

A RELIABLE ROUTING MECHANISM WITH ENERGY-EFFICIENT NODE SELECTION FOR DATA TRANSMISSION USING A GENETIC ALGORITHM IN WIRELESS SENSOR NETWORK

Sateesh Gudla^{1,2}, Nageswara Rao Kuda³

¹Dept. of Computer Science and Engineering, JNTUK, Kakinada, AP, India

²Dept. of Computer Science and Engineering,

Lendi Institute of Engineering and Technology(A), JNTU Kakinada, Vizianagaram, AP, India

³Dept. of Computer Science and Systems Engineering, AUCE(A), Andhra University,
AP, India

Abstract. *Energy-efficient and reliable data routing is critical in Wireless Sensor Networks (WSNs) application scenarios. Due to oscillations in wireless links in adverse environmental conditions, sensed data may not be sent to a sink node. As a result of wireless connectivity fluctuations, packet loss may occur. However, retransmission-based approaches are used to improve reliable data delivery. These approaches need a high quantity of data transfers for reliable data collection. Energy usage and packet delivery delays increase as a result of an increase in data transmissions. An energy-efficient data collection approach based on a genetic algorithm has been suggested in this paper to determine the most energy-efficient and reliable data routing in wireless sensor networks. The proposed algorithm reduced the number of data transmissions, energy consumption, and delay in network packet delivery. However, increased network lifetime. Furthermore, simulation results demonstrated the efficacy of the proposed method, considering the parameters energy consumption, network lifetime, number of data transmissions, and average delivery delay.*

Key words: *Genetic algorithm, Energy-efficient routing path, Data transmissions, Lifetime, Wireless sensor networks*

1. INTRODUCTION

Wireless sensor networks (WSNs) are networks of many static or mobile sensors that use self-organization and multi-hop communication. WSNs use cooperative sensing, collecting, computation, and transmission to cover the data of sensing objects in a specific region and then deliver it to the sink node or base station. WSN emerged as a

Received August 29, 2022; revised November 26, 2022, December 02, 2022 and December 24, 2022;
accepted January 18, 2023

Corresponding author: Sateesh Gudla

Department of Computer Science and Engineering, JNTUK, Kakinada, AP, India

E-mail: sateesh.research@gmail.com

significant paradigm for the rapid collection of data because of its widespread use in time-critical applications such as tsunami warnings, chemical assault detection, forest fire detection, and infiltration detection in military surveillance [1, 2, 3], etc. Considering the significance of data acquired, many of these applications necessitate reliable information quickly. A sensor network's primary purpose is information dissemination, but this cannot take place if essential data is lost due to unforeseen node failure or the unpredictable nature of a wireless communication channel. Hence achieving reliable data delivery is a challenging issue.

On the other hand, the wireless sensor node is engineered on tiny circuits, and the sensors are low-cost and power-restricted [4]; energy is a critical resource since it affects the lifetime of individual sensor nodes and the entire network. Henceforth current WSN routing techniques focus on discovering an efficient route to the sink to diminish energy usage and prolong the network lifetime.

Hence in WSNs, it is challenging to deliver energy-efficient, reliable, low-latency data routing. Reliable data delivery can be obtained by reducing packet drops in data communication. A packet could be lost for several reasons, including lousy connectivity, overflowing buffers, or a node running out of energy. Retransmission and redundancy are two standard methods used in WSNs to ensure network reliability. Network energy usage and packet delivery delays are exacerbated by retransmission-based techniques because the number of transmissions increases [5]. Since a bit lost within a packet can be retrieved using the coding scheme, the research community needs to pay more attention to employing redundancy to achieve reliability in WSNs. If packets could be repaired to recover any lost or incorrect bits, the transmission overhead created by retransmitting an entire packet would be significantly reduced. However, selection of any node as a successor node in the routing path plays a vital role in designing an energy-efficient routing protocol. The packet dropping rate may be reduced by selecting a resourceful (which has the highest residual energy, free available buffer, good link quality, and less distance) node as the successor node.

The current work aims to design energy-efficient and reliable data-gathering algorithms in WSNs while adhering to strict data delivery time constraints by considering parameters such as residual energy, buffer capacity, link quality, and distance between nodes. In recent years, there has been considerable research to find the solution to reliable communication between sensor nodes while managing their energy consumption [6]. The primary keys in this area are addressed by familiar optimization techniques inspired by swarm intelligence, fuzzy logic system, heuristic search algorithms, reinforced learning approach, and genetic algorithms (GA).

Genetic Algorithm (GA) is an evolutionary heuristic, stochastic optimization algorithm that learns about its universe by analyzing data to eliminate wrong solutions and increase the number of excellent ones. It discovers near-optimal solutions quickly and heuristically for either massive or substantially smaller populations [7]. The literature states that GA provided optimized solutions for node placement, network coverage, clustering, data aggregation, routing, etc, in WSNs. Hence, GA is considered in the present research for developing a heuristic approach-based efficient and reliable routing mechanism.

The proposed research's novelty lies in developing a heuristic approach-based energy-efficient and reliable routing mechanism using a genetic algorithm that ensures the reduction of packet drops by selecting a resourceful node as a successor node in the route. The authors proposed a genetic algorithm with a roulette wheel selection process, value encoding, and fitness function evaluation using the quality of the link. The network

parameters considered to state the resourcefulness of a node in the evaluation of a proposed algorithm are residual energy, an available buffer of the node, link quality, and distance. The above approach demonstrated a decrease in retransmissions, packet delivery delays, and energy consumption. Further, an increase in the network lifetime.

The rest of this document is structured as follows. Section 2 discusses related work and motivation. The proposed mechanism is described in Section 3. The chromosome representation is presented in Section 3.2.1, initialization of the population is shown in Section 3.2.2, the fitness function is defined in Section 3.2.3, Section 3.2.4 offers the selection of chromosomes, the crossover and mutations are presented in Section 3.2.5 and 3.2.6 respectively and repairing of the chromosomes are explained in Section 3.2.7. Accordingly, performance was measured using the simulations in Section 4. Finally, in Section 5, we summarise our findings.

2. RELATED WORK AND MOTIVATION

This part of the manuscript provides an overview of existing works on network lifetime improvement and energy-efficient and reliable routing algorithms.

Celimuge Wu et al. [5] proposed a redundancy-based technique for reliable and rapid data collection in sensor networks to achieve high reliability and minimal end-to-end delay. The proposed protocol employs a network-based coding strategy to increase packet redundancy when a link is defective or when there is a strict end-to-end latency requirement. The protocol automatically modifies the redundancy level to suit the application requirements and the link failure rate. Fatma H. Elfouly et al. [6] used SWARM knowledge to suggest a routing scheme that grows the life of WSNs while reducing energy consumption per node. In addition, this model accounts for data reliability by guaranteeing that the sensed data will arrive at the sink node consistently. Finally, the buffer size reduces packet loss and energy expenditure from retransmitting duplicate packets. Mahmood et al. [8] addressed several reliability schemes that use retransmission and redundancy techniques to recover lost data through either hop-by-hop or end-to-end procedures. They analyzed these schemes by looking into the optimal mix of these techniques, methods, and desired reliability levels to propose an efficient mechanism for resource-constrained WSNs. Bhardwaj et al. [9] discussed the factors affecting the lifetime of WSNs. They determined that one of the most critical issues in wireless sensor networks is how long a network can survive if each node is limited in energy consumption.

Using an evolutionary genetic algorithm, Bhatia et al. [10] suggested an approach termed GADA-LEACH, which seeks to enhance CH selection in the conventional LEACH routing protocol for WSN such that to facilitate communication between the central hub (CH) and the base station (BS) by considering relay node. According to Wang et al. [11], a multipath routing technique for wireless sensor networks uses genetic algorithms to boost the network's fault tolerance and reduce energy consumption. They used only the distance between nodes to compute the genetic algorithm's fitness function. Muruganatham et al. [12, 13] have presented simulating results analysis of classic and genetic-based routing strategies to examine the performance of a wireless sensor network. The Genetic approach has been shown to enhance the WSN's lifespan in the presence of faulty nodes.

Mujtaba Romoozi et al. [14] explored innovative strategies for node positioning to reduce power consumption without compromising coverage. They proposed a genetic algorithm-based node positioning in wireless sensor networks to optimize energy consumption and

extend the network's lifetime. T.Abirami et al. [15] used a genetic algorithm to improve the network lifetime by creating spanning trees for data aggregation, which are stable and economical with power consumption. Hasanien Ali Talib et al. [16] presented a honey-bee optimization with a genetic algorithm approach to developing a system for sharing data among individual nodes in a one-to-one network setup. Ioana Apetroaei et al. [17] applied genetic algorithms for routing protocols in WSNs. B. Baranidharan et al. [18] to improve FND, HND, and LND presented a new clustering technique called the Genetic Algorithm based Energy-efficient Clustering Hierarchy (GAECH). Their fitness function is computed by considering the most critical aspects of a cluster. Ajay Khunteta et al. [19] designed an approach using a genetic algorithm with leach protocol for cluster head selection in WSN to mitigate energy consumption. Trong-The Nguyen et al. [20] proposed an approach based on a genetic algorithm with self-configuration chromosomes in the cluster formation of a sensor network. M. K. Somesula et al. [21] established contact duration-aware cooperative cache placement using a genetic algorithm-based heuristic search technique for practical scenarios to improve the hit and acceleration ratios.

Table 1 Summary of existing work vs proposed work

Reference	Algorithms / Techniques used	Parameters used	Comparison of the proposed work with related work
Proposed Work	Genetic Algorithm – with roulette wheel selection, value encoding, and route quality as the fitness function.	Available buffer, residual energy, Link quality, and distance.	GA-based route construction with resourceful successor nodes in the path by considering available buffer, residual energy, Link quality, and distance. The proposed work significantly decreased packet drops, improved the network's lifetime, reduced energy consumption and packet delivery delay, and decreased the number of transmissions.
[5]	Network coding assisted redundancy in improving redundancy based on link quality.	Link Quality.	A network coding-based approach to improve packet redundancy when a link is unreliable, or there is a strict end-to-end delay requirement. It has considered only link quality to update redundancy levels. But Available buffer, residual energy, and distance are also important parameters to be considered to improve the network further by reducing packet dropping rate.
[11]	Genetic Algorithm where the fitness function is determined using distance.	Distance between nodes.	Multipath routing uses a genetic algorithm with only distance as a parameter. Available buffer, residual energy, and Link quality are not considered in routing decisions which plays a vital role in reducing packet drops.
[26]	Fuzzy approach and the A-star algorithm.	Residual energy, traffic load, hop count.	A-star routing approach based on fuzzy logic to evaluate node weight using residual energy, traffic load, and hop count. Here the node's available buffer is not considered, which may lead to packet dropping.

Park et al. [22] introduced a scalable architecture to achieve reliability in downstream data delivery efficiently by considering the unique characteristics of wireless sensor networks to ensure the reliability of data transfer from sensing devices to base nodes. Le et al. [23], with the help of a statistical reliability metric, suggested an energy-efficient and reliable transport protocol (ERTP) to minimize the number of retransmissions in WSNs. ERTP guarantees that more than enough data packets are sent to the sink.

D. Jiang et al. [24] described a multi-constraint routing strategy for smart city applications using load balancing to achieve significant energy efficiency. Lee et al. [25] investigated cluster-based wireless sensor networks' upper bound on network lifespan by addressing a solution to the energy hole problem with spatial correlation in networked clusters. Alshawi et al. [26] suggested a new routing strategy for WSNs that combines the fuzzy approach and the A-star algorithm to increase the network's lifetime. The concept seeks to find the best route from the source to the destination, prioritizing routes with the most available energy, the fewest possible hops, and the low traffic loads.

The redundancy-based technique for reliable and rapid data collecting in sensor networks employed a network-coding-based strategy to increase packet redundancy when a link is defective or when there is a strict end-to-end latency requirement. The GA-based technique used only distance as a parameter to choose which routing paths to explore. But when selecting the most efficient routes for data transmission, it is crucial to consider the network link quality, remaining energy, and available buffer, as these are the primary factors that cause packet loss. The network's lifetime increases when packet loss decreases due to fewer transmissions.

In this work, a data routing strategy with parameters remaining energy, the available node buffer, link quality, and distance based on a genetic algorithm that is both energy-efficient and reliable has been proposed. This study's significance resides in using a genetic algorithm with a roulette wheel selection process, value encoding, and fitness function evaluation using the quality of the link. The resulting route ensures a selection of a resourceful node as a successor node in the routing path to minimize the number of end-to-end data transmissions, energy consumption, packet delivery delay, and an improved network lifetime.

2.1. Motivation

Obtaining accurate data collection (packet level reliability) with carefully enforced end-to-end delay requirements is a substantial difficulty in several time-critical applications of wireless sensor networks, such as tsunami warnings, chemical assault detection, forest fire detection, and infiltration detection in military surveillance. Hence apart from improving the lifetime of the WSN, reliable data should be delivered within the time boundaries. Retransmission and redundancy are two standard methods used in WSNs to ensure network reliability. Even though the retransmission approach achieves reliable data transfer as it consumes more energy, it is unsuitable for WSNs. The research community has placed less focus on using redundancy to achieve reliability in WSNs since, in redundancy-based reliability mechanisms, a bit lost within a packet can be recovered by adopting some form of the coding scheme. The transmission overhead generated by retransmitting a whole packet would be drastically reduced if packets could be repaired to restore any lost or malformed bits. So that, to meet the challenges here, we need to build a route such that the nodes in the route should be strong enough to mitigate packet drops. The primary causes of packet drops are bad connectivity, overflowing buffers, or a node running out of energy.

The Genetic Algorithm (GA) is well-known in stochastic optimization because it employs the idea of natural evolution to develop optimization solutions. Thus, a genetic algorithm is suitable for representing and resolving many complex issues. It is even proved that GA is used in wireless sensor networks to improve the positioning of nodes, the extent to which a network is covered, the organization of clusters, and the collection and aggregation of data. Hence, in this paper, an energy-efficient and reliable data routing solution using a genetic algorithm has been proposed to tackle the stated challenges in WSNs. The proposed technique improves the network's lifetime and minimizes the number of data transmissions while achieving reliable data delivery and strict packet delivery delay. The proposed mechanism also considered the available buffer, residual energy, and link quality while selecting the data routing paths. The primary objective of this study is (a) To design a WSN data routing method based on a genetic algorithm that is both energy-efficient and reliable and (b) Evaluation of the proposed algorithm's performance using simulation results.

3. PROPOSED MECHANISM

The proposed research work develops an energy-efficient and reliable routing mechanism using the genetic algorithm with a roulette wheel selection process and value encoding; by considering the sensor node's residual energy, available buffer, link quality, and distance. This section of the manuscript describes the proposed work and associated methodologies - Genetic Algorithm-based Routing Scheme, Chromosome representation, Initialization of population, Fitness function, Selection, Recombination (Crossover), Mutation, and Repair.

3.1. Network Model

As indicated in Fig.1, a WSN has been considered as a graph $G(V, E)$, where V is a set of vertices of the graph representing a set of sensor nodes of the WSN, and E is the set of edges of the graph illustrating a set of wireless communication links of WSN. Each node is shown in Fig.1. with associated residual energy 'e' and buffer availability 'b.' Using multi-hop communication, a sensor node gets the data and directs it to the sink node. Retransmissions were expected to be repeated until all the packets arrived at the sink node.

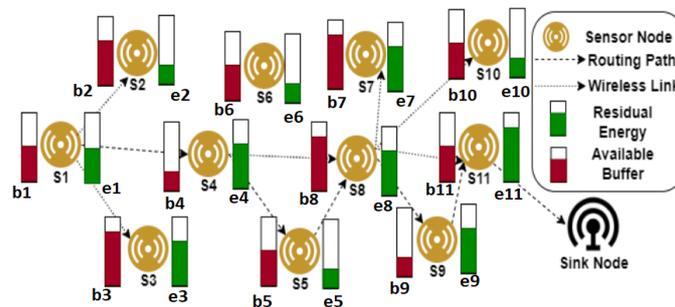


Fig. 1 Network diagram

3.2. Genetic Algorithm-based Routing Scheme

The Genetic Algorithm (GA) is a stochastic optimization tool based on natural evolution [21,32,33,35]. GA was found to be suitable for parallel optimization [29]. GA is an incremental method in which each round is referred to as a generation. GA is a population-based method that considers all possible individuals. GA derives individuals randomly from a specified population, and these individuals are encoded into genetic form. Every chromosome can be viewed as a single string or an array of genes covering a piece of the solution. Alleles are different variations of a gene's value. The evolution of encoded individuals is accomplished by continuing the processes below until the termination conditions are met.

- The fitness function identifies the fittest members with the best fitness levels, and these most qualified individuals are chosen as parents for the following generation.
- The determined parents were subjected to genetic operators (crossover and mutations) to generate new offspring from the existing ones.

The chromosomes for the following generations are chosen after a population's crossover and mutation mechanisms. Some of this generation's best performers might be changed out for the worst performers from prior generations in the same proportion to ensure that the current generation is, at maximum, as fit as the preceding generation. This is referred to as elitism. This process is continued till the algorithm's halting requirement is fulfilled. We present a genetic method for an energy-efficient routing algorithm in WSNs based on this survival of the fittest concept in Algorithm 1.

Algorithm-1: Genetic Algorithm for Routing

Input: 'P(N)': Size of Population

'CP': Crossover-Probability

'MP': Mutation-probability

'G': Number of iterations

Output: Routing path from sensor node to sink node

```

1: for all Nodes  $n \in N$ , do
2:    $t=0$ ;
3:   Generate the initial population by randomly initializing P(N);
4:   Repair the randomly generated population P(N);
5:   Evaluating the individual's fitness from the population P(N);
6:   Store best solutions of P(N) in old B(M);
7:   while  $t < G$  do
8:     Choose the individuals as parents chromosomes from P(N) (i.e., Selection);
9:     Perform the crossover on the selected individuals to produce new offspring
      (i.e., Recombination);
10:    Perform mutation on the new offspring based on CP;
11:    Repair the individual chromosomes produced after mutation;
12:    Evaluating the individual's fitness from the new population;
13:    Store the best Fitness individuals of P(N) in new B(M);
14:    if  $FIT(\text{old } B(M)) > FIT(\text{new } B(M))$  then
15:       $\text{new } B(M) = \text{old } B(M)$ ;
16:    end if
17:     $\text{old } B(M) = \text{new } B(M)$ ;
18:    compute worst fitness value in P(N) and change it with new B(M);
19:     $t=t+1$ ;
20:  end while
21: end for

```

GA aims to discover a strategy for gathering data in WSN that maximizes the network's lifetime, given a set of intermediate nodes and a base station with their coordinates. Every data collection period is referred to as a round, and the number of repetitions defines how long the first intermediate node will operate before running out of energy. We further consider that each relay node has the same initial energy.

3.2.1. Chromosome Representation

In the initial population, every chromosome belongs to a feasible genetic strategy. A sequence of positive integers indicating the IDs of sensor nodes in a route from the source to the destination is called a chromosome [28]. The structure of the initial routing path depends on the position of the central nodes, which can move the content from the source to the destination (Fig.2). The first locus gene is usually given the location of the source. The chromosome size is sensitive, but it must stay within the allowable level, the number of nodes in the network because the channel will never have more than the value. Based on virtual network information (route table), the chromosome (path) represents a problem by assigning node IDs from its source location to the destination.

The first locus's gene encodes the source node, and the nodes linked with the source node designated by the front gene's allele are picked randomly or heuristically by the succeeding locus's gene. A selected node is removed from the topological information database to prevent it from being selected again. This procedure will be repeated until the destination node has been reached. It is worth noting that encoding is only possible if each route step is carried out over an actual network link. To represent a chromosome, we used the ids of nodes in the routing sequence from source to destination. The same is encoded, as shown in Fig.2. There are different encoding schemes from the literature [7], such as Binary encoding, Octal encoding, Hexadecimal encoding, Permutation encoding, Value encoding, and Tree encoding. In the proposed work, the chromosome is a sequence of node IDs of nodes in the route from the source node to the destination node. Hence value encoding is considered to avoid complexities and further conversions.



Fig. 2 Encoding Scheme

Consider an example of a chromosome from source to destination. The chromosome is a set of nodes along the formed path (S1 – S4 – S8 – S11 – Sink), as shown in Fig. 1. The chromosome length is denoted as the count of nodes (genes) in the constructed route.

3.2.2. Initialization of Population

Population initialization of genetic algorithm assumes two points: the process to initialize the population and its size. To develop sensible solutions, it was believed that the size of the population needs to grow significantly with the complexity of the problem. However, recent research has demonstrated that satisfactory results can be produced with substantially smaller populations. To conclude, having a significant population is beneficial,

but it comes with a high cost in terms of memory and time [35, 36]. As one might think, determining an appropriate population size is critical for effectiveness.

Furthermore, the initial population can be generated in random or heuristic initialization. Due to the lacking of variety in the population, the optimal global solution is never achieved, and it examines only a tiny portion of the solution space. Hence, random initialization is used in this study to construct the initial population using the encoding approach described in Section 3.2.1.

3.2.3. Fitness Function

The fitness function analyses chromosomes regarding the physical description and assesses their viability in the solution based on desirable features. However, the fitness function should precisely estimate the population's chromosomal quality. Hence, the fitness function is crucial. In our work, the fitness function is the number of data transmissions (including ACK and retransmissions) needed to transmit the packet to the sink node and is defined as follows.

$$N = \left(1 + \sum_{x=1}^{nr} G_x\right) + \left(1 + \sum_{x=1}^{nr} G_x\right) * P \quad (1)$$

where $G_x = (1 - P^H)^x$

N - signifies the total number of transfers (including ACK and data transfer),

' nr ' represents the largest number of retransmissions (' nr ' = 8),

' P ' represents the probability of successful link transfer,

' H ' represents the number of hops between the source and destination,

' G_x ' represents the probability that a node will not get an ACK for an ' x^{th} ' data transmission.

In the transfer method, the first part of Equation 1 specifies the data transfer value, and the next part specifies the ACK transfer value per packet of data.

3.2.4. Selection

The reproduction operation aims to increase the population's average quality by increasing the possibility of high-quality chromosomes being transferred to the subsequent generation [35,38]. As a result of the selection, the exploration is focused on profitable regions in the solution space.

Algorithm-2: Selection Process

- 1: Compute the selection probability of each individual using Eq. (2);
 - 2: **for** all Generations $g \in G$, **do**
 - 3: $C_i=0$;/*chromosome index*/
 - 4: $P(r)=0$;/*accumulation probability of roulette*/
 - 5: **while** $P(r) < \text{random}(0,1)$ **do**
 - 6: $C_i=C_i+1$;
 - 7: $P(r)=P(r)+P_s(C_i)$;
 - 8: **end while**
 - 9: Selected individual= C_i ;
 - 10: **end for**
-

The selection systems are characterized by selection pressure, the ratio of the probability of selecting the best chromosome in a population to the probability of selecting an average chromosome. As a result of the tremendous selection pressure, the population reaches stability swiftly, although genetic diversity is inevitably sacrificed. Various approaches for selection are available in the literature [45]: Proportionate selection (Roulette Wheel selection), Stochastic Selection, Tournament selection, and Truncation selection. The Roulette Wheel selection approach is the optimal choice when deciding on a selection method for genetic algorithm. Because of its simplicity of understanding and coding with correctness at runtime, there is a strong bias toward selecting the fittest elements. Hence the proposed work uses roulette wheel approach to choose the elite individuals. Here individuals are picked depending on the selection probability of the fitness function. The selection probability of an individual is well-defined as:

$$P_s(j) = \frac{FIT(j)}{\sum_{j \in N_{pop}} FIT(j)} \tag{2}$$

where ‘*FIT(j)*’ represents the fitness value of *j*.

3.2.5. Recombination (Crossover)

Blending the existing chromosome’s genetic information as parents to produce the new chromosomes (children) is known as crossover (recombination). Fig.3 shows the crossover process. In the routing problem, recombination is defined as swapping each partial route of two individuals so that the child produced by the crossover reflects only one route.

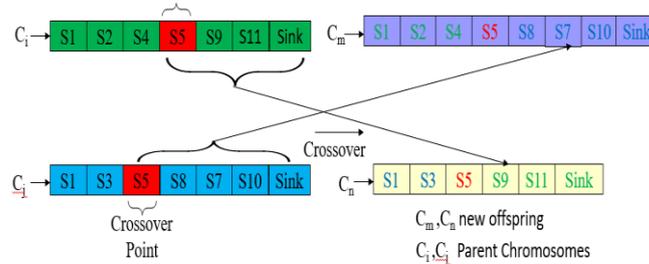


Fig. 3 Crossover example

As a result, the one-point crossover is a suitable method for the proposed GA. The source node is connected to a relay node by one partial route, and the relay node is connected to the destination node by the other partial route. The recombination of two prevalent parents picked through the selection process (Algorithm 3) increases the likelihood of generating children with prevailing characteristics.

The proposed manuscript considered a one-point crossover different from the traditional one-point recombination scheme in the proposed GA. Since the proposed problem deals with routing, the solution should contain the sequence of nodes that form a path from a source to the target node. Hence, the recombination operation must have a minimum of one common node (gene) in the chromosomes other than the source and destination nodes in the considered chromosomes. However, the position of genes in the chromosome needs to be different. Algorithm 3 shows the steps to perform the crossover operation.

Algorithm-3: Crossover Process

Input: Each chromosome contains variable Length d genes E.g.,
 $C_i = \{g_1, g_2, \dots, g_d\}$
 d : Length of the chromosome

Output: C_m, C_n : New chromosomes after crossover

- 1: **for** all Chromosomes **do**
- 2: pick two chromosomes (C_i, C_j) from the given population;
- 3: CPS=Find the set of crossing point pairs by calling Common(C_i, C_j)
- 4: **if** Length(CPS) ≥ 1 **then**
- 5: **if** CP $>$ random[0,1] **then**
- 6: ex=Choose randomly a pair from CPS;
- 7: Produce the two new routes by switching the partial routes
of C_i and C_j with other
(i.e., $C_m = (C_i[1] \text{ to } C_i[\text{ex}[1]]) + (C_j[\text{ex}[2]+1] \text{ to } C_j[d])$
and $C_n = (C_j[1] \text{ to } C_j[\text{ex}[2]]) + (C_i[\text{ex}[1]+1] \text{ to } C_i[d])$);
- 8: **end if**
- 9: **end if**
- 10: **end for**

3.2.6. Mutation

The mutation operation involves either flipping the randomly chosen gene based on mutation probability or explicit modification, and the mutation causes a slight bias. Fig. 4 shows mutation. The mutation helps to achieve the global optima by avoiding the optimal local value. Flipping a gene may produce a partial and incomplete path in the suggested scheme. Hence, when a random node is picked, we identify the common nodes connected to the next and before the chosen node. The chosen node is replaced with one of the common nodes identified. Algorithm 4 specifies the steps involved in the mutation process.

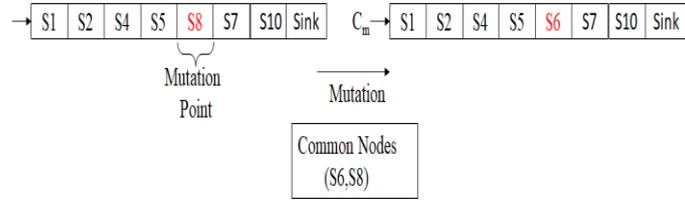


Fig. 4 Mutation example

Algorithm-4: Mutation Process

- 1: **if** MP $>$ random[0,1] **then**
- 2: Randomly choose a gene $C_i[m]$ (i.e., node in the routing path);
- 3: Identify the previous and next genes to $C_i[m]$ (i.e., $C_i[m-1]$ and $C_i[m+1]$) in the chromosome;
- 4: Choose a node randomly from the list of nodes that are common to $C_i[m-1]$ and $C_i[m+1]$);
- 5: **end if**

3.2.7. Repairing

The recombination could produce infeasible individuals, which contradicts the proposed buffer and residual energy availability constraints. Even loops may be generated while performing crossover operations. Each chromosome produced after crossover and mutation should be reasonable. Hence, repairing the violations should be done to make the infeasible chromosomes feasible. First, the nodes involved in the looping path should be removed to eliminate the loops in the routing path. Second, each node should have buffer availability (AB) and residual energy (RE) more significant than the given threshold; otherwise, considering those nodes would result in a losing path or dropping packets. Therefore, each gene should be evaluated in the repairing process to determine whether 'AB' and 'RE' exist sufficiently. If any gene is not meeting the criteria, then replace the gene (node) with one of the common nodes common to the previous and next nodes to a node in position 'G' having available buffer greater than τ_{ab} and residual energy greater than τ_{re} ; The process of repairing is present in Algorithm 5.

Algorithm-5: Repairing Process

```

1: for all Chromosomes, do
2:   if Chromosome is infeasible (i.e., loops in routing path),
   then
3:     Identify nodes that form a cycle;
4:     Remove the cycle
5:   end if
6:   for  $g \in C_i$  do
7:     if ( $g_{ab} < \tau_{ab}$ ) or ( $g_{re} < \tau_{re}$ ), then
8:        $S_{comm}^{ab}$  = Nodes that are common to the
       previous and next nodes to a node in position g
       having available buffer greater than  $\tau_{ab}$  and
       residual energy greater than  $\tau_{re}$ ;
9:       Replace g with a node randomly chosen from
        $S_{comm}^{ab}$ ;
10:    end if
11:  end for
12: end for

```

4. EVALUATION OF PERFORMANCE

At this stage, the proposed data collection method has been simulated in Network Simulator-3 (NS3) [46]; it is compared with the retransmission-based strategy and multipath routing using GA in terms of power consumption, transmission rate, packet delivery delay, and network lifetime. To ensure the most reliable data transfer, each packet is repeatedly transferred in the simulation until it reaches the sink node. In this paper, the data collection round refers to each node that senses the packet and sends it to the sink node.

4.1. Simulation Setup

In the simulation, the nodes in the network are between 50 and 500, the initial energy of a node is 25 KJ, and the buffer capacity is 2.5 KB. The link quality is assigned a value between 0 and 1. Table 2 describes the simulation parameters.

Table 2 Parameters for the simulation

Parameter	Value
Number Nodes	50 to 500 nodes
Range Transmission	40 meters
Node's initial energy	25 KJ
Packet Size	960 bits
$E_s = \alpha_3$	$\alpha_3 = 50 \times 10^{-9}$ Joules / bit
$E_r = \alpha_{12}$	$\alpha_{12} = 0.787 \times 10^{-6}$ Joules / bit
$E_t = \alpha_{11} + \alpha_2 d^n$	$\alpha_{11} = 0.937 \times 10^{-6}$ Joules / bit
	$\alpha_2 = 10 \times 10^{-12}$ Joules / bit / meters ²
	$d = 85$ meters

4.2. Results and Discussions

To analyze the network, the metrics considered are energy consumption, average delivery delay of a packet, the number of retransmissions, and network lifetime in terms of the first node dying round and half nodes dying round.

The packet delivery delay, also known as latency, refers to the time a data packet travels from one node to another. The average packet delivery delay for the three approaches is shown in Fig. 5. It is observed from Fig. 5 that, compared to the other two techniques, the retransmission-based strategy has the most latency due to more retransmissions. The distance-based GA algorithm considered the distance as the parameter for finding the route. But our proposed work also considered link quality, residual energy, and available buffer parameters while finding the data routing paths.

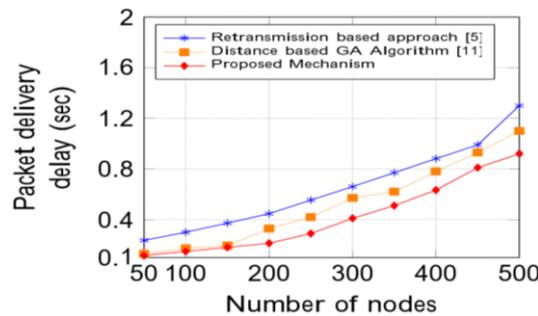


Fig. 5 Performance comparison between the average packet delivery delay

Fig. 6 compares the number of transmissions among the distance-based GA algorithm, retransmission-based approach, and the proposed energy-efficient and reliable GA-based data gathering mechanism. The number of retransmissions in the network increases in proportion to the packet loss rate. Each dropped packet is retransmitted in the simulation to achieve high data delivery reliability. As seen in Fig. 6, our approach minimizes the number of retransmissions compared to the other two mechanisms. Our mechanism identifies resourceful data routing paths to reduce total data transmissions. The fitness function of the proposed GA-based approach considers the link quality while choosing the routing paths. Hence, the proposed algorithm selects good link-quality routing paths; thereby, it minimizes the rate of

packet dropping and reduces the number of retransmissions. The proposed mechanism also considered the residual energy and available buffer parameters in the genetic algorithm.

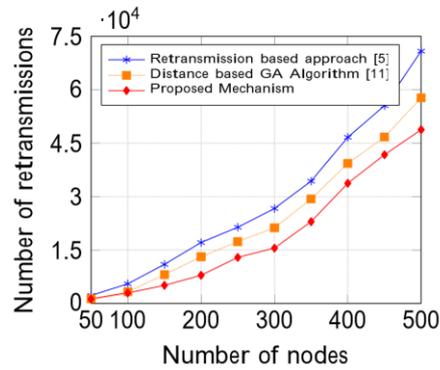


Fig. 6 Performance comparison between the number of data and ACK retransmissions

Energy consumption is the energy needed to send data from one node to another (transmission energy). Fig.7 shows the network's energy consumption for the proposed technique, retransmission method, and distance-based GA mechanism. The number of retransmissions is higher in the retransmission-based approach than in our proposed mechanism. So that the network's consumption of energy is more in a retransmission-based approach. A distance-based GA mechanism considered only distance as a parameter while constructing a route. Hence in contrast to the retransmission method and distance-based GA mechanism, the proposed mechanism used less energy. The GA-based proposed mechanism considered the link quality as the parameter that reduces the number of transmissions, resulting in reduced energy consumption.

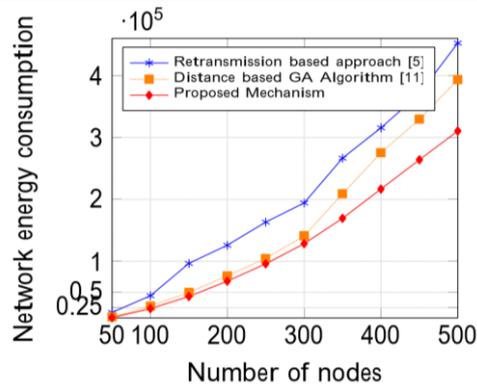


Fig. 7 Performance comparison between Network energy consumption

Network lifetime has different definitions by various researchers. Here we considered that the lifetime of a network is measured in terms of how long it takes for some predetermined fraction of sensors to die from lack of energy, such as the first node dying round and half

nodes dying round. Fig. 8 and Fig. 9 show the total life of the network (first dead node) and half of the dead nodes. A half-node dead round is a data collection round in which half of the total number of nodes dies. The proposed approach has improved network lifetime compared to the retransmission method and the distance based on the GA algorithm.

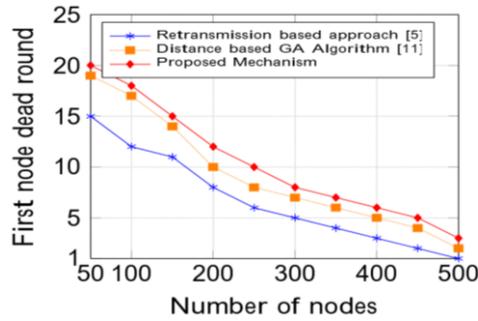


Fig. 8 Performance comparison between First node dead round

The proposed technique provides the optimal data routing path in terms of available energy, available buffer, and quality of the link. Thus, the packet loss rate is minimized. Fig. 8 and Fig. 9 depict that - with an improvement in the node count, the node's lifetime decreases due to increased network energy consumption. In comparison to existing techniques, the proposed mechanism exhibited considerable longevity improvement.

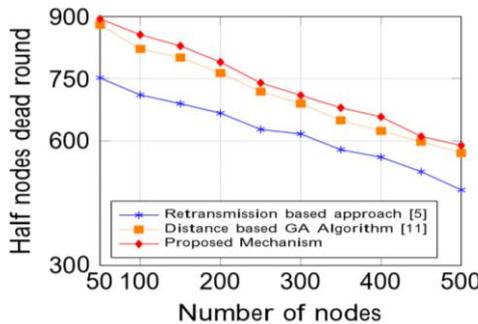


Fig. 9 Performance comparison of Half nodes dead round

Table 3 signifies that the proposed method is an improvement over the retransmission-based strategy, the distance-based genetic algorithm, and the A-star-based approach. This is because the proposed work developed an energy-efficient and reliable data delivery routing using a genetic algorithm with a roulette wheel selection approach and value encoding for encoding chromosomes, considering the parameters like node's current energy state, distance, link quality, and available buffer. It is demonstrated when compared with the related works from Table 3, the delay in packet delivery is reduced, the network's lifetime is extended, and the energy consumption is reduced.

Table 3 Comparison of results with existing works

Reference	Number of Nodes	Average packet delivery delay (seconds)	First node died round	Half number of Nodes died round
[5]	50	0.24	15	750
[11]		0.12	19	880
[26]		0.13	16	860
Proposed Work		0.11	20	890
[5]	200	0.49	8	670
[11]		0.34	10	760
[26]		0.22	11	755
Proposed Work		0.21	12	780
[5]	300	0.67	5	620
[11]		0.58	7.5	690
[26]		0.44	7	700
Proposed Work		0.41	8	710
[5]	400	0.89	3	560
[11]		0.79	5	620
[26]		0.68	5	634
Proposed Work		0.63	6	650

5. CONCLUSION

There are two primary challenges in WSN: power efficiency and the availability of reliable data transmissions. In this study, a genetic algorithm-based data collection strategy is developed to improve the ability to use robust and reliable data systems in WSNs to extend network life. The genetic algorithm with a roulette wheel selection approach and value encoding regulates the most energy-efficient and reliable routes by considering the node's remaining energy, the link quality, the node's free available buffer, and distance as key factors that cause packet loss. Simulated results show that the proposed method reduces network energy consumption by 40 percent. In addition, the proposed technique significantly increases the network's lifetime and reduces the packet delivery delay. Possible future expansion to the study includes considering node mobility and energy harvesting.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless sensor networks: a survey", *Comput. Netw.*, vol. 38, no. 2, pp. 393-422, March 2002.
- [2] I. F. Akyildiz and I. H. Kasimoglu, "Wireless sensor and actor networks: research challenges", *Ad Hoc Netw.*, vol. 2, no. 4, pp. 351-367, Oct. 2004.
- [3] T. Rault, A. Bouabdallah and Y. Challal, "Energy efficiency in wireless sensor networks: A top-down survey", *Comput. Netw.*, vol. 67, pp. 104-122, April 2014.
- [4] B. Singh and D. K. Lobiyal, "An energy-efficient adaptive clustering algorithm with load balancing for wireless sensor network", *Int. J. Sensor Networks*, vol. 12, no. 1, pp. 37-52, July 2012.
- [5] C. Wu, Y. Ji, J. Xu, S. Ohzahata and T. Kato, "Coded packets over lossy links: A redundancy-based mechanism for reliable and fast data collection in sensor networks", *Comput. Netw.*, vol. 70, pp. 179-191, Sept. 2014.

- [6] F. H. Elfouly, R. A. Ramadan, M. I. Mahmoud and M. I. Dessouky, "Swarm intelligence based reliable and energy balance routing algorithm for wireless sensor network", *FU: Elec. Energ.*, vol. 29, no. 3, pp. 339-355, Sept. 2016.
- [7] U. Mehboob, J. Qadir, S. Ali and A. Vasilakos, "Genetic algorithms in wireless networking: techniques, applications, and issues", *Soft Computing*, vol. 20, no. 6, pp. 2467-2501, June 2016.
- [8] M. A. Mahmood, W. K. G. Seah and I. Welch, "Reliability in Wireless Sensor Networks: A Survey and Challenges Ahead", *Comput. Netw.*, vol. 79, pp. 166-187, March 2015.
- [9] M. Bhardwaj, T. Garnett and A. P. Chandrakasan, "Upper Bounds on the Lifetime of Wireless Sensor Networks", In Proceedings of the IEEE International Conference on Communications (ICC), 2001, pp. 785-790.
- [10] T. Bhatia, S. Kansal, S. Goel and A. Verma, "A genetic algorithm-based distance-aware routing protocol for wireless sensor networks", *Comput. Electr. Eng.*, vol. 56, pp. 441-455, Nov. 2016.
- [11] S. Wang, "Multipath routing based on genetic algorithm in wireless sensor networks", *Hindawi Math. Prob. Eng.*, vol. 2021, pp. 1-6, June 2021.
- [12] N. Muruganantham and H. El-Ocla, "Genetic algorithm-based routing performance enhancement in wireless sensor networks", In Proceedings of the IEEE 3rd International Conference on Communication and Information Systems (ICCIS), 2018, pp. 79-82.
- [13] N. Muruganantham and H. El-Ocla, "Routing using genetic algorithm in a wireless sensor network", *Wirel. Pers. Commun.*, vol. 111, pp. 2703-2732, Jan. 2020.
- [14] M. Romozi and H. Ebrahimpour-komleh, "A Positioning Method in Wireless Sensor Networks Using Genetic Algorithms", In Proceedings of the International Conference on Medical Physics and Biomedical Engineering, 2012, pp. 1042-1049.
- [15] T. Abirami and P. Priakanth, "Energy Efficient Wireless Sensor Network using Genetic Algorithm based Association Rules", *Int. J. Comput. Appl.*, vol. 91, no. 10, April 2014.
- [16] H. A. Talib, R. Alotman and M. K. Farhan, "Optimization Approach to Optimal Power Efficient Based On Cluster Top Option In Wireless Sensor Networks", *Turkish J. Comput. Math. Education*, vol. 12, no. 4, pp. 970-979, April 2021.
- [17] I. Apetroaei, I.-A. Oprea, B.-E. Proca and L. Gheorghe, "Genetic algorithms applied in routing protocols for wireless sensor networks", In Proceedings of the 10th RoEduNet International Conference, 2011, pp. 1-6.
- [18] B. Baranidharan and B. Santhi, "GAECH: Genetic Algorithm Based Energy Efficient Clustering Hierarchy in Wireless Sensor Networks", *Hindawi J. Sensors*, vol. 2015, pp. 1-8, Aug. 2015.
- [19] A. Khunteta and A. Bajpai, "Genetic algorithm with leach protocol for cluster head selection in wireless sensor networks", *ICTACT J. Commun. Technol.*, vol. 11, no. 2, pp. 2182-2186, June 2020.
- [20] T.-T. Nguyen, C.-S. Shieh, M.-F. Horng and T.-K. Dao, "A Genetic Algorithm with Self-Configuration Chromosome for the Optimization of Wireless Sensor Networks", In Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia, 2014, pp. 413-418.
- [21] M. K. Somesula, R. R. Rout and D. Somayajulu, "Contact duration-aware cooperative cache placement using genetic algorithm for mobile edge networks", *Comput. Netw.*, vol. 193, April 2021.
- [22] S. J. Park, R. Vedantham, R. Sivakumar and I. F. Akyildiz, "GARUDA: Achieving Effective Reliability for Downstream Communication in Wireless Sensor Networks", *IEEE Trans. Mobile Comput.*, vol. 7, no. 2, pp. 214-230, Feb. 2008.
- [23] T. Le, W. Hu, P. Corke and S. Jha, "ERTP: Energy efficient and Reliable Transport Protocol for data streaming in Wireless Sensor Networks", *Comput. Commun.*, vol. 32, pp. 1154-1171, Jan. 2009.
- [24] D. Jiang, P. Zhang, Z. Lv and H. Song, "Energy-efficient multi-constraint routing algorithm with load balancing for smart city applications", *IEEE Internet of Things J.*, vol. 3, no. 6, pp. 1437-1447, Sept. 2016.
- [25] S. Lee and H. S. Lee, "Analysis of Network Lifetime in Cluster-Based Sensor Networks", *IEEE Commun. Letters*, vol. 14, no. 10, pp. 900-902, Oct. 2010.
- [26] I. S. Alshawi, L. Yan, W. Pan, B. Luo, "Lifetime enhancement in wireless sensor networks using fuzzy approach and a-star algorithm". *IEEE Sensors J.*, vol. 12, no. 10, pp. 3010-3018, Oct. 2012.
- [27] S. K. A. Imon, A. Khan, M. D. Francesco and S. K. Das, "Energy-efficient randomized switching for maximizing lifetime in tree-based wireless sensor networks", *IEEE/ACM Trans. Netw.*, vol. 23, no. 5, pp. 1401-1415, Oct. 2015.
- [28] C. W. Ahn and R. S. Ramakrishna, "A Genetic Algorithm for Shortest Path Routing Problem and the Sizing of Populations", *IEEE Trans. Evolutionary Comput.*, vol. 6, no. 6, pp. 566-579, Dec. 2002.
- [29] P. Lin, Q. Song and A. Jamalipour, "Multidimensional cooperative caching in comp-integrated ultra-dense cellular networks", *IEEE Trans. Wirel. Commun.*, vol. 19, no. 3, pp. 1977-1989, Dec. 2019.
- [30] W. Shen, T. Zhang, F. Barac and M. Gidlund, "Priority-MAC: A Priority-Enhanced MAC Protocol for Critical Traffic in Industrial Wireless Sensor and Actuator Networks", *IEEE Trans. Industr. Inform.*, vol. 10, no. 1, pp. 824-835, Feb. 2014.

- [31] N. Alsindi and K. Pahlavan, *Node localization: Wireless Sensor Networks: a networking perspective*, John Wiley & Sons, 2009, chapter 8.
- [32] J. Patel and H. El-Ocla, "Energy Efficient Routing Protocol in Sensor Networks Using Genetic algorithm", *MDPI Sensors*, vol. 21, no. 21, p. 7060, Oct. 2021.
- [33] M. Shokouhifar and A. Hassanzadeh, "An Energy Efficient Routing Protocol in Wireless Sensor Networks using Genetic Algorithm", *Adv. Environ. Biol.*, vol. 8, no. 21, pp. 86-93, Oct. 2014.
- [34] Y. Liu, A. Liu, Y. Li, Z. Li, Y. June Choi, H. Sekiya and J. Li, "APMD: A fast data transmission protocol with reliability guarantee for pervasive sensing data communication", *Pervasive Mob. Comput.*, vol. 41, pp. 413-435, 2017.
- [35] K. Sastry, D. Goldberg and G. Kendall, *Genetic algorithms in Search methodologies*, Springer, 2005, chapter-4, pp. 97-125.
- [36] U. Dohare, D. K. Lobiyal and S. Kumar, "Energy Balanced Model for Lifetime Maximization in Randomly Distributed Wireless sensor networks", *Wirel. Pers. Commun.*, vol. 78, no. 1, pp. 407-428, April 2014.
- [37] B. Singh and D. K. Lobiyal, "An energy-efficient adaptive clustering algorithm with load balancing for wireless sensor network", *Int. J. Sensor Networks*, vol. 12, no. 1, pp. 37-52, July 2012.
- [38] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley Publishing, October 1989.
- [39] S. Gudla and N. R. Kuda, "Learning automata-based energy efficient and reliable data delivery routing mechanism in wireless sensor networks", *J. King Saud Univ. – Comput. Inform. Sci.*, vol. 34, no. 8, pp. 5759-5765, April 2021.
- [40] A. Rastogi and S. Rai, "A novel protocol for the stable period and lifetime enhancement in WSN", *Int. J. Inform. Technol.*, vol. 13, pp. 777-783, Jan. 2021.
- [41] D. K. Sharma, D. Kukreja, S. Bagga et al., "Gauss-sigmoid based clustering routing protocol for wireless sensor networks", *Int. J. Inform. Technol.*, vol. 13, pp. 2569-2577, Nov. 2019.
- [42] F. Ullah, M. Zahid Khan, M. Faisal, H. U. Rehman, S. Abbas and F. S. Mubarek, "An energy-efficient and reliable routing scheme to enhance the stability period in wireless body area networks", *Comput. Commun.*, vol. 165, no. 1, pp. 20-32, Jan. 2021.
- [43] D. Deepakraj and K. Raja, "Markov-chain based optimization algorithm for efficient routing in wireless sensor networks", *Int. J. Inform. Technol.*, vol. 13, pp. 897-904, March 2021.
- [44] J. Agarkhed, V. Kadrolli and S. Patil, "Fuzzy based multi-level multi-constraint multi-path reliable routing in a wireless sensor network", *Int. J. Inform. Technol.*, vol. 12, pp. 1133-1146, June 2020.
- [45] R. Champlin, "Selection Methods of Genetic Algorithms", *Student Scholarship - Computer Science*, 2018. Available at: https://digitalcommons.olivet.edu/csis_stsc/8 (Accessed: 2022-01-02).
- [46] Network Simulator 3. Available at: <https://www.nsnam.org> (Accessed:2022-01-02).