Received: July 20, 2024. Revised: August 17, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Data Aggregation Scheme Using Differential Evolution with Sailfish Optimization for Clustering and Routing in IoT

Srilakshmi Puli[1]* **Srinivasu Nulaka[1]** **Lavanya Patnala[2]** **Sangita Mishra[2]**
Simhadri Venkata Meena[2]

[1]*Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India*
[2]*Department of Computer Science Engineering-DS, AI & ML, Avanthi Institute of Engineering and Technology, Tagarapuvalasa, India*
* Corresponding author's Email: thotasrilakshmi1@gmail.com

Abstract: Internet of Things (IoT) facilitates connectivity in businesses and smart homes by integrating embedded technology, wireless sensor networks and data aggregation. Regular monitoring of energy usage in IoT networks is crucial due to the high energy consumption and delays in transmitting data to the Base Station (BS) by the sensor nodes. The most significant challenges in IoT include energy depletion and transmission delays. In this research, the proposed Differential Evolution with Sailfish Optimization (DESFO) model addresses large network handling, achieves maximum convergence rates, and reduces energy consumption. The Differential Evolution (DE) mutation and crossover operators enhance exploration capabilities, while SFO adaptive movement strategies improve the exploitation of the search space. Together, they achieve high convergence rates, prevent falling into local optima, provide iterative control and manage high-dimensional networks effectively. The DESFO method exhibits superior performance when compared to the existing methods, Firefly Optimization and Aquila Optimization (FF-AO), Fixed-Parameter Tractable Approximation Clustering (FPTAC), and Cluster based Reliable Data Aggregation- Sunflower Optimization (CRDA-SFO). The proposed DESFO method yields impressive results, achieving a Packet Delivery Ratio (PDR) of 96.12% at 250 nodes, a Delay of 3ms at 250node, Energy consumption of 12J at 250 respectively.

Keywords: Base station, Differential evolution, Data aggregation, Internet of Things and Sailfish optimization.

1. Introduction

Internet of Things (IoT) connects with the physical environment using sensors and actuators. Sensors collect the data which is then processed to understand the current environment [1]. IoT nodes are typically powered by limited battery sources that are difficult to charge, and the diversity of smart nodes and the demand for ubiquitous connectivity create energy efficiency challenges. Consequently, energy-efficient methods to enhance IoT networks have received significant attention [2, 3]. Furthermore, when moving an item, a signal actuator sends data to the sensor node. Sensor nodes must transmit data to the gateway node, also known as the Base Station (BS). The data from each cluster is forwarded by the sensor nodes to a Cluster Head (CH), which collects the information and transmits it to the BS [4, 5]. Optimization algorithms used for clustering can be either static or dynamic [6]. The sensor nodes in the network are capable of sensing, data aggregation, and information transmission, efficiently manipulating and converting physical parameters of distance, energy and delay [7].

To conserve energy within the network, various mechanisms for IoT with Wireless Sensor Networks (WSN) have been developed. A major mechanism utilized is data clustering and aggregation [8]. In this context, CHs are responsible for collecting information from their cluster sensor nodes and data aggregation. Cluster-based communication schemes integrate data aggregation functions to eliminate duplicate data through redundancy checks, thus avoiding multiple data transmissions [9]. In advanced

IoT systems, various configurations for efficient data processing and minimal data recovery have been recommended [10]. These configurations involve centralized data storage, such as in cloud systems or nearby distribution systems. The selection of a node as the CH within each cluster involves computational and probabilistic operations [11, 12]. Nodes with maximum residual energy are more likely to be chosen as CHs. These CHs play a crucial role in collecting data from their respective clusters and transmitting the aggregated data [13, 14]. Therefore, clustering is an effective solution for improving environmental sustainability and energy efficiency, as it extends the lifespan of the network and enhances energy efficiency [15]. The most significant challenges in IoT include high energy consumption and delays in transmitting data to Base Station (BS) by sensor nodes. In this research, the proposed Differential Evolution with Sailfish Optimization (DESFO) model addresses large network handling, achieves maximum convergence rates and reduces energy consumption.

The main contributions of the research are discussed below:

- The DE approach is employed for selecting CH in WSN-IoT due to its maximum stability and minimum energy consumption. DE selects the CH based on network nodes to handle complex networks in high dimensions.
- The SFO method identifies shortest path between CH and BS because SFO has high convergence in discovering solutions in WSN.
- In this research, the proposed DESFO approach reduces energy consumption based on population size and control iterations while transmitting data packets.

The paper is organized as follows: Section 2 provides a literature review that summarizes clustering and routing in IoT, Section 3 introduces proposed method utilized by DESFO, while Section 4 discusses the result and comparative analysis, and Section 5 discusses the conclusion.

2. Literature review

The related work about clustering and routing in IoT based on techniques are discussed along with their advantages and disadvantages.

Hosseinzadeh [16] presented a clustering and routing approach in the IoT ecosystem that focused on minimizing the power consumption. They utilized Firefly Optimization and Aquila Optimization (FF-AO) algorithms for clustering and routing. This hybrid FF-AO method enhanced the power usage and throughput while increasing the number of clusters, thereby extending network consumption and producing lower overhead than the competitors. However, the FF-AO approach had a slow convergence rate due to oscillations during clustering and routing.

Agbulu [17] implemented a Power-Efficient Compressive Data Fusion and Cluster-Based Multi-Hop Relay-Assisted Protocol (PECDF-CMRP) for IoT sensor networks. In this protocol, CHs were nominated using a multi-weight function, and the K-means approach was modified to evenly assign sensor nodes. Data aggregation was performed using single-level wavelet sparsity-based fusion. However, compressive data fusion involved collecting and aggregating data from multiple nodes, leading to increased latency due to multiple transmission stages.

Kiamansouri [18] developed a Fuzzy-based Clustering protocol called Fixed-Parameter Tractable Approximation Clustering (FPTAC) to analyze clustering and routing in IoT. This approach balanced the clusters to rapid energy consumption and extended network lifetime, selecting the CHs based on sensor information for real-time decision-making, thereby increasing the packet delivery rates and reduced delay. However, the FPTAC approach faced challenges of energy consumption and reliable information transferred due to heterogeneity of network.

Mohseni [19] introduced a Cluster-based Energy-aware Data Aggregation Routing (CEDAR) protocol for IoT using Capuchin Search Algorithm (CapSA) approach and a fuzzy logic system. These optimization algorithms addressed the constrained and global optimization, efficiently routing massive amount of information transmitted by sensor nodes between the CHs and the Base Station (BS). However, CapSA method for routing and transmitting information faced challenges of unbalanced energy consumption and uneven distribution in the initial population, leading to minimum levels of global search performance.

Guguloth Ravi [11] presented a Cluster based Reliable Data Aggregation (CRDA) scheme for IoT network to ensure the data aggregation in the energy efficient manner. These was considering the Sunflower Optimization (SFO) algorithm toe design constraints for efficiently cluster and routing in the data aggregation. However, CRDA involved the SFO algorithm suffer from premature convergence, fall in the local optimum then led to maximized the computational time.

In the overall analysis, the existing approaches have limitations of high energy consumption and delays in transmitting data to the BS by sensor nodes. In this research, the proposed DESFO model addresses large network handling, achieving maximum convergence rates and reduced energy consumption.

3. Proposed methodology

In this research, CHs are selected using the DESFO method to achieve a balanced distribution of CHs. This approach selects the intermediate and CH nodes required for routing and transmitting data. In the IoT-WSN, nodes transmit data and use gathering schemes to analyze capacity. The proposed DESFO method handles large networks, balancing exploitation and exploration. Fig. 1 illustrates block diagram of the proposed method.

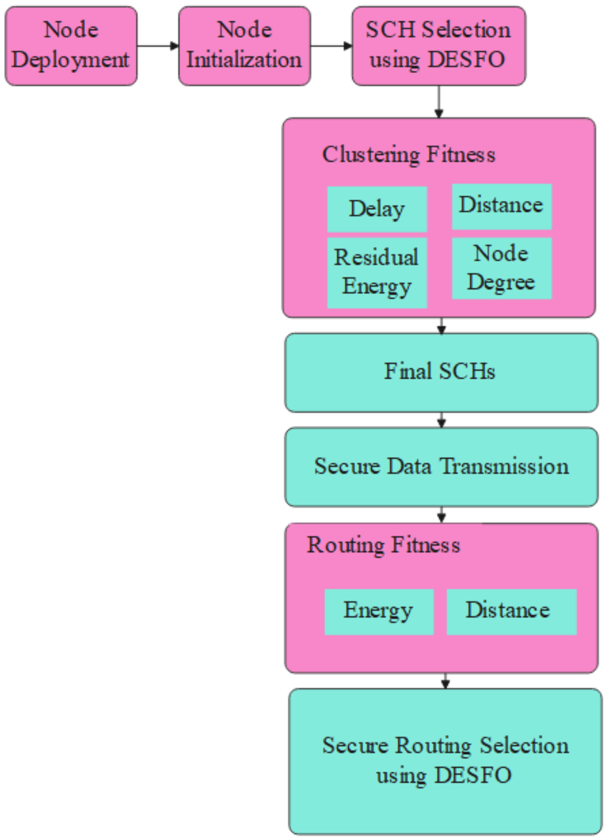

Figure. 1 Block diagram of proposed method

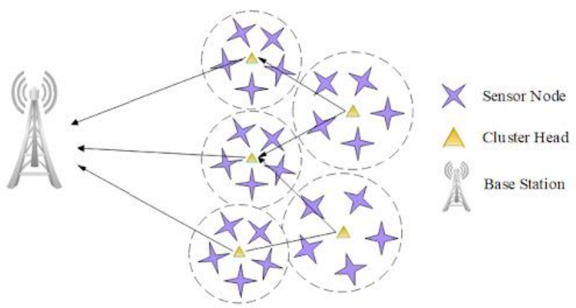

Figure. 2 The architecture of clustered WSN

3.1 Sensor deployment

Initially, nodes are arbitrarily located in IoT. Secure cluster heads and secure paths are discovered using DESFO to attain a reliable information broadcasting in the network.

Each sensor node is assigned to a gateway within its communication range, ensuring that every sensor node maintains a record of the gateways to which it is allocated. When the gateways receive data, they aggregate it to eliminate the duplicated and uncorrelated data before sending it to the BS via additional gateways acting as next-hop relay nodes.

3.2 Network model

This research considers a WSN-based IoT system with an unlimited contributing to the BS connected to network model, which indicates the arbitrarily spread nodes in the network [21]. While every node and BS are stored, their information is transmitted in power changes based on the distance. The sensor nodes store information in every round and then send it to BS. Fig. 2 depicts architecture of clustered WSN.

The proposed network model's routing protocols aim to achieve various objectives, including maximizing PDR, minimizing energy consumption and extending network's lifespan. This is done by reducing the number of message swaps between nodes and optimizing aggregation data at CHs. The data duplication is eliminated and entire information set is compressed into a single package, further enhance network efficiency and resource utilization.

3.3 Energy model

The Euclidean distance between network components serve as foundation for energy consumption model. When Euclidean distance falls inside a predetermined range (ri), the data transfer between sensor nodes is assessed. To ensure successful data transfer, simultaneous transmission data from nodes within interference level of receiving node j is avoided. The energy model calculates how much energy is needed to send a specific amount of bits l across the network. Eqs. (1) and (2) define energy consumption for data transmission(E_{tr}).

$$E_{tr}(l, D_0) = \begin{cases} lE_{elec} + l \in_{fs} D_{ij}^2 D_{ij} < D_0 \\ lE_{elec} + l \in_{fs} D_{ij}^4 \; otherwise \end{cases} \quad (1)$$

$$E_{tr}(l) = lE_{elec} \quad (2)$$

In Eq. (1), several factors contribute to energy consumption during information transmission by considering E_{elec}, which is energy consumed by electronic circuitry of sensor node.

3.4 Secure cluster head selection using DESFO

In this research, a DE-based clustering algorithm is employed for selecting CHs and allowing non-CH nodes to join their nearest CHs. DE, a stochastic and population-based evolutionary algorithm, is widely utilized for solving optimization problems. The DE aims to explore different configurations of CHs by leveraging mutation and crossover to discover the optimal solutions that balance energy efficiency and data aggregation. The SFO refines and exploits the promising solutions to be found by DE. The SFO helps in fine-tuning the selection and positioning processes of CH to enhance the network reliability and scalability. The DE approach involves three stages: mutation, crossover, and selection, iteratively contributing to improvement based on the population size, and its ability to handle complex networks and high dimensions. Each vector in the population is evaluated using a fitness function to assess the solution quality.

Mutation: During this stage, the DE algorithm applies a mutation operator to cause a recent donor vector (DV) for every target solution in each epoch (n). The DV is built by the scaling difference vector between 2 other vectors by including outcomes to a 3rd solution, as shown in Eq. (3). This approach allows DE algorithm to explore the solution space, effectively improving selection of Cluster Heads and optimizing the network's performance.

$$K_{t,H+1} = y_{j1,H} + F\left(y_{j2.H} - y_{j3.H}\right) \tag{3}$$

In the process, tri distinct integers $j1, j2, and\ j3$ are arbitrarily collected and $\in [1, NP]$, where NP is a positive integer greater than or equal to 4. Moreover, these integers are dissimilar from running index t. The amplification of differential $\left(y_{j2.H} - y_{j3.H}\right)$ is then amplified by a standard element F, ranging from $0\ to\ 2$.

Crossover: Following mutation, a crossover search operation manufacturing offspring is taken in a vector form from target solution. The most frequently used crossover operators are exponential and binomial crossovers, which are relatively uncomplicated. The decision variable is $m,$ and in the scenario where $(rand \leq C_j)$, the expression represented by Eq. (4).

$$x_{a,b,H} = \begin{cases} x_{a,b,H}\ if\ rand(b) \leq C_j or\ b = b_{rand} \\ \quad x_{a,b,H}\ \ otherwise' b = 1,2, \dots . D \end{cases} \tag{4}$$

Here, arbitrarily value b_{rand} is collected from a range of specified values, indicating that V_z is chosen randomly, and the b^{th} calculation is indicated by $rand(b)$ from a constant arbitrarily number of [0,1] range. The created variable obtained from trial vector ensures that at least one variable is changed. The crossover rate C_j controls the count of variables obtained from DV and guarantees that $K_{n,H+1}$ provides at least 1 parameter to $x_{a,b,H}$.

Selection: In this section, the operator evaluates optimal solution by balancing objective function values of both offspring and parent. If the offspring has a minimum solution, it is maintained for upcoming iterations. The generation of parent vector is represented by Eq. (5).

$$y_{a,H+1} = \begin{cases} x_{a,H} \quad if\ \left(f\left(x_{a,H}\right) \leq \left(\, y_{a,H}\right)\right) \\ y_{a,H} \qquad\qquad\qquad otherwise \end{cases} \tag{5}$$

To determine $H+1$, $y_{a,H+1}$ indicates the trial vector calculated against target vector $y_{a,H}$ utilized greedy criterion. The trial vector $y_{a,H+1}$ replaces the target vector $x_{a,H}$, while the original target vector $x_{a,H}$ values are kept.

3.4.1. Sailfish optimizer

This section introduces Sailfish Optimizer (SFO), a population-based algorithm rooted in swarm intelligence. In group hunting scenarios, predators exert less effort to capture prey when compared to the solitary hunting endeavors. Group hunting strategies vary from simple, uncoordinated attacks to more complex patterns involving coordinated alternation. Such strategies enable hunters to conserve energy while effectively balancing between exploration and exploitation. Sailfish, known as the fastest fish in ocean, achieve speeds of up to approximately $100\ km/h$. They involve group hunting scheme by herding smaller fish like sardines towards the water surface. Sailfish employ tactics such as slashing with their rostrum to injure multiple sardines at once or by using precise taps to destabilize individual sardines [22].

Initialization: The SFO is a population-based metaheuristic algorithm where the sailfish represent the candidate solutions and problem variables corresponding to their positions in search space. The algorithm operates by populating solution space with sailfish, navigating through a tri-dimensional hyperspace using their position vectors. Sailfish are pivotal entities scattered across the search space, while sardines collaborate to improving their positions within this domain. Notably, sailfish consume sardines during their search, updating their positions based on the success of their attack alternation strategy in finding optimal solutions.

Attack Alternation Strategy: The SFO algorithm recognize sardines with better fitness values as injured fish, with updated location indicated as L_{srdina}^{a} at a^{th}iteration. During every iteration, the locations of both sardines and sailfish are improved. The location of a sailfish in the a^{th} iteration is improved using elite sailfish $L_{SIfbest}^{a}$, while an injured sardine based on particular criteria.

The positions of sailfish and sardines are enhanced in every iteration, denoted by $a+1$, with elite and injured statuses altering or improving location of sailfish to a recent 1, indicated as L_{Slf}^{a+1}. The update is done according to Eq. (6).

$$L_{Slf}^{a+1} = L_{SIfbest}^{a} - \mu_a \left(rand * \frac{L_{SIfbest}^{a} + L_{srdina}^{a}}{2} - L_{Slf}^{a} \right) \quad (6)$$

Where, the values of $rand \in (0,1)$ are arbitrarily values, and coefficient μ_a is formulated by Eq. (7).

$$\mu_a = (3 * rand * PrD - PrD) \quad (7)$$

In every iteration, the prey density (PrD) denotes the number of prey available, evaluated using Eq. (3). The number of preys minimized during the hunting phase, leading to a corresponding minimization of PrD values. Sailfish and sardine numbers are represented by H_{Slf} and H_{Srd} , respectively. The Num_{SlF} evaluated according to Eqs. (8) and (9), respectively.

$$PrD = 1 - \frac{H_{Slf}}{H_{Slf} - H_{Srd}} \quad (8)$$

$$H_{Slf} = H_{Srd} * Prcent \quad (9)$$

Where, the percentage denotes the proportion of population sardine that makes up begin sailfish population. It is also assumed that begin count of sardines exceeds the sailfish. The updated position of sardines in every iteration is given by Eq. (10). The last position and upgrade location of the sardine are corresponding by means of H_{Srd}^{a} and H_{Srd}^{a+1} , respectively. The attack power (AP) of the sailfish at each iteration a is evaluated by Eq. (11).

$$L_{Srd}^{a+1} = rand * \left(L_{SIfbest}^{a} + L_{srd}^{a} + AP \right) \quad (10)$$

$$AP = A * (1 - 2 * itr * v)) \quad (11)$$

The AP is crucial in evaluating the number of sardines that upgrade their location and displacement that is extended to the sardines. Minimizing AP facilitates convergence of search agents. The parameters γ indicate the updated position and δ of sardines, as computed using Eqs. (12) and (13).

$$\gamma = AP * H_{Srd} \quad (12)$$

$$\delta = AP * k \quad (13)$$

Where, L_{Srd} and k indicate number of variables and sardine, correspondingly. If a sardine exceeds fitness quality of any sailfish, its location is adjusted to follow that sardine. Conversely, if a sardine does not surpass fitness level, it is eliminated from the population. To effectively explore search space, both sailfish and sardines are selected arbitrarily. Minimizing the AP parameter after every iteration allows sardines to escape from most aggressive sailfish. The SFO approach helps perform a balance between exploration and exploitation of search space, with AP variable utilized to search for the optimal solutions, wherein balance between these 2 aspects is considered.

3.4.2. Proposed differential evolution with sailfish optimization

In this research, the integration of DE with SFO enhances the performance of CH selection by improving exploration and exploitation capabilities of the optimization process. The DE mutation and crossover operators enhance exploration capabilities, while SFO adaptive movement strategies improve exploitation of the search space. They are combined to achieve a high convergence rate, avoid falling into local optima, iterate control, and handle high-dimensional networks. The DESFO is combined to find an optimal set of CHs that maximizes the network performance while minimizing energy consumption and delay. DE provides global exploration through mutation and crossover, while SFO offers local refinement through adaptive movement. The robust and efficient solution is achieved by leveraging DE's exploration capabilities and SFO's adaptive optimization. DE helps maintain population diversity, reducing the risk of premature convergence that the SFO approach encounters in complex optimization landscapes. The proposed DESFO algorithm allocates an equal number of iterations to DESFO with 50 iterations each. DE optimizes by choosing the initial iterations to obtain an optimal solution. This solution is then passed to SFO, which enhances the choosing of relevant features. Below is a detailed explanation of those stages:

Initial Population Generation: The initial stage of the DESFO algorithm causes X of the population's location denoted as values in the dimensional space of D. The size of the population is evaluated utilizing a particular mathematically expressed in Eq. (14).

$$X = Round\big(10 + 2 * \sqrt{D}\big). \tag{14}$$

Where, X indicates the overall count of position and D denoted issues dimensionality, the location matrix is defined in Eq. (15), the j^{th} solution is represented by $M_{i,j}$ where j denoted j^{th} component. M is begin population caused within the predefined bounders.

$$M = \begin{bmatrix} m_{1,1.}m_{1,2,} \dots . m_{1,p} \\ m_{2,1.}m_{2,2,} \dots . m_{2,p} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ m_{X,1.}m_{X,2,} \dots . m_{X,p} \end{bmatrix}$$

$$M_i^u = u(0,1) * (UB - LB) + LB \tag{15}$$

Where, UB and LB indicate the upper and lower bounds, respectively.

Position Updated in DESFO: The improving position utilizes the equation of DE and SFOs determined in Eq. (10) after updating the position of population from DESFO. Then, the movement strategies are adjusted to minimize the energy consumption, control iteration, and balance exploration and exploitation.

3.5 Fitness for SCH selection

To reduce energy consumption and improve communication efficiency, a fitness function incorporating node degree, delay and communication distance within the cluster is formulated. Nodes with the lowest fitness function values are selected for CH roles. CHs are distributed evenly across the monitoring area, contributing to reduced energy consumption among member nodes.

Residual Energy: The Residual Energy ($f1$) usage of the SCH is essential because it carries out different functions of information collection and data aggregation to be distributed over the network. The sensor node with maximum residual energy is preferred to be selected as an SCH to ensure that the aforementioned tasks are performed appropriately to achieve a reliable transmission. Eq. (16) mathematically denotes the residual energy.

$$f_1 = \sum_{i=1}^{m} \frac{1}{E_{SCH_i}} \tag{16}$$

Where, E_{SCH_i} is the residual energy of i^{th}SCH.

Distance: The distance transmitter between CHs and sink is evaluated by the amount of Euclidean distance among every CH in the path and demonstrated through distance ($f2$), as expressed in Eq. (17).

$$f2 = \sum_{j=1}^{m} \big(\sum_{i=1}^{lj} dis(S_i, CH_j)/l_j\big) \tag{17}$$

Here, $CH_i(x)$ and $CH_i(y)$ characterize x and y coordinates of i^{th}CH in path correspondingly, and l_i is the count of sensor nodes belonging to CH_j.

Node Degree: The node degree considers the number of sensor node belonging to CH with fewest sensors, as the minimum energy loss over time is higher in clusters with more members with CHs. The node degrees ($f3$) are expressed in Eq. (18).

$$f3 = \sum_{i=1}^{m} l_i \tag{18}$$

Where, l_i represents the count of sensor nodes belonging to CH_i.

Delay: The delay defines the $f4$ interval [0.1] with a ratio of CH in the WSN to the total number of sensor nodes (Y). Delay ($f4$) is expressed in Eq. (19).

$$f4 = \frac{max_{i=1}^{x}(CH_i)}{Y} \tag{19}$$

The delay should be minimum which is attained by lowering quantity of nodes in cluster.

The weight values are allocated for every objective value and multiple objectives that are modified into a single objective function. The weights are indicated as $\delta_1, \delta_2, \delta_3$ & δ_4, and a single objective function is shown in Eqs. (20) and (21).

$$f = \delta_1 f_1 + \delta_2 f_2 + \delta_3 f_3 + \delta_4 f_4 \tag{20}$$

$$\text{Where, } \sum_{i=1}^{4} \delta_i = 1,\ \delta_i \epsilon (0,1) \tag{21}$$

Where, the values of $\delta_1, \delta_2, \delta_3$ and δ_4 are 0.35, 0.22,0.1 and 0.3, respectively.

3.6 Cluster formation using potential function

After selecting CHs by SFO, the sensor node is assigned to CHs using the constant function denoted in (22) equation. The clustering process also helps in balancing the load among the CHs, preventing any single CH from becoming overwhelmed and ensuring even distribution of energy consumption across the network. To enhance the scalability for supporting large-scale deployments, the communication overhead is reduced with a simplified network management. The data transmission efficiency performs lower latency and increased throughput by optimizing data transmission paths. The cluster stage computes the overall CH to collect the sensor nodes, balance the cluster to accommodate the size of the nodes, and strike a balance in energy consumption. These nodes are concerned with the correlation between distance and energy as displayed in Eq. (22).

$$SN_p = \frac{z \times Energy\ (CH_j)}{Distance\ (s_i, CH_j)} \tag{22}$$

Where, the sensor node's potential is denoted as SN_p, Z is indicated as a proportionality constant with (s_i, CH_j), particularly with a distance among the cluster head CH_j and sensor s_i. The energy (CH_j) indicated residual energy of CH with allocated sensor specifying CH with maximum potential.

3.7 Route recovery stage using DESFO

The performance of the DE approach is sensitive in population size to the transmitted parameters in the intermediate node, such as the attractiveness coefficient and the randomization factor. Routing the recovery stage using DESFO minimizes the energy consumption while maintaining acceptable distance metrics and dynamically adjusting routing paths to enhance the network performance. This approach maintains network connectivity and continuous data flow despite node failures or changes in topology, thereby improving the overall reliability of the network. The SFO does not converge quickly because of some other optimization algorithms, especially on the complex or high-dimensional problems. This slower convergence is a limitation when efficiency is a critical factor. The DESFO method is also used to perform route discovery. The steps involved in route discovery are as below:

- The initial path solution is transmitted using the Cluster head for the route discovery. The dimension path considers an equal amount provided that to the cluster existing in the route.
- The fitness metric computations utilize the residual energy, distance, delay and node degree, as shown in Eq. (23), thereby improving the location that the route discovery is performed based on the iterative method in DESFO.

$$Routing\ fitness = \delta_1 f_1 \times \sum_{i=1}^{m} \frac{1}{E_{SCH_i}} + \delta_2 f_2 \times \sum_{i=1}^{m} dis(S_i, CH_j)/l_j + \delta_3 f_3 \times \sum_{i=1}^{m} l_i + \delta_3 f_3 \times \frac{max_{i=1}^{x}(CH_i)}{Y} \tag{23}$$

Where, δ_1, δ_2 and δ_3 are assigned to the fitness metrics in the routing process. These metrics are crucial for selecting an optimal and secure route to enhance security of IoT-WSN while improving the efficiency of improving information delivery.

3.8 Cluster maintenance

In this section, cluster preservation is determined as an important phase to balance load between clusters. Therefore, the cluster preservation stage is necessary to prevent node failure and increase throughput during transmission data from source node to BS. The selected updated best node is chosen as the CH, with routing path between BS and CHs evaluated using DESFO algorithm. In this proposed methodology, DESFO algorithm is used to achieve an effective CH selection by considering various parameters of node degree, distance, energy, and delay. To avoid node failure during data transfer, BS stably monitors the nodes' residual energy. Thus, an energy-efficient IoT-WSN is employed to improve throughput and maximum total count of packets transmitted to BS during data transmission.

4. Experimental setup

In this research, DESFO techniques are simulated using MATLAB R2020b with a system configuration of i7 processor, 16GB RAM and Windows 10 OS. The performance of the proposed method is assessed using several performance metrics namely, Packet Delivery Ratio (PDR), Energy Consumption, Throughput, and Delay. The mathematical expressions of these performance metrics are provided in Eqs. (24) and (25).

$$PDR = \frac{\sum Packet\ received}{\sum Packet\ transmitted} \tag{24}$$

$$Delay = \frac{\sum Transaction\ time - Receiving\ time}{Number\ of\ packets} \tag{25}$$

In this context, data loss is determined by calculating the percentage decrease and is assessed using metrics of PDR, Throughput, Energy consumption and Delay.

4.1 Performance analysis

In this section, the proposed DESFO approach observes a relevant feature that results in the population size causing an increase in energy between the source and destination. Furthermore, the proposed techniques demonstrate a better throughput, PDR, minimized energy consumption, and delay. The existing algorithm considers the load on the CH, average distance clustering, and routing formation. Table 1 represents the quantitative analysis of delay utilizing the number of nodes. The performance of proposed methods DESFO is measured and compared with existing methods: FF, AO, TLBO and SFO. The DESFO achieves a superior delay scoring of 5ms, 6ms, 5.5ms, 5ms, and 7ms at the number of nodes of 100, 200, 300, 400, 500, respectively. The DESFO approach minimizes the delay by efficiently selecting the optimal routing paths and dynamically adjusting to the network conditions. The optimization technique ensures low transmission delay and the improved overall network performance.

Table 2 represents a quantitative analysis of PDR utilizing the number of nodes. The performance of proposed methods DESFO is measured and compared with existing methods: FF, AO, TLBO and SFO. The DESFO achieves better PDR scoring of 96.12%, 96.11%, 96.02%, 95.96%, and 95.90% at number of nodes of 100, 200, 300, 400, 500, correspondingly. The DESFO approach achieves the highest PDR across different number of nodes due to its efficient route selection, dynamic path adjustments and optimal pathfinding, enhancing the overall network performance and maintaining high PDR.

Fig. 3 illustrates the graphical representation of the throughput performance. The performance of the proposed methods DESFO is measured and i7compared with the existing methods namely, FF, AO, TLBO and SFO. The DESFO attains a superior throughput scoring of 73kbps, 85kbps, 80kbps, 82kbps, and 78kbps simultaneously at the number of nodes of 100, 200, 300, 400, 500. The DESFO approach transfers the maximum information bits in contrast to the existing algorithm due to energy efficient CHs. The dynamic balancing in the network load reduces the packet loss and maximizes the data transmission rates. The proposed method achieves maximum throughput by successfully transferring the information packets to the BS.

Table 1. Performance Analysis of Delay (ms)

Methods	Number of Nodes				
	100	**200**	**300**	**400**	**500**
FF	7	10	9	9	11
AO	6	9	8	8	10
TLBO	7	8	7	7	9
SFO	6	7	6	6	8
DESFO	**5**	**6**	**5.5**	**5**	**6**

Table 2. Performance Analysis of PDR (%)

Methods	Number of Nodes				
	100	**200**	**300**	**400**	**500**
FF	92.83	92.08	94.25	91.58	91.15
AO	93.42	93.02	93.86	92.98	92.08
TLBO	94.63	94.63	94.12	93.25	93.73
SFO	95.85	95.12	95.85	94.74	94.02
DESFO	**96.12**	**96.11**	**96.02**	**95.96**	**95.90**

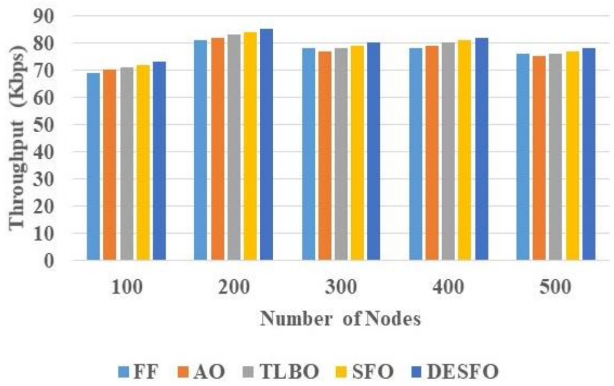

Figure. 3 Graphical representation of Energy Consumption performance

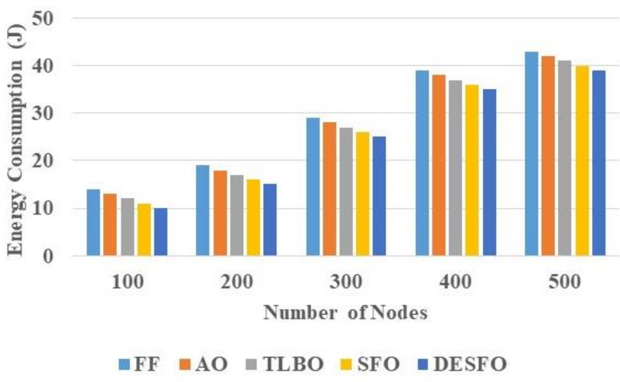

Figure. 4 Graphical representation of Energy Consumption performance

Fig. 4 presents a graphical representation of the energy consumption performance. The performance of proposed method, DESFO is measured and compared with existing methods named FF, AO, TLBO and SFO. The DESFO achieves commendable energy consumption with the scoring of 10J, 15J, 25J, 35J, and 39J at the number of nodes correspondingly being 100, 200, 300, 400, 500. The DESFO minimizes energy consumption by CHs and balanced energy usage across nodes with a secure route with lesser transmission distance to BS and minimized energy usage over the network.

4.2 Comparative analysis

The comparative analysis of DESFO model with the existing techniques such as FF-AO [16], FPTAC [18], CapSA [19] and CRDA-SFO [11] is provided in this section. The comparison is carried out under different scenarios, as specified in Tables 3 to 7. The performance metrics used are, PDR, delay, energy consumption and throughput. The different scenarios are considered based on the existing methods' comparison with the DESFO approaches. Table 3 to 7 show the comparison of DESFO with HAOFA [16], FPTAC [18], CapSA [19] and CRDA-SFO [11], respectively.

Table 3. Specification of different scenarios

Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4
	HAOFA **[16]**	**FPTAC [18]**	**CapSA [19]**	**CRDA-SFO [11]**
Area	$500 \times 500m^2$	$100 \times 100m^2$	$100 \times 100m^2$	$1000 \times 1000m^2$
No of node	400	100 - 500	100 -500	200-1000
Initial Energy	0.5J	0.5J	0.1J	2J

Table 4. Comparison of DESFO with Scenario 1

Methods	**Number of Rounds**	**Energy Consumption (J)**	**PDR (%)**	**Delay (ms)**	**Throughput (Kbps)**
HAOFA [16]	250	25	90	5	84
	500	26	86	5.1	NA
	750	50	82	5.2	NA
	1000	52	78	5.5	NA
	1250	70	80	6.1	NA
	1500	78	70	6.5	NA
	1750	79	68	6.6	NA
	2000	80	67	6.9	NA
Proposed DESFO methods	**250**	**12**	**96.12**	**3**	**75**
	500	**25**	**89**	**4**	**80**
	750	**43**	**83**	**4.5**	**85**
	1000	**50**	**91**	**4**	**90**
	1250	**69**	**80**	**6.0**	**91**
	1500	**75**	**74**	**6.1**	**92**
	1750	**72**	**71**	**6.2**	**93**
	2000	**79**	**70**	**6.3**	**94**

Table 5. Comparison of DESFO with Scenario 2

Methods	**Number of nodes**	**Delay (ms)**	**Network Lifetime (Rounds)**	**PDR**
FPTAC [18]	100	3	1500	0.75
	200	8	1400	0.85
	300	10	1386	0.88
	400	19	1300	0.98
	500	28	1285	NA
Proposed DESFO methods	**100**	**2**	**1525**	**0.76**
	200	**7**	**1485**	**0.86**
	300	**9**	**1389**	**0.89**
	400	**18**	**1325**	**0.99**
	500	**27**	**1291**	**0.96**

Received: July 20, 2024. Revised: August 17, 2024.

Table 6. Comparison of DESFO with Scenario 3

Methods	Number of nodes	Delay (s)	Energy Consumption (J)	PDR
CapSA [19]	100	0.8	8	0.99963
	200	0.7	12	0.9961
	300	1.9	21	0.9957
	400	2.5	32	0.9952
	500	3.1	38	0.9953
Proposed DESFO methods	**100**	**0.7**	**7**	**0.9964**
	150	**0.6**	**11**	**0.9962**
	200	**1.8**	**20**	**0.9958**
	250	**2.4**	**31**	**0.9953**
	500	**3.0**	**37**	**0.9954**

Table 7. Comparison of DESFO with Scenario 4

Methods	Number of nodes	Throughput (Mbps)	Energy Consumption (J)	Network Lifetime (%)
CRDA-SFO [11]	200	8876	95.45	99.79
	400	7878	254.43	88.80
	600	7052	275.68	86.34
	800	6205	354.18	83.71
	1000	5667	403.56	82.01
Proposed DESFO methods	**200**	**8881**	**94.12**	**99.85**
	400	**7881**	**253.35**	**87.89**
	600	**7059**	**274.55**	**85.42**
	800	**6212**	**353.11**	**82.79**
	1000	**5671**	**402.39**	**81.12**

4.3 Discussion

The investigation of the outcomes from the DESFO approach is discussed in detail in this section. DESFO is involved in clustering and routing in IoT, initially performed by sensor nodes to assume the CHs to transmit to the BS. Clustering and routing using optimization techniques are compared with the existing methods such as PSO, SFO, and ACO. DESFO performs the clustering phase by balancing the cluster nodes through enhanced exploitation and exploration search capacities, achieved by altering the updating position and population size. This analysis confirms that DESFO has a minimum energy consumption of 9J and a PDR of 98.90%, outperforming the HAOFA, FPTAC, CapSA, and CRDA-SFO approaches. The scenario 1 consider the 500×500 and energy is $5j$, DESFO consistently consumes minimized energy than HAOFA across all node densities indicating better energy efficiency. The PDR show better result than HAOFA across all node densities indicating more reliable packet delivery. The delay is consider minimized in DESFO to efficient routing path reducing the time packets spend in transit. The throughput values increasing with high node densities suggests that algorithm scales and increase data transmission demands effectively. The Scenario 2 consider the $100 \times 100m^2$ and energy is $5j$, DESFO minimized delay in efficient routing paths and better handling of packet transmission. The network lifetime in DESFO algorithm is better effective energy management, which optimize node activity and routing to prolong network operation. The PDR in DESFO is imporved reliability in packet transmission due to better preformed in the proposed method that minimize packet lose. The Scenario 3 consider the $100 \times 100m$ and energy is $0.1J$, DESFO algorithm show minimized delay compared to CapSA, more efficient packet delivery. The DESFO consistently consumes minimized energy than CapSA across all node densities achieve better energy efficiency. The PDR achieve better result in the proposed method than CapSA. The scenario 4 consider the $1000 \times 1000m$ and energy is $0.2J$, DESFO algorithm achieve better throughput value efficient routing and data transmission. The minimized energy consumption in DESFO approach, which optimize node and routing to minimize energy usage. The network lifetime in DESFO better energy management compared to CRDA-SFO approach.

Received: July 20, 2024. Revised: August 17, 2024.

5. Conclusion

In this section, the proposed DESFO model handles the large network with maximum convergence rate and reduced energy consumption. The DE mutation and crossover operators augments the exploration capabilities, while the SFO adaptive movement strategies improve the exploitation of search space. Together, they achieve high convergence rates, prevent fall into local optima, provide iterative control, and manage the high-dimensional networks effectively. The DESFO method exhibits superior performance when compared to the existing methods: HAOFA, FPTAC, and CRDA-SFO. The proposed DESFO method yields impressive results, achieving a PDR of of 96.12% at 250 nodes, a Delay of 3ms at 250node, Energy consumption of 12J at 250 respectively. In the future, various parameters should be considered for analysing data aggregation using hybrid techniques.

Notation

Notation	Description
E_{tr},E_{elec}	Energy consumption for data transmission
$(y_{j2.H} - y_{j3.H})$	Amplification f differential
m	Decision variable
b_{rand}	Range of specified values
V_z	Randomly value
C_j	Crossover rate control the count of variable
$y_{a,H+1}$	Trial vector
$y_{a,H}$	Target vector
a^{th}	iteration
$L^{a}_{SIfbest}$	Updated elite sailfish
$rand \in (0,1)$	Arbitrarily value
μ_a	Coefficient
(PrD)	Number of prey available
H^{a}_{Srd} and H^{a+1}_{Srd}	Last and upgrade position
γ	Updated position
δ	Sardines
L_{Srd} and k	Variable and sardine
X	Overall count position
UB and LB	Upper and lower bounds
$f1, E_{SCH_i}$	Residual Energy
$f2$	Distance
$f3$	Node Degree
l_i	Number of sensor node
$f4$	delay
SN_p	Sensor node
Z	Proportionality constant
CH_j	Cluster head
s_i	Sensor
δ_1, δ_2 and δ_3	Fitness metrics

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation done by 3rd and 4th, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st and 5th author. The supervision and project administration, have been done by 2nd author.

References

[1] N. Chandnani, and C. N. Khairnar, "An analysis of architecture, framework, security and challenging aspects for data aggregation and routing techniques in IoT WSNs", *Theoretical Computer Science*, Vol. 929, pp. 95-113, 2022.

[2] M. Sadrishojaei, N. J. Navimipour, M. Reshadi, and M. Hosseinzadeh, "A new clustering-based routing method in the mobile internet of things using a krill herd algorithm", *Cluster Computing*, Vol. 25, No. 1, pp. 351-361, 2022.

[3] 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.

[4] M. K. Roberts, and P. Ramasamy, "An improved high performance clustering based routing protocol for wireless sensor networks in IoT", *Telecommunication Systems*, Vol. 82, No. 1, pp. 45-59, 2023.

[5] K. Jaiswal, and V. Anand, "FAGWO-H: A hybrid method towards fault-tolerant cluster-based routing in wireless sensor network for IoT applications", *The Journal of Supercomputing*, Vol. 78, No. 8, pp. 11195-11227, 2022.

[6] L. Rajesh, and H. S. Mohan, "Adaptive group teaching based clustering and data aggregation with routing in wireless sensor network", *Wireless Personal Communications*, Vol. 122, No. 2, pp. 1839-1866, 2022.

[7] P. P. I. Vazhuthi, A. Prasanth, S. P. Manikandan, and K. K. D. Sowndarya, "A hybrid ANFIS reptile optimization algorithm for energy-efficient inter-cluster routing in internet of things-enabled wireless sensor networks", *Peer-to-Peer networking and applications*, Vol. 16, No. 2, pp. 1049-1068, 2023.

[8] G. Saranraj, K. Selvamani, and P. Malathi, "A novel data aggregation using multi objective based male lion optimization algorithm (DA-

MOMLOA) in wireless sensor network", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 13, No. 12, pp. 5645-5653, 2022.

[9] N. Chandnani, and C. N. Khairnar, "Quality of Service (QoS) Enhancement of IoT WSNs Using an Efficient Hybrid Protocol for Data Aggregation and Routing", *SN Computer Science*, Vol. 4, p. 762, 2023.

[10] V. S. Badiger, and T. S. Ganashree, "Data aggregation scheme for IOT based wireless sensor network through optimal clustering method", *Measurement: Sensors*, Vol. 24, p. 100538, 2022.

[11] G. Ravi, M.S. Das, and K. Karmakonda, "Reliable cluster based data aggregation scheme for IoT network using hybrid deep learning techniques", *Measurement: Sensors*, Vol. 27, p. 100744, 2023.

[12] S. Yang, X. Long, H. Peng, and H. Gao, "Optimization of heterogeneous clustering routing protocol for internet of things in wireless sensor networks", *Journal of Sensors*, Vol. 2022, p. 4327414, 2022.

[13] M. Bilal, E. U. Munir, and F. K. Alarfaj, "Hybrid Clustering and Routing Algorithm with Threshold-Based Data Collection for Heterogeneous Wireless Sensor Networks", *Sensors*, Vol. 22, p. 5471, 2022.

[14] R. Nagaraju, C. Venkatesan, J. Kalaivani, G. Manju, S. B. Goyal, C. Verma, C. O. Safirescu, and T. C. Mihaltan, "Secure Routing-Based Energy Optimization for IoT Application with Heterogeneous Wireless Sensor Networks", *Energies*, Vol. 15, No. 3, p. 4777, 2022.

[15] G. Ravi, M. S. Das, and K. Karmakonda, "Energy Efficient Data Aggregation Scheme using Improved LEACH Algorithm for IoT Networks", *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 11, No. 2s, pp. 174-188, 2023.

[16] M. Hosseinzadeh, L. Ionescu-Feleaga, B.-Ș. Ionescu, M. Sadrishojaei, F. Kazemian, A. M. Rahmani, and F. Khan, "A Hybrid Delay Aware Clustered Routing Approach Using Aquila Optimizer and Firefly Algorithm in Internet of Things", *Mathematics*, Vol. 10, p. 4331, 2022.

[17] G. P. Agbulu, G. J. R. Kumar, V. A. Juliet, and S. A. Hassan, "PECDF-CMRP: A Power-Efficient Compressive Data Fusion and Cluster-Based Multi-hop Relay-Assisted Routing Protocol for IoT Sensor Networks", *Wireless Personal Communications*, Vol. 127, No. 4, pp. 2955-2977, 2022.

[18] E. Kiamansouri, H. Barati, and A. Barati, "A two-level clustering based on fuzzy logic and content-based routing method in the internet of things", *Peer-to-Peer Networking and Applications*, Vol. 15, No. 4, pp. 2142-2159, 2022.

[19] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic", *Peer-to-Peer Networking and Applications*, Vol. 16, No. 1, pp. 189-209, 2023.

[20] K. Lakshmanna, N. Subramani, Y. Alotaibi, S. Alghamdi, O. I. Khalafand, and A. K. Nanda, "Improved Metaheuristic-Driven Energy-Aware Cluster-Based Routing Scheme for IoT-Assisted Wireless Sensor Networks", *Sustainability*, Vol. 14, p. 7712, 2022.

[21] C. Lei, "An energy-aware cluster-based routing in the Internet of things using particle swarm optimization algorithm and fuzzy clustering", *Journal of Engineering and Applied Science*, Vol. 71, p. 135, 2024.

[22] M. Rizwanullah, H. Alsolai, K. M. Nour, A. S. A. Aziz, M. I. Eldesouki, and A. A. Abdelmageed, "Hybrid Muddy Soil Fish Optimization-Based Energy Aware Routing in IoT-Assisted Wireless Sensor Networks", *Sustainability*, Vol. 15, p. 8273, 2023.