the required level of data collection and aggregation.

The fundamental tenets of energy-conscious data aggregation are as follows:

In-network Processing: The recommendation is to perform processing and aggregation jobs within the network itself, in close proximity to the source nodes, in order to optimize energy consumption. This restricts the quantity of delivered data over extended distances and decreases communication expenses, therefore preserving energy.

Data fusion refers to the process of combining comparable or redundant data readings from several nodes into a single representative value. By consolidating comparable data points, there is a reduction in the total volume of transmitted information, resulting in power savings. Fusion algorithms can utilize statistical techniques such as averaging or summing, as well as more sophisticated methods like cluster-based approaches.

Sleep schedule is when the sensor nodes in WSN are put into a low-power sleep mode for long periods and wake up as required. This will save considerable energy because, using sleep scheduling, the multiple nodes have their sleep cycles synchronized so that they start at the same time. Since energy-efficient data aggregation algorithms, the sleep scheduling techniques allow only a specific group of nodes to get involved in the aggregation process. The rest of the nodes are made to remain inactive by entering the sleep state, therefore, energy saving.

Routing optimization is the core process for efficient data aggregation in energy-aware routing. Routing algorithms may be designed with optimization to pick the most optimal routes where the data are aggregated and sent to the base stations. Such algorithms depend on various parameters, such as the energy levels of nodes, transmission distances, and the network topology, in such a way as to minimize the total energy consumed during data transfer.

Data compression techniques aim to reduce the size of combined data prior to distribution. The compression algorithms aim to remove redundancy within datasets by exploiting information patterns and efficiently compressing the data. By reducing the quantity of transfer of information, the power is effectively saved.

Service Quality (QoS) characteristics need to be taken into account. Energy conscious data aggregation algorithms need to balance, reducing energy with the required data quality or QoS. To be utilized by specific applications there's the need to address quality of service aspects such as latency, data correctness, and dependability; it guarantees the obtained results boast of the right standards and is still designed to use the least amount of energy.

Due to the limited battery power, wireless sensor networks require energy-aware data aggregation. Optimizing the collection, transmission and processing of data extends the life of networks and reduces battery charge and replacement. Power efficiency is required to extend the life of sensors in independent locations with a limited battery population. Data aggregation solutions consume less power through data reduction, sleep scheduling, and adaptive sampling. Compression, fusion, and aggregation bring down the transmitted data size. Sleep scheduling maintains the mode by shifting between active and sleep modes to save energy. Adaptive sampling provides an important reduction in energy use by adjusting the sampling rates if there is nothing happening. Low Energy Adaptive Clustering Hierarchy (LEACH) and Threshold-Sensitive Energy Efficient Sensor Network (TEEN-N) are energy-saving routing protocols for sensing data delivery. It saves energy because the specific "node" is activated only when it is synchronized with the needs and conditions of the network. Cross-layer resilience adjusts transmit power and routing decisions through protocol layers. Solar and heat energy enhance power from the battery. These solutions optimize energy efficiency, hence ensuring that sensor node lifespans are expanded and wireless sensor networks are sustainable across applications.

Secure Data Aggregation

The term "Secure Data Aggregation" (SDA) refers to the process of combining and summarizing data from multiple sources, all while preserving the integrity, privacy, and confidentiality of individual data points. This method is widely utilized in a variety of fields, including healthcare, finance, and applications related to the Internet of Things (IoT).

The process of aggregating data in a secure manner can be accomplished through the use of a variety of different approaches. The following are some examples that are frequently used:

Key encryption: There are a variety of methods that can be utilized to encrypt the information before its transmission in order to safeguard its confidentiality.

Differential privacy protects the right to privacy of the individual while ensuring that the aggregated results will continue to have statistical reliability. These techniques, if put to use, can serve this purpose, and these are the techniques of randomized response and Laplace noise addition.

Secure Multi-Party Computation (MPC): Secure Multi-Party Computation is a programming technique that enables many parties to jointly compute an aggregate function without disclosing their individual inputs. The data of each side will be encrypted, and computations will be performed on the values that have been encrypted. Each of the parties encrypts their own data, and then computations are performed on the values that have been encrypted. It is guaranteed by MPC protocols that the aggregated result is produced without any party being aware of the data of other participants.

Trusted Execution Environments (TEEs): Trusted Execution Environments provide a secure calculation space for performing calculations on delicate information. There are specific extensions for the establishment of the TEEs, such as Intel Software Guard Extensions (SGX) and ARM Trust-Zone that protect data from superior malicious software. Aggregation computations can be performed practically locally within the TEE so as to preserve data confidentiality and to guard against unauthorized alteration.

Secure Protocols: Several secure aggregation protocols that include secure sum protocols, protocols for secure average, and secure counting protocols have been proposed. These protocols use mathematical algorithms to perform the computation securely without compromising data confidentiality.

Hence a need to develop an efficient and secure data aggregation scheme in these networks that does not only tolerate fault but also conserves energy. It is in these aspects that the present work is unfolded and a new method for data aggregation proposed here is outlined. The main contributions of this work are as follows:

To incorporate compressed sensing for data aggregation and present a hybrid compressed sensing mode to improve the aggregation performance.

We present a tree-based routing along with the hybrid compressed sensing.

We also incorporate a secure communication link establishment model to consider the security aspects.

The rest of the manuscript is organized into the following segments: section II presents the literature review of existing data aggregation techniques, section III presents the proposed solution to overcome the issues of existing aggregation schemes, section IV presents the comparative analysis and section V presents concluding remarks about this work.

LITERATURE REVIEW

This section presents a brief literature review of existing techniques in this domain of efficient data aggregation in wireless sensor networks. As discussed before, DA plays an important role in WSN by reducing the energy requirement, minimizing the network traffic and extending the lifespan of the network. several methods have been discussed such as spatial aggregation, temporal aggregation, hierarchical Aggregation Tree and many more. This topic still remains a hot topic for the research community due to its significant impact on the communication performance of sensor networks.

Devi et al. [1] addressed the issues such as energy balancing, packet loss, and reducing latency in WSN. The authors claimed that the existing time-slot assignment-based approaches do not focus on packet loss and latency issues. So, the authors proposed a new two-phase scheme in which, during phase 1, every sink node subjects comprehensive aggregation on the data delivered from its surrounding nodes. Afterwards, in conjunction with the minimum spanning tree technique, the sink constructs an aggregation tree. This phase supports the effective aggregation of data within clusters and also establishes a hierarchy structure for data transceiving. This scheme, in phase 2, takes packet loss rate and latency into consideration while prioritizing and timeslot allocation to the node that holds aggregated data. Because of this combined assumption, the scheme reduces the chance of undistinguished retransmission and waiting and consequently results in better network performance in WSNs.

Along with energy efficiency, privacy preservation and reliability also play important roles in Hu et al. [2] focused on reliable and efficient data aggregation and presented a chain-based data aggregation approach where sensor nodes are arranged in a tree topology. In this approach, the leaf nodes in a tree structure establish multiple chain topologies by reconnecting with each other sequentially. To ensure data privacy, after data gathering, the tail nodes (nodes at the end of the chains) divide their sensing data into J fragments. They keep one fragment for themselves and distribute the remaining J-1 data fragments to their neighbouring nodes. Additionally, each tail node inserts fake fragments into the J-1 fragments to disrupt potential adversaries or unauthorized access to the data. Similarly, Naghibi et al. [3] introduced Secure Hybrid Structure Data Aggregation (SHSDA) by a combination of star and tree structure. Moreover, it divides the network into four equal portions and a star structure is formulated where lightweight symmetric encryption is also employed to ensure the secure data exchange. The data from leaf nodes is encrypted and transmitted to their respective parent nodes. The encrypted data is then propagated through the network, gradually reaching the root node via the star structure. Ullah et al. [4] introduced a self-organized map neural network to minimize data redundancy and outliers. Moreover, this method uses cosine similarity measurement to enhance the clustering performance. Similarly, optimization schemes are also adopted in this domain because these methods try to find the optimal solutions by iterative methods. Based on this concept, Yousefpoor et al. [5] presented a novel secure and optimized DA approach by using the dragonfly approach. The complete process is divided into three main phases including intra-cluster, inter-cluster and data transfer. The intra-cluster DA employs the fuzzy scheduling mechanism to adjust and assign the appropriate transmission rate of cluster member nodes. In the inter-cluster phase, this method constructs an aggregation tree where a dragonfly optimization strategy is applied to obtain the optimal aggregation tree. Finally, uses columnar transposition cipher mechanism is employed to establish the secure connection between cluster members and Cluster Head (CH). Further, a lightweight encryption method is implemented which is based on the Residue Number System (RNS) and improved as RNS+. Based on this concept of intra-cluster aggregation, Bongale et al. [6] introduced an energy-efficient DA approach where it constructs an aggregation path from source to corresponding CH and it also aggregates data packets from relay nodes. Currently, machine learning and intelligent learning approaches have been adopted in various offline and online applications. By leveraging the learning mechanism, therefore, Gandhi et al. [7] presented a novel approach by combining grid clustering and fuzzy Reinforcement Learning (RL) to maximize the network lifetime along with energy-efficient DA. Finally, the fruit fly optimization strategy is used to facilitate the dynamic relocation of the mobile sink node. Based on this concept of machine learning, Ullah et al. [8] introduced a machine learning approach by using a modified radial basis function neural network which is used to classify the data at cluster head and discard the redundant and outlier data. Moreover, this method uses cosine similarity measurement to formulate the clusters and RBF is used with Mahalanobis distance to detect and classify the multivariate data. Dhanaraj et al. [9] proposed a sink-originated hybrid and dynamic clustering mechanism. This approach focuses on node handling capability, CH selection and forwarder node selection. This method uses space and time correlation data collection and performs the data aggregation to improve the communication efficiency. Liu et al. [10] reported that query processing in sensing also consumes excessive resources and ensuring privacy for queries also plays an important role. Therefore, the authors introduced a Queries Privacy Preserving mechanism for Data Aggregation (QPPDA) which aggregates all queries in a single packet and employs homomorphic encryption to obtain reliable and energy-efficient aggregation. In [11] authors focused on security aspects and analyzed Amin Biswas's approach to identifying the drawbacks. Further, the authors have introduced a lightweight key agreement scheme by using symmetric key cryptography to ensure secure data exchange. Xiong et al. [12] introduced a lightweight authentication protocol by using hash chain and pseudonym identity which facilitates the authentication, user anonymity and forward secrecy. Jung et al. [13] studied existing mechanisms and reported that the method is susceptible to smart card attacks. Therefore, to overcome this issue authors introduced anonymous user authentication and key agreement schemes by using symmetric key cryptography.

Chandnani et al. [14] introduced ANT Particle Swarm Optimization Adhoc On-demand Distance Vector (ANTPSOAODV): A Trust-Based Secure Data Aggregation Method and an Energy-Efficient Secure Routing Protocol for Multi-Hop Environments. The proposed ANTPSOAODV protocol aims at ensuring secure data aggregation by closely monitoring the behavior of nodes within the network, evaluating their trustworthiness, and optimizing data collection methods. This routing protocol is composed of three basic operations: Zone Segmentation: It starts by dividing the IoT WSN into outer and inner zones based on the location of a node. The concept of this zoning approach is the basis for data routing in a very effective way. Cluster Formation: Clusters are dynamically formed within every single zone based on node proximity. These clusters are headed by CHs and prove to be a vital part of the data transmission process. For secure data transmission, a strong secret-sharing scheme is applied to provide confidentiality and integrity throughout the travelling of information from CHs to the central sink. This scheme secures the data in transit and, in turn, increases the overall security of the network. Ataei Nezhad et al. [15] proposed a secure DA approach, which is divided into three main phases, with the first one constructing the star structure in each cluster. Another unique key is shared between a child and its parent to encrypt the data. The second phase, i.e., intra-cluster communication, is based on cluster members transmitting their data to the assigned cluster head through a multi-hop process. During this phase, data is subjected to encryption at every hop with a unique key, and right after any data transfer, key updates are made for the sake of the data's security. The third phase enhances the security of the inter-cluster communications and is based on the introduction of an AP. It ensures that before any information is passed on, firstly, a cluster head has to be authenticated. This mechanism reduces the potential risk of malicious nodes within the network. Ahmed et al. [16] introduced an Energy-Efficient Data Aggregation Mechanism (EEDAM) in which data aggregation is secured using block chain technology. It works at the cluster level by and large for the sake of saving energy resources. In this context, edge computing provides a platform for the deployment of reliable services to IoT with the least possible latency. Block chain ecosystem is implemented within a cloud server for ensuring this trustworthy connection. This ecosystem validates the provision of secured services to the IoT ecosystem by authenticating the edge via the block chain. Dao et al. [17] focused on the challenge of optimizing network lifetime through the minimization of overall transmission and reception energy consumption across all sensor nodes. Additionally, the authors introduce a heuristic algorithm known as "Reduce Redundant Packet Tree" (RRPT) designed to mitigate redundancy in the network. The RRPT algorithm operates on the premise that fewer packets lead to reduced energy consumption. Subsequently, an Evolutionary Algorithm called "PGA", which leverages RRPT as a heuristic initialization to enhance overall performance further is also introduced.

This section has discussed various schemes for secure DA in WSN where we have identified several challenges such as security, the tradeoff between security and energy efficiency, scalability of these mechanisms in heterogeneous environments, poor adaptability to dynamic environments and interoperability. Given Table 1 below shows some analysis based on the literature review.

Table 1. Literature Review Analysis

Ref No.	Critical Analysis
[1]	Failure in reconstruction can lead to packet loss and security aspects are not considered.
[2]	Tree topology is used for network deployment and a chain-based model is used for connection but user authentication is not considered which can lead to different attacks.
[3]	It uses secure symmetric encryption for data security however it doesn't support mobile nodes.
[4]	This method doesn't use dimensionality reduction which impacts its accuracy and increases computational complexity due to redundancy.
[5]	It used dragonfly optimization however the performance purely relies on objective function and performance is affected due to the convergence mechanism.
[6]	It uses fuzzy and reinforcement learning mechanisms. The fuzzy logic model fails to handle uncertain scenarios.
[7]	This method presented the combination of RL with FF optimization where performance is affected by reward policy and convergence of optimization.
[8]	This method used RBF machine learning which can be beneficial for prediction but it requires a huge amount of training data. Low quality data can lead to increased false alarms.

METHODOLOGY

Previous sections have described various aspects of data aggregation and several methods have been discussed to augment the performance of WSN communication. However, maintaining the fault-tolerance, energy-aware and secure nature of data aggregation schemes has a significant impact on these networks. Therefore, we present a combined approach to deal with these issues. The first subsection presents the network model and problem formulation, later, the fault-tolerant model is presented followed by energy-aware data aggregation. Finally, we introduced the concept of privacy preserving during the aggregation process. The overall aggregation mechanism is presented in Figure 2 below.

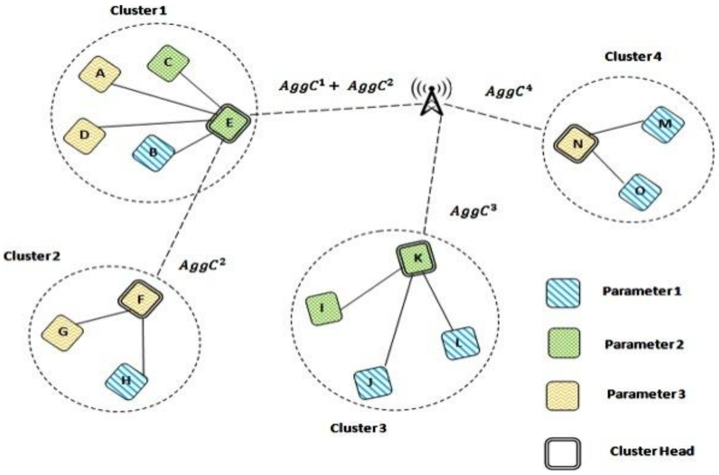

Figure 2. Data Aggregation Process

Compressed Sensing

This section presents a brief discussion of the basic theory of compressed sensing and its implementation for data collection in sensor networks. Let us consider that u is a signal and the signal vector is denoted as:

$$\mathrm{u} = [\mathrm{u1} \ldots. \mathrm{un}]^{\mathrm{T}} \tag{1}$$

This signal is represented sparsely under the bias $\psi = [\psi_1, \ldots, \psi_n]$ subjected to $\mathrm{u} = \sum_{i=1}^{m} \mathrm{w}_i \psi_i$ and $\mathrm{m} \ll \mathrm{n}$. According to the concept of compressed sensing, in a given certain conditions, instead of collecting the data u directly, we only need to assemble the $\mathrm{k} = \mathrm{O}(\mathrm{m} \log \mathrm{n})$ measurements $\mathrm{v} = \phi\mathrm{u}$ where $\phi = [\phi_1, .., \phi_n]$ represents the sensing matrix of size $\mathrm{k} \times \mathrm{n}$. With the help of this, we can recover the signal u from v by solving the optimization problem as:

$$\min_{\mathrm{w} \in \mathbb{R}^{\mathrm{n}}} \|\mathrm{w}\|_{l_1} \mathrm{s.t.}\ \mathrm{v} = \phi\psi\mathrm{w} \tag{2}$$

Where $\mathrm{u} = \psi\widehat{\mathrm{w}}$ represents the optimal solution. The effectiveness of this mechanism relies on the sparsity of the signal and the choice of reconstruction approach. Given that networked data typically exhibits a high degree of sparsity, Compressed Sensing (CS) is a highly suitable method for data collection in WSNs. This concept of CS is extended for WSNs in this work where we consider that the total n number of nodes are present in the network boundary and each node acquires sample data u_i and the goal of WSN is to collect all data $\mathrm{u} = [\mathrm{u}_1, \ldots, \mathrm{u}_n]^{\mathrm{T}}$ at sink node. To address the issue of heavy traffic congestion around the sink caused by each node sending its sample without data aggregation, a potential solution is to apply CS to data collection. This approach offers a promising way to alleviate this. This can be expressed as:

$$\mathrm{v} = \phi\mathrm{u} = \mathrm{u}_1\phi_1 + \ldots + \mathrm{u}_n\phi_n \tag{3}$$

DA plays a crucial role in WSNs by reducing the amount of transmitted data and minimizing energy consumption. Instead of each individual node sending its data directly to the sink, data aggregation allows nodes to combine and summarize their collected information before forwarding it. However, compressed sensing differs from traditional aggregation methods. According to CS based aggregation, each node utilizes a column vector ϕ_i to expand it to k dimensional vector. Further, instead of sending the raw data, it encodes and formulates an encoder vector for transmission. Aggregation occurs when these coded vectors meet, and the aggregation path carries a consistent traffic load of k. Ultimately, the sink receives the accumulated k-dimensional vector, rather than n raw samples, which can then be decoded to recover the original n raw samples. The encoding process is distributed across all nodes. Moreover, this process includes some basic mathematical operations resulting in negligible computational cost. This approach allows for efficient data transmission and reduces the traffic load on the network. By encoding and aggregating the data, the network benefits from lower energy consumption and improved scalability. The distributed encoding process ensures that each node performs minimal computations, contributing to the overall efficiency of the system.

CS offers a promising approach to efficient data collection in WSNs by enabling the acquisition and transmission of sparse or compressible signals at reduced sampling rates. The underlying principle of CS includes that it exploits the inherent sparsity or compressibility of signals to reconstruct them from a small number of linear measurements. Further, it relies on the assumption that the signal of interest can be represented in a sparse domain, where only a few coefficients are significant, even though the signal itself might be of high dimensionality. CS employs random or structured measurement matrices to acquire linear projections of the signal, allowing for efficient sampling and reconstruction. In WSNs, where energy efficiency is critical, CS reduces the amount of data transmitted by directly sampling and compressing sensor readings at the node level. CS-based data collection architectures often employ distributed data aggregation and fusion techniques to further reduce communication overhead and energy consumption. Various CS reconstruction algorithms, such as Basis Pursuit, Iterative Hard Thresholding, or Compressed Sensing Matching Pursuit, are adapted to reconstruct sparse sensor data at the sink node or base station. The CS has several advantages such as energy efficiency, bandwidth conservation, and scalability. However, as the WSN grows larger or more complex, scalability becomes a concern due to increased communication overhead, computational burden, and network congestion for real-time processing or dynamic network topologies. Therefore, designing a scalable CS can overcome this issue.

Hybrid Compressed Sensing Mechanism

This segment describes the concept of a hybrid compressed sensing mechanism in the context of wireless sensor networks. According to Figure 3, $\mathrm{n}-1$ nodes are transmitting data samples to the n^{th} node thus the outgoing link carries a total n number of samples if no aggregation mechanism is employed. In the same scenario, if we employ lossy aggregation, then it carries 1 sample. Similarly, if we apply the compressed sensing based aggregation then each link will be forced to carry k samples which leads to an increase in the traffic at early stage transmission. Thus, the traditional CS based aggregation mechanism suffers from the congestion issues. Hence, the correct approach to implementing CS is to initiate CS coding only when the number of outgoing samples exceeds or equals a threshold value, denoted as “k”. Otherwise, the collection of raw data (without aggregation) is employed. Therefore, in this work, we introduce a hybrid compressed sensing approach. The concept of hybrid CS aggregation conglomerates the benefits of non-aggregation and traditional CS aggregation. By doing so, it effectively diminishes the burden on network traffic while maintaining the integrity of the original dataset. It is worth noting that the hybrid CS aggregation introduces two distinct types of traffic: encoded and raw traffic. These can be distinguished by a flag carried within the data packet.

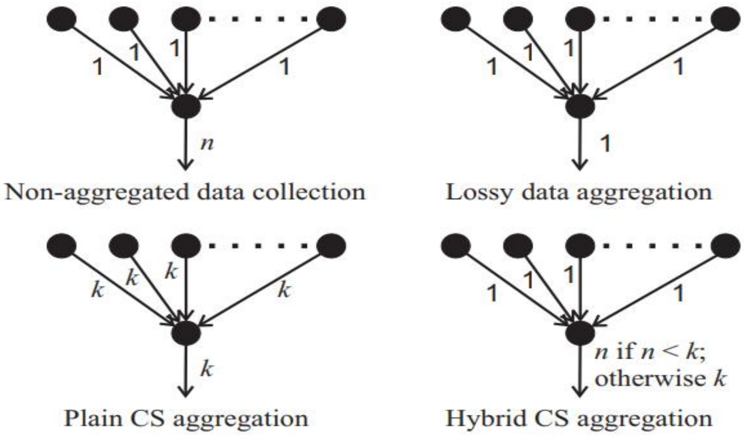

Figure 3. Different Data Collection Schemes

In the network deployment, all nodes are initiated with a non-aggregation mode. Let's consider a specific node, denoted as node "i". Given a predetermined threshold value, denoted as "k" node i will wait to receive data from all its downstream neighbours. It will accumulate the data sent by its neighbors until it has received more than k−1 raw samples or any encoded samples. Once node i has received the required number of raw or encoded samples, it switches to CS aggregation mode. In this mode, node i creates a vector $u_j\phi_j$ for each uncoded sample u_j it receives, including its own sample u_i. The threshold value is determined by data sparsity, transmission cost and energy efficiency which facilitate the transition to CS aggregation node. After aggregating the vectors, node i proceeds to send out exactly k encoded samples. Each encoded sample corresponds to a column vector representing the aggregated data.

Problem Formulation and Network Model

In this work, a WSN is denoted by a connected graph $G\ (V, E)$. The set of vertices, V, represents the nodes within the network, while the set of edges, E, corresponds to the wireless links connecting these nodes. Within this network, there exists a unique node, denoted as $s \in V$, which is referred to as the sink. The sink node is responsible for gathering data from the entire network. The cardinality of the vertex set V is denoted by n, while the cardinality of the edge set E is denoted by l. Let us consider that $c: E \rightarrow \mathbb{R}_0^+$ represents the cost assignment. Similarly, $x: E \rightarrow \mathbb{R}_0^+$ represents the load allocation as data traffic load caused by a certain aggregation scheme on the given link $(i. j)$. Here, our main aim is to diminish the total cost by reducing energy consumption, this can be expressed as:

$$\sum_{(i,j)\in E} c_{ij} x_{ij} \tag{4}$$

In this scenario, we make the assumption that all nodes are approximately synchronized in time, and the process of collecting data occurs in rounds. At the start of each round, each node generates a single unit of data, also known as a sample. At the conclusion of the round, the sink node gathers all the information from the nodes. To ensure that no data packets are lost during transmission, it is crucial to schedule the network effectively by implementing efficient MAC mechanisms (Figure 4).

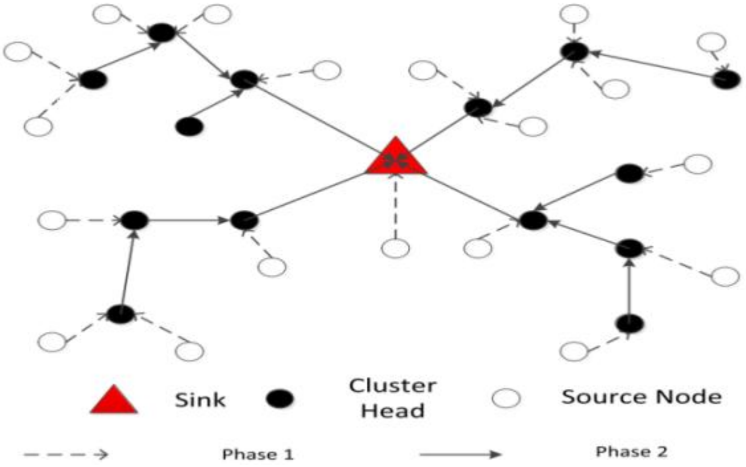

Figure 4. Compressed Sensing Mechanism

To maintain generality, we limit the data aggregation process to a tree structure that originates from the sink node. In order to simplify, we make the following assumptions:

The total N number of nodes is distributed in a given 2D region with L radius and sink node present at the center of the designated sensing area.

The sink node has enough space and processing capability.

Initially, all sensor nodes are assigned identical initial energy and transmission rates.

Nodes are capable of knowing their location by employing relative location estimation.

For the given scenario where the sink node is present at the center where j^{th} cluster consists of $m_j^{'}$ nodes which are involved in the aggregation process is expressed as:

$$m_j^{'} = \sum_{i=1}^{m_j} s \times 1 = m_j s \tag{5}$$

At this stage, $m_j^{'}$ needs to forward their weights therefore cluster head node receives $m_j^{'}$ packets. Hence, the average energy consumption in j^{th} cluster can be expressed as:

$$E_{intra}^{j} = \sum_{i=1}^{m_j^{'}} E_{Tx}^{i}\left(k, E(d_i)\right) + m_j^{'} E_{Rx}(k)$$

$$= k\sum_{i=1}^{m_j^{'}} \left(E_{ele} + \epsilon_{amp}E(d_i^2)\right) + m_j^{'}kE_{ele}$$

$$= 2m_j^{'}kE_{ele} + k\epsilon_{amp}\sum_{i=1}^{m_j^{'}} E\left(d_i^2\right) \tag{6}$$

Where $E_{Tx}^{'}\left(k, E(d_i)\right)$ denotes the energy consumed by i^{th} node to forward the k bit data to its corresponding CH, $E(d)$ is the distance between node to CH. Moreover, we also assume that the nodes are capable of adjusting their energy levels according to the transmission distance.

Routing Tree Construction and Compressed Sensing Aggregation

The hops are transferred from cluster head to another cluster head and it is determined based on the communication radius and distribution of cluster heads. Thus, the CH receives the $h-1$ data packets and energy-consumption of inter-cluster can be expressed as:

$$E_{inter} = \sum_{i=1}^{h} E_{Tx}^{i}(k, d_i) + (h-1)E_{Rx}(k)$$

$$= k\sum_{i=1}^{h} \left(E_{ele} + \epsilon_{amp}d_i^2\right) + (h-1)kE_{ele}$$

$$= (2h-1)kE_{ele} + k\epsilon_{amp}\sum_{i=1}^{h} d_i^2 \tag{7}$$

Where d_i denotes the transmission distance of i^{th} data packet. This process is repeated to construct the inter-cluster routing tree. Considering the significant data correlation among nodes within a cluster, we can employ a random space sparse matrix to reduce the measurement values. In conventional data aggregation approaches relying on compressive sensing, the cluster head generates the measurement matrix for the CS process. In this process, both the data and measurement matrix are forwarded from the cluster head to the sink node. However, by utilizing a sparse seed vector, the sink node can directly generate the random space sparse matrix. This allows each CH to produce its respective sub-matrix using the seed vector obtained from the sink node. However, implementing a hybrid CS mechanism involves overcoming several technical such as the design of custom algorithms to integrate multiple sensing modalities while ensuring compatibility and synergy between them. Synchronization and calibration are essential to match the streams significantly and to penalize the discreteness of the sensors. Computational management requires the optimization of suitable algorithms, which probably may entail parallel computing or employing hardware add-ons appropriate for the heterarchical system. Knowledge is the main point in getting important and productive information from numerous and dispersed data sources in providing influence from uncertainty and variability data. However, resource limitations greatly owning to data constraints in embedded systems or IoT devices make the algorithm optimization for efficiency and energy consumption. All these challenges can only be met by a multidisciplinary approach that embraces skills in signal processing, optimization techniques, data fusion and system integration. It is evident that implementing mechanisms for hybrid computing in an organisation requires practitioners and researchers with expertise across different fields to collaborate. In this work, we have modified the aggregation process with the help of an efficient routing tree construction which is obtained based on the communication radius and distribution of cluster heads.

Incorporating Secure Data Aggregation Module

The proposed HCSFTDA in WSN has three main phases: establishing the secure link, aggregating the data and authenticating the integrity of aggregation results.

Establishment of a Secure Communication Link

This stepinvolves the establishment of a secure communication infrastructure among sensor nodes. These nodes may have been initialized with certain secret information but have not had any direct contact with each other before. The conventional method is key pre-distribution, where keys are preloaded into sensors before their deployment. Once deployed, each sensor establishes a confidential connection with its neighbouring sensor using a pre-shared pair-wise key. The key connectivity, which refers to the likelihood of a sensor node finding a common key with its neighbour, plays a vital role in pair-wise key distribution schemes and requires careful consideration. However, these methods are not resilient to different types of attacks.Therefore, we adopt the random key distribution mechanism where each node acquires a random subset of k keys. The next stage includes the shared-key discovery phase where two nodes engage in a process to identify a common key for communication. They search within their respective subsets to find a mutually shared key, which then serves as their secret key for secure communication.

Aggregation Functions

The sensing data denoted from different sensing nodes can be aggregated by employing the following functions in WSN:

$$Sum : f(s_1, .. s_n) = \sum_{i=1}^{n} s_i \tag{8}$$

$$Average : f(s_1, .. s_n) = \sum_{i=1}^{n} \frac{s_i}{n} \tag{9}$$

$$Median : f(s_1, .. s_n) = s_{(r)}, r = \frac{(n+1)}{2} \tag{10}$$

$$Minimum : f(s_1, .. s_n) = \min\{s_i | i = 1, .. n\} \tag{11}$$

$$Maximum : f(s_1, .. s_n) = \max\{s_i | i = 1, .. n\} \tag{12}$$

$$Count : f(s_1, .. s_n) = |\{s_i | i = 1, .. n\}| \tag{13}$$

Data Integrity

In the next stage, we focus on checking the data integrity to ensure data protection before transmission to the sink node. Here, we assume that each node is initiated before the deployment and symmetric pair-wise key distribution where keys are shared with the based station and sensor node. In the data transmission phase, any given leaf node assigns the temporary key and forwards its data, node ID along with message authentication code to its parent. The aggregation process begins with the parent node, which computes the aggregate value of its children nodes' readings. It then transmits the computed result, labelled as, along with the node ID and a message authentication code (MAC) generated using the key K_i Base Station (BS). The data is then forwarded to the parent node, which in turn sends the final aggregation value along with its corresponding MAC to the BS. In the data validation phase, the BS verifies the integrity of the final aggregation result. Furthermore, the BS disseminates temporary keys, enabling the intermediary aggregators to authenticate the intermediary aggregation results using these pair-wise keys. This method reduces communication costs by obviating the necessity to transmit the data readings of every individual sensor node to the base station.

RESULTS AND DISCUSSION

This section presents the outcome of the HCSFTDA approach and the obtained performance of the HCSFTDA approach is compared with the traditional approaches of data aggregation.

Simulation Setup

The proposed approach is simulated by using the MATLAB simulation tool. In this work, we have considered that a minimum of 50 nodes are deployed and a maximum of 450 nodes are deployed. The deployment is done in a 2D region of 450x450m2 by following Gaussian distribution. Distribute the nodes in the deployment area following a Gaussian distribution, which simulates a more realistic spatial distribution of nodes. The initial energy of the node is considered as 7.2J. Given Table 2 below demonstrates the various simulation parameters adopted in this simulation setup.

Table 2. Simulation Parameters Used in This Work

Simulation Parameters	Considered Value
Total number of nodes	50-450 nodes
Transmission range	30m
Network deployment area	450 x 450 m^2
Transmission power	0.650 mw
Received power	0.375 mw
Initial energy	7.2 J
Simulation time	400 sec

Performance Measurement Parameters and Comparative Analysis

In this section, we briefly describe the different parameters used to measure the performance of the HCSFTDA scheme such as communication overhead, energy consumption, and packet delivery rate.

Communication overhead: In the context of WSNs, communication overhead refers to the additional energy consumption, computational resources, and time required for communication activities within the network.

WSNs consist of a large number of low-power wireless sensors deployed in an area to collect and transmit data to a central BS or sink node. Communication overhead is a critical factor in WSNs because energy conservation and network lifetime are crucial considerations due to the limited resources of the sensor nodes.

Energy consumption: The average energy consumption in a WSN depends on various factors, including the characteristics of the sensor nodes, the application requirements, the network topology, the data processing and communication protocols used, and the specific energy-saving mechanisms implemented. The energy consumption performance is affected due to several parameters such as Sensing Energy, Processing Energy, Communication Energy, ideal listening energy etc.

Packet delivery rate: Packet delivery in a WSN refers to the successful transmission of data packets from source nodes to the intended destination, typically a base station or sink node. The packet delivery performance is affected due to several parameters.

First of all, we measure the performance of the HCSFTDA model in terms of communication overhead that is evaluated by adding the packets from all sensor nodes during the aggregation phase. As discussed before, the communication overhead is caused due to multiple processes during the aggregation and secure communication establishment phase. Below given Figure 5 demonstrates the overall communication overhead performance. This experiment shows that the HCSFTDA approach induced minimum overhead due to its faster aggregation process which reduces the waiting period for packets at the aggregator node. For a 100-node scenario, the average communication overhead is obtained as 400, 490, 2500, 2700, and 3100 by using the HCSFTDA Model, Aggregating Secure Data -Separate MAC (SDA-SM), Aggregate MAC (AMAC), Energy Efficient Interest Based Reliable Data Aggregation (EIRDA), and Integrity-Protecting Private Data Aggregation (IPDA) respectively.

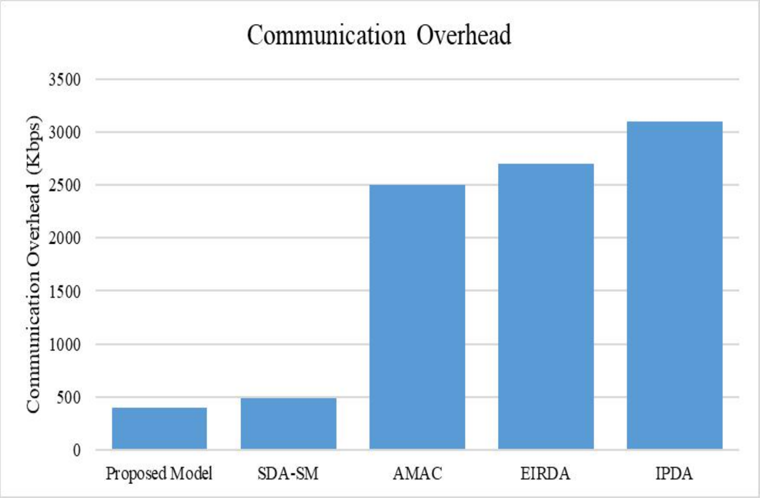

Figure 5. Communication Overhead Performance

Further, we measure the performance of the HCSFTDA model in terms of average energy consumption for three different scenarios where the number of sensor nodes is considered as 150, 300 and 450. The primary source of energy consumption in wireless sensor networks is the transmission of packets. As the amount of data transmitted increases, the energy consumed by the sensors also increases, leading to a faster depletion of energy and a shorter lifespan for the sensor nodes. Given Figure 6 below depicts the comparative analysis for this experiment.

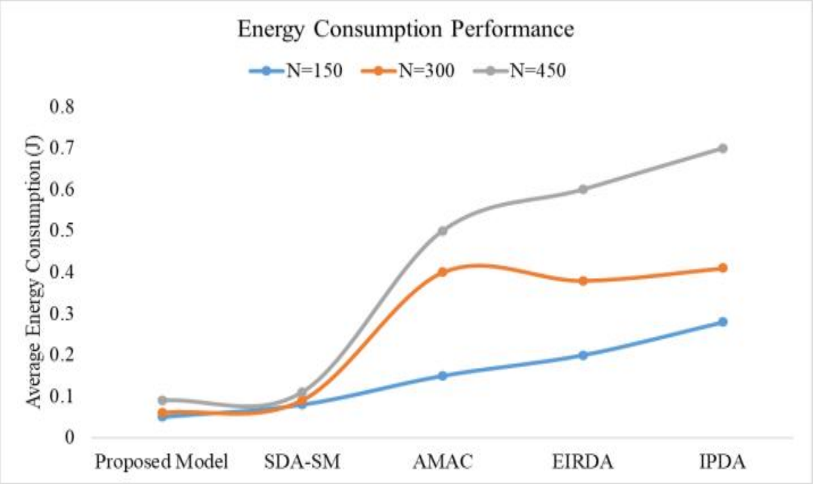

Figure 6. Average Energy Consumption Performance

As demonstrated in Figure 6, the average energy consumption is obtained as 0.0667,0.0934, 0.35, 0.394, and 0.464 for HCSFTDA, SDA-SM, AMAC, EIRDA, and IPDA schemes, respectively. This experiment shows that the increased number of nodes increases the energy consumption however, the HCSFTDA approach outperforms by minimizing the overall energy consumption.

Similarly, we extend this experiment and measure the energy consumption performance for a given simulation time. Given Figure 7 below depicts the comparative analysis of the HCSFTDA approach with other existing schemes.

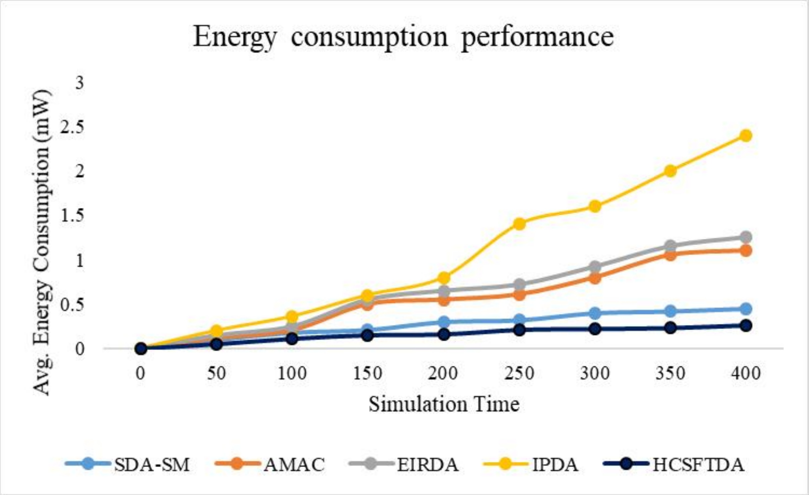

Figure 7. Energy Consumption Performance

According to this experiment, the average energy consumption performance is obtained as 0.288, 0.634, 0.71, 1.191, and 0.17 by employing SDA-SM, AMAC, EIRDA, IPDA, and HCSFTDA respectively.

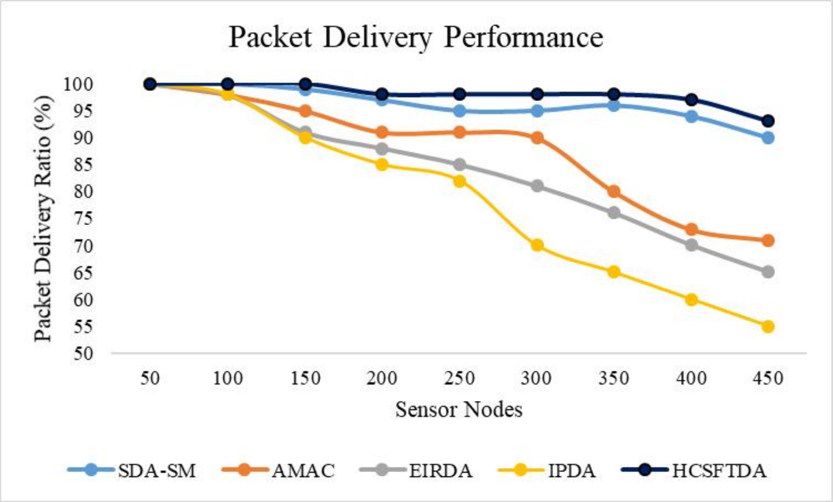

Figure 8. Packet Delivery Performance

Finally, we measure the packet delivery performance to show the robustness of the proposed approach. Figure 8 depicts the comparative analysis for this experiment. The experiment shows that the average packet delivery is obtained as 96.22%, 87.66%, 83.78%, 78.38% and 98%, respectively.

Security Analysis

In this section, we present the security analysis of the HCSFTDA approach and compare it with existing secure protocols for WSN communication.

Message Replay Attack

In our system, we have implemented a protective function to address the potential threat of an attacker impersonating a valid user node. This protective function ensures that the sensor node does not receive any previously delivered messages that it has already recognized as originating from a legitimate node. The sensor node has a maximum time restriction for receiving extended messages from user nodes. If a message takes longer than this limit, the sensor node will not react. This solution effectively mitigates the risk of a replay assault. In order to further mitigate the impact of message replay attacks, this system incorporates the utilization of a time factor during the exchange of information and authentication between nodes. A timestamp is valuable for verifying the freshness of communications and detecting the presence of replayed messages. The implementation of timestamp-based authentication adds an additional level of identity to the conversation, ensuring that messages that are simply repetitions or falsifications are not considered genuine.

Man-in-the-middle (MITM) Attack

In a man-in-the-middle attack scenario, the adversary's goal is to intercept and manipulate authentication messages, aiming to influence communication between authentication entities. They achieve this by forging authentication messages and then retransmitting them, creating an illusion of direct communication between the entities. However, our technique effectively counters such attacks by employing authentication and encryption to safeguard all communication between network entities. By utilizing unique values that the attacker is unaware of, the verification message cannot be counterfeited. As a result, our technique remains secure against man-in-the-middle attacks. Even if a man-in-the-middle attacker captures, replicates, and modifies authentication communications between authentication entities, the proposed solution is most safe. We use authentication and encryption on all authentication messages to assure secrecy, validity, and integrity. These precautions prevent the attacker from altering or deceiving intercepted messages. Importantly, our methods use new values unknown to the attacker. The adversary cannot forge the verification message, assume an identity, or alter the authorization process without understanding these values.

These tactics make man-in-the-middle assaults easy to avoid because their technique is immediately identifiable. So, authentication, encryption, and identity values are used to safeguard the communication environment and ensure that all messages are authentic.

Flooding Attacks

This attack scenario begins with the attacker sending a HELLO packet to the victim to drain network resources. The threat to our technique's legitimacy is greatly mitigated by it. Sensor nodes receive and send data, simplifying network design. I work with the dynamically selected CH to maintain network topology.

At the same time, the base station provided for data exchange is engaged in its maintenance. Secured technique employs node authentication by the base station to ensure the security of the data transferred. This narrow-down of the authentication step helps in increasing the security of our system and canceling the Hello Flood Assault opportunity.

Also, our system successfully prevents the selectivity forward attack which aims to limit the flow of data from certain sensor nodes. This threat is addressed in our framework by having each cluster be headed by one node that is tasked with forwarding the data received from the several sensor nodes. This threat is avoided in our approach by assigning unique cluster heads to forward data from each sensor node it possesses. Therefore, the selective forwarding attack no longer poses threats to the network. In conclusion, it minimizes the attacker's attempt to exhaust network resources created by HELLO packet and offers a strong defence against Hello flood and selective forwarding attacks.

Denial of Service Attack

Thus, in this attack scenario, the adversary seeks to prevent users from accessing the services/resources and exploit the transmission packet and capacity of a telecommunication network. But through several strategic measures mentioned below, the threat from this sector is fought rather effectively by our solution. Firstly, we follow a feature where the cluster head is rotated periodically, and the change is done at the end/restart of every round of transmission. This proactive approach makes it difficult for the adversary continuously to attack a certain cluster head and, therefore, increases the stability of our system. Furthermore, we provide for enthusiastic alerting of the newly assigned cluster head with regard to the content of its cluster. In this regard, the cluster head needs to take preemptive measures so that any adversary's attempt at attacking the cluster is easily identified The use of early warning systems allows for the identification of these problems within the cluster as they occur and have measures that limit the adversary's influence upon the cluster. Also, for safe and efficient communication, we have included an acknowledgment message system that will be implemented by the base station. This is a useful mechanism because it helps the base station understand that the transmission is successful, and if there is an interruption from the opponent, it is quickly identified and dealt with. If we practice how the network functions and if we acknowledge our messages delivered we are in a position to contain and counter threats that may pose a threat to the security of the system arena.

Given Table 3 below demonstrates the comparative security analysis of the HCSFTDA approach's resiliency to various attacks.

Table 3. Security Analysis: Resiliency to Various Attacks

Security Feature	**[11]**	**[13]**	**[12]**	**HCSFTDA**
Smart card loss attack	×	×	×	✓
Replay attack	✓	✓	✓	✓
User anonymity		✓	✓	✓
Mutual authentication	✓	✓	✓	✓
Desynchronization attack	×	×	✓	✓
MITM attack	✓	✓	✓	✓
Insider attack	✓	✓	✓	✓
Impersonation attack	✓	✓	✓	✓
Subsequence authentication	×	×	×	✓
Forward secrecy	×	×	✓	✓

The aforementioned security analysis shows the importance and novelty of the HCSFTDA secure data aggregation approach for WSNs. Real-time fault tolerance can be investigated through four distinct avenues, outlined as follows:

Redundancy: Redundancy is integrated into the Wireless Sensor Network (WSN) by deploying multiple sensor nodes to monitor the same machine or equipment. If a node experiences a failure, redundant nodes seamlessly assume data collection responsibilities, thereby ensuring uninterrupted data continuity.

Dynamic Routing: The network utilizes a dynamic routing protocol that continuously evaluates the performance of sensor nodes. In case of a node failure or performance degradation, the routing algorithm dynamically redirects data to operational nodes, thereby averting data loss or delays.

Data Fusion: The central monitoring system utilizes sophisticated data fusion techniques to harmonize data from redundant nodes. This fusion process guarantees that data from multiple sources is amalgamated into a unified and precise representation of the machine's status.

Predictive Maintenance: Machine learning algorithms are employed to analyze the collected data, forecasting potential equipment failures. This predictive maintenance strategy facilitates proactive scheduling of maintenance tasks, mitigating the risk of unexpected downtime.

Balancing privacy preservation in data processing with performance objectives, such as energy efficiency and data accuracy, often requires trade-offs. To address these challenges, various mechanisms have been introduced, including encryption and access control mechanisms. In our study, we focused on two critical aspects: establishing secure links, aggregating data, and authenticating the integrity of aggregation results. This approach ensures energy efficiency while maintaining secure aggregation.

CONCLUSION

WSN plays an important role in various real-time applications. Therefore, the demand for these networks has increased drastically. These networks are resource-constrained and deployed in a harsh environment therefore maintaining a resource becomes an important aspect of these networks. Data aggregation has been considered as a promising solution to minimize energy consumption and prolong the network lifetime by efficiently using the available resources. However, security and maintaining energy-efficient aggregation have remained a challenging issue for the research community. Therefore, we present a novel approach that considers energy efficient and secure data aggregation by using compressed sensing and cryptographic schemes. However, some advanced key exchange mechanisms can be used to increase the robustness during user authentication. Moreover, block chain schemes also can be explored in the context of secure data exchange.

ETHICAL DECLARATION

Conflict of interest: No declaration required. **Financing:** No reporting required. **Peer review:** Double anonymous peer review.

REFERENCES

[1] V. S. Devi, T. Ravi, and S. B. Priya, "Cluster based data aggregation scheme for latency and packet loss reduction in WSN," *Computer Communications*, vol. 149, pp. 36-43, 2020.

[2] S. Hu, L. Liu, L. Fang, F. Zhou, and R. Ye, "A novel energy-efficient and privacy-preserving data aggregation for WSNs," *IEEE Access*, vol. 8, pp. 802-813, 2019.

[3] M. Naghibi and H. Barati, "SHSDA: Secure hybrid structure data aggregation method in wireless sensor networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 12, pp. 10769-10788, 2021.

[4] I. Ullah and H. Y. Youn, "A novel data aggregation scheme based on self-organized map for WSN," *The Journal of Supercomputing*, vol. 75, pp. 3975-3996, 2019.

[5] 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*, vol. 14, no. 4, pp. 1917-1942, 2021.

[6] A. M. Bongale, C. R. Nirmala, and A. M. Bongale, "Energy efficient intra-cluster data aggregation technique for wireless sensor network," *International Journal of Information Technology*, vol. 14, pp. 1-9, 2020.

[7] G. Sanjay Gandhi, K. Vikas, V. Ratnam, and K. Suresh Babu, "Grid clustering and fuzzy reinforcement — Learning based energy — Efficient data aggregation scheme for distributed WSN," *IET Communications*, vol. 14, no. 16, pp. 2840-2848, 2020.

[8] I. Ullah, H. Y. Youn, and Y. H. Han, "An efficient data aggregation and outlier detection scheme based on radial basis function neural network for WSN," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-17, 2021, doi: 10.1007/s12652-020-02703-7.

[9] R. K. Dhanaraj, K. Lalitha, S. Anitha, S. Khaitan, P. Gupta, and M. K. Goyal, "Hybrid and dynamic clustering based data aggregation and routing for wireless sensor networks," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 6, pp. 10751-10765, 2021.

[10] X. Liu, X. Zhang, J. Yu, and C. Fu, "Query privacy preserving for data aggregation in wireless sensor networks," *Wireless Communications and Mobile Computing*, pp. 1-10, 2020, doi: 10.1155/2020/9754973.

[11] Y. Lu, L. Li, H. Peng, and Y. Yang, "An energy efficient mutual authentication and key agreement scheme preserving anonymity for wireless sensor networks," *Sensors*, vol. 16, no. 6, p. 837, 2016.

[12] L. Xiong, D. Peng, T. Peng, H. Liang, and Z. Liu, "A lightweight anonymous authentication protocol with perfect forward secrecy for wireless sensor networks," *Sensors*, vol. 17, no. 11, p. 2681, 2017.

[13] J. Jung, J. Kim, Y. Choi, and D. Won, "An anonymous user authentication and key agreement scheme based on a symmetric cryptosystem in wireless sensor networks," *Sensors*, vol. 16, no. 8, p. 1299, 2016.

[14] N. Chandnani and C. N. Khairnar, "Bio-inspired multilevel security protocol for data aggregation and routing in IoT WSNs," *Mobile Networks and Applications*, vol. 27, no. 3, pp. 1030-1049, 2022.

[15] M. Ataei Nezhad, H. Barati, and A. Barati, "An authentication-based secure data aggregation method in Internet of Things," *Journal of Grid Computing*, vol. 20, no. 3, p. 29, 2022.

[16] A. Ahmed, S. Abdullah, M. Bukhsh, I. Ahmad, and Z. Mushtaq, "An energy-efficient data aggregation mechanism for IoT secured by blockchain," *IEEE Access*, vol. 10, pp. 11404-11419, 2022.

[17] T. C. Dao, N. T. Tam, N. Q. Quy, and H. T. T. Binh, "An energy-efficient scheme for maximizing data aggregation tree lifetime in wireless sensor network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12329-12344, 2023.