A Genetic Algorithm-Based Clustering and Watchdog Selection Framework for Heterogeneous Wireless Sensor Networks

Sarah Abdulelah Abbas[1], Leili Farzinvash[1*], Mina Zolfy[1]

1-Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran.

Abstract: In Wireless Sensor Networks (WSNs), relying solely on Cluster Heads (CHs) for management exposes the system to vulnerabilities from compromised or malfunctioning nodes. Introducing WatchDog (WD) nodes provides critical independent monitoring. They enhance security and ensure data integrity. Previous studies in this area have largely overlooked Heterogenous WSNs (HWSNs), where nodes differ in roles and resources. Furthermore, existing methods on homogenous WSNs fall short of achieving optimal clustering and monitoring efficiency. This paper presents a two-phased novel clustering, routing, and WD selection algorithm designed for HWSNs. In the first phase, normal nodes are assigned to super nodes acting as CHs which facilitates structured data aggregation and intra-cluster management. Additionally, in the same phase, predefined number of super nodes are tasked to act as WDs, charged with independent monitoring tasks such as anomaly detection, data integrity assurance, and verification of task execution. Notably, the roles of CH and WD are mutually exclusive to avoid functional overlap and optimize energy resources usage. In the second phase, a spanning tree is constructed over non-WD super nodes. The proposed method employs Genetic Algorithm (GA) for network configuration. Additionally, two fitness functions that take to the account the required energy for data transmission and monitoring are devised to find the best structure for the network in each phase. Simulation results show the effectiveness of the approach in improving data delivery, network lifetime and minimizing energy consumption, by in average 25%, 59 time slices, and 10% respectively.

[*] Corresponding Author: Leili Farzinvash, Associate professor, Postal address: Advanced Computer Networks Lab, Faculty of Electrical and Computer Engineering, University of Tabriz, 29 Bahman Blvd, Tabriz, Iran, PO Box: 5166616471.

Keywords: Heterogeneous wireless sensor network; Genetic algorithm; Cluster Head; Malicious node; Watchdog.

1. Introduction

Heterogeneous Wireless Sensor Networks (HWSNs) enable pervasive monitoring in domains such as industrial control, environmental sensing, and smart infrastructure [1]. However, these networks suffer from limitations like constrained energy, limited bandwidth, and security issues. Cluster-based architectures - where super nodes acted as Cluster Heads (CHs) - are a well-established approach for reducing communication overhead and prolonging network lifetime. Many surveys and recent works continue to treat CH as a core optimization problem for both homogeneous and Heterogeneous WSNs [2,3]. Population-based optimizers such as Genetic Algorithm (GA) have been used to find near-optimal CH selection and routing structures in HWSNs because they can jointly optimize multiple metrics under diverse constraints [4,5].

At the same time, security and integrity remain open challenges for cluster-based WSNs: CHs that aggregate and forward data are attractive targets for attackers. Additionally, conventional intrusion-detection and WatchDog (WD) schemes have been designed for homogenous WSNs and have not accounted for heterogeneity [6]. Surveys of Intrusion Detection System (IDS) techniques for WSNs and studies of WD placement highlight that (i) dedicated monitoring by WD nodes significantly improve detection, and (ii) The way WDs are placed and selected has a major impact on both detection coverage and energy consumption [7, 8]. More recent works also showed that secure or intelligent WD-selection mechanisms are necessary to maintain high availability of monitoring under attack or node failure, rather than assuming that any CH or neighbor can safely act as a monitor [9, 10].

Trust management has emerged as a reasonable mechanism to enhance the resilience of cluster-based HWSNs against attacks. Trust models provide a decentralized way to evaluate the reliability of nodes by observing their past behavior and recommending trustworthiness for forwarding, aggregation, or monitoring roles. In particular, trust-assisted clustering approaches aim to prevent the election of malicious or selfish nodes as CHs, while trust-aware WD selection ensures that nodes with consistent and credible monitoring histories are prioritized. However, existing trust-based solutions often assume homogeneous capabilities or overlook the dual challenges of energy constraints and heterogeneity in HWSNs, which can lead to biased trust computation or additional communication costs [11, 12]. Integrating trust evaluation with optimization-based clustering and WD selection presents a promising yet underexplored direction for balancing security, energy efficiency, and performance guarantees in heterogeneous deployments.

Motivated by these observations, this paper proposes a lightweight, GA-driven framework that jointly optimizes clustering, routing, and WD selection for HWSNs. We enforce mutual exclusivity between CH and WD roles among high-capability super nodes. Our method encodes for CH selection and WD selection in a single chromosome in the first phase. Additionally, WD selection for monitoring each super nodes behavior and routing (i.e. constructing spanning tree over non-WD super nodes) together are encoded also in a single chromosome in the second phase. Furthermore, two fitness functions are devised that try to reduce transmission energy and monitoring effectiveness. In short, we can list our contributions as follows:

- Role-separation architecture: We introduce a practical architecture for HWSNs that partitions higher-capability nodes into two disjoint sets of CHs and WDs. In this setup, CHs are responsible for aggregation and routing, and WDs are responsible for independent monitoring and anomaly detection.
- Chromosome representation for joint optimization: We propose two compact chromosome encodings for each phase. In the first phase, the chromosome simultaneously represents CH selection and WD selection. The proposed chromosome in the second phase selects WDs to monitor the behavior of super nodes and a parent per non-WD super node to deliver the data to Base Station (BS).
- Two new devised fitness functions: We design and evaluate two fitness functions used for each phase. In the first phase, fitness function balances energy of CHs and trust in the network. Additionally, in the second phase, the fitness function reduces transmission-energy and routing cost in addition to ensuring a reliable data transmission by avoiding malicious nodes. These two fitness functions permitting explicit tradeoffs between longevity and intrusion-resilience.
- Empirical validation: Through simulation on representative HWSN scenarios, we show that the proposed GA framework achieves better combined energy/monitoring performance than

baseline clustering schemes, and that role separation improves detection coverage and resilience to CH compromise without prohibitive energy cost.

2. Related Works

This section surveys recent literature related to (i) HWSNs, (ii) metaheuristic approaches for clustering and routing in WSNs, and (iii) WD-based monitoring and security mechanisms. We focus on works that address energy-aware clustering, routing, heterogeneity, metaheuristic optimization, and WD selection or intrusion-detection in sensor networks.

2.1. HWSNs

Anusuya et al. [13] provided a comprehensive review of sensor-node deployment and optimization strategies, emphasizing coverage and energy-efficiency tradeoffs in heterogeneous deployments. Reference [14] modeled HWSNs for coverage and reliability, proposing deployment strategies that consider multiple node classes. Authors in [15] proposed a heterogeneous routing protocol aimed at balancing load across nodes with different capabilities to extend network lifetime. Divya et al. [16] investigated optimized routing and CH selection using optimization techniques in heterogeneous scenarios. Khedhiri et al. [17] examined clustering techniques aimed at extending network lifetime through heterogeneity-aware modeling and simulations. A recent survey [18] highlighted modern metaheuristics approaches in WSNs, focusing on heterogeneous node roles.

In [19], two node types were defined: super nodes with multiple radios and high transmission range and energy, and normal single-radio nodes for sensing. The algorithm included two phases - clustering and tree construction - with separate GA-based channel assignments for each node type. Building on this, reference [20] integrated transmission power control with a multi-radio, multi-channel scheme, enabling super nodes to reduce excess energy use and extend network lifetime. In [21], super nodes acted as CHs, first forming a spanning tree and then assigning normal nodes using PSO. Similarly, reference [22] applied PSO to cluster normal nodes and build a super-node data tree, considering both energy and reliability. Wang et al. [23] jointly optimized clustering and routing via two-part chromosomes, with a chaos logistic map generating strong initial populations.

2.2. Metaheuristics for Clustering, Routing, and CH Selection

Population-based and other metaheuristic methods have jointly optimized multiple, often conflicting objectives (e.g. energy, coverage, load balance). Houssein et al. [24] surveyed metaheuristic algorithms and their specific applications in WSNs, summarizing strengths/limitations of GA, PSO, ACO, and hybrid methods. Kaedi et al. [25] performed simultaneous CH selection and routing optimization via GA. Authors in [26] proposed a PSO-based energy-efficient clustering scheme that jointly decided clusters and reduced communication cost. Sahoo et al. [27] presented a GA-based optimized CH-selection method with adaptive crossover and binary encoding for energy-aware clustering. Reference [28] adapted chaotic GA for energy-efficient, load-balanced clustering and routing in WSNs. Sharada et al. [29] introduced an adaptive ant-colony clustering algorithm that determines ideal cluster counts and improves energy performance. Authors in [30] described a two-phase metaheuristic framework for cluster-based routing that reduces energy consumption and improves lifetime. A recent work [31] used metaheuristic optimization for energy-aware clustering with integrated routing to maximize lifetime.

The metaheuristic literature provides many encodings and operators for CH and routing optimization; comparatively fewer works embed an explicit, separate monitoring/WD selection as part of the same optimization problem - an opportunity that our work addresses.

2.3. WD Placement, Intrusion Detection, and Security in WSNs

Security-focused studies highlight WD strategies, IDS designs, and the cost/coverage tradeoffs of monitoring. Reference [32] introduced a visualization technique to enhance administrators' ability to identify hidden attacks. Sivagaminathan et al. [33] surveyed IDSs for WSNs and present IDS architectures that combine signature, anomaly, and hybrid detectors tailored for constrained nodes. Shirvani and Akbarifar [34] provided a recent survey of IDS methods in WSNs, cataloguing lightweight approaches suitable for resource-limited nodes. Authors in [35] optimized barrier and sensor placement for intrusion detection using ACO-based metaheuristics, showing that placement has a strong effect on detection coverage and resource cost. Reference [36] combined PSO variants and Bayesian-game analysis to model intrusion-detection and defense strategies, illustrating how optimization and game-theoretic reasoning can improve IDS robustness.

WD and IDS research in WSNs generally emphasizes achieving high monitoring coverage while maintaining low communication and energy overhead. However, many existing WD mechanisms simplify the monitoring model by assuming that the WD nodes are either the CHs themselves or randomly selected neighboring nodes, without optimizing their placement or considering energy constraints. This assumption often limits scalability and reduces overall network lifetime. Recent discussion papers consider secure, availability-aware WD selection mechanisms - ideally co-designed with clustering/routing - which motivates our mutually exclusive CH/WD optimization.

Taken together, the surveyed works show (i) mature metaheuristic methods for CH selection and routing in HWSNs, and (ii) a growing body of IDS/WD research emphasizing coverage and reliability. What remains less explored is a joint optimization that enforces role separation (dedicated WD nodes disjoint from CHs), encodes CH/WD or WD/routing decisions in a single metaheuristic chromosome, and evaluates tradeoffs between transmission energy and monitoring coverage in heterogeneous deployments - the goal that the present paper targets.

3. System Model

In this section, we describe the system model considered in this study. We first define the network model, introduce the node types, and outline the notations used throughout the paper. We then present the energy consumption model adopted in our framework. Finally, we outline network operation timeline.

3.1. Network Model

The network consists of two categories of nodes namely, super nodes and normal nodes along with a BS. The set of all super nodes and normal nodes are shown by SN and NN, respectively. Normal nodes are tasked to monitor the environment. They deliver the gathered data to their corresponding CHs. Super nodes serve as CHs, tasked with receiving data from normal sensors and forwarding it toward the BS. Additionally, a subset of super nodes is designated as WDs, denoted by WD. It is important to note that the set of CHs (denoted by CH) and the set of WDs are mutually exclusive. This means a super node cannot simultaneously act as both a CH and a WD within the same round. Additionally, note that $SN = CH \cup WD$.

Super nodes are provisioned with higher initial energy and extended transmission range compared to normal nodes. The number of super nodes is significantly less than the number of normal nodes. Since direct communication with the BS may not always be feasible, data delivery from distant nodes is enabled through multi-hop forwarding where nearby super nodes act as relays to progressively route information toward the BS.

3.2. Energy Consumption Model

Both normal and super nodes deplete energy during data communication, whether transmitting or receiving packets. The amount of consumed energy for data transmission and data reception depends on the packet size l and is modeled using the equations outlined in (1) and (2), respectively. These equations account for the energy required by the internal circuitry e_{elec} and the additional energy consumed for signal amplification during transmission. In (1), d_{tr} stands for distance between sender and receiver.

$$e_{trs}(l, d_{tr}) = \begin{cases} l\, e_{elec} + l\, \varepsilon_{fs}\, d_{tr}^{2},\ d_{tr} < d_0 \\ l\, e_{elec} + l\, \varepsilon_{mp}\, d_{tr}^{4},\ d_{tr} \geq d_0 \end{cases} \tag{1}$$

$$e_{rec}(l) = l\, e_{elec} \tag{2}$$

The energy consumption of amplifier is distance-dependent: For short transmission distances, the free-space model applies, with amplification cost denoted by ε_{fs}. For transmissions exceeding a threshold distance d_0, the multipath fading model is used, with amplification cost denoted by ε_{mp}. The threshold distance d_0 is defined as $\sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$.

3.3. Adversary and Trust Model

The attack analyzed in this paper is the Selective Forwarding Attack (SFA), where each malicious node discards a portion of the packets it receives. Specifically, a compromised sensor (either normal or super) drops a received packet with probability θ, and forwards it to the parent with probability $1 - \theta$. No restrictions are placed on the behavior of malicious nodes, and it is assumed that the attacker may compromise any sensor in the network.

The trust level is calculated at the BS using (3). In this equation, notation t_i which shows the rate of data successfully forwarded by node i compared to its received data (Equation (4)). Additionally, t_i^r represents the trust level of node i at round r. The value of t_i^r is computed as an average of the trust levels of node i up to the previous round (t_i^{r-1}), combined with the trust value evaluated at the current round (i.e., t_i). Finally, parameter α presents the importance of t_i^{r-1} in computing t_i^r.

$$t_i^r = \alpha\, t^{r-1}{}_i + (1-\alpha)\, t_i \tag{3}$$

$$t_i = \frac{Forwarded\ data\ by\ node\ i}{Received/generated\ data\ of\ node\ i} \tag{4}$$

3.4. Network Operation Timeline

The network operation begins with a bootstrapping phase, during which each sensor is assigned a unique ID and its location is determined. Through the exchange of hello packets, nodes identify their neighbors, while the BS collects location information and neighbor lists of sensors for use in subsequent operations.

As illustrated in Figure 1, the network functions in rounds, each comprising two main periods: network setup and data collection. At the start of every round, the BS executes the proposed algorithm, taking into account the current network status, such as the remaining energy, availability of both normal sensors and super nodes, and the degree for trust of each node. The network setup is adjusted based on the results produced by the algorithm. Each round is divided into time slices which are further segmented into time slots that allocate resources for intra-cluster and inter-cluster data transmission. This iterative process continues until the network becomes non-operational, either due to complete energy depletion of nodes or other critical failures.

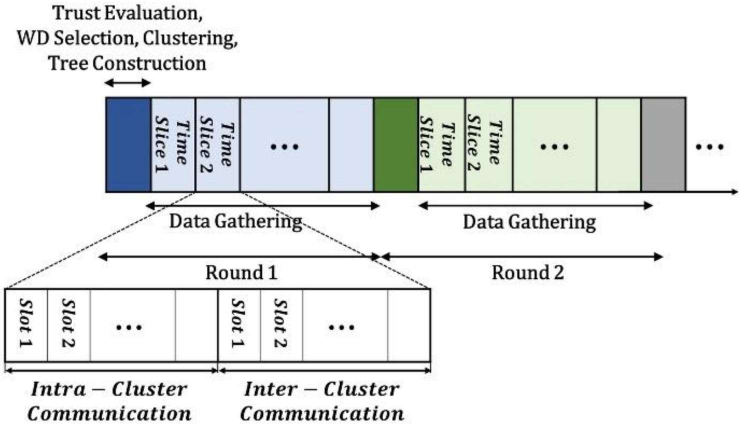


Figure 1. Network operation time.

WD nodes play a key role in maintaining the security and reliability of the network. Their main responsibility is to monitor the behavior of nearby nodes, detect any abnormal or malicious activities, and evaluate the trust level of each node based on observed actions. By continuously assessing communication patterns and packet forwarding behavior, WDs help identify misbehaving or compromised nodes, ensuring that only trustworthy nodes participate in data transmission. This trust evaluation process enhances the overall resilience and stability of the network.

4. The Proposed Method

In this section, we provide a comprehensive discussion over the proposed method. As mentioned earlier, it operates in two distinct phases:

1. Clustering and WD Selection: The network is partitioned into clusters. Additionally, predefined number of super nodes are selected to act as WDs.
2. Tree Construction and WD Assignment: A routing tree is built over the non-WD super nodes, with the BS acting as the root of the tree. Furthermore, WDs are assigned to monitor the security of the network and the correctness of the work of the super nodes.

The mentioned phases are described in the following.

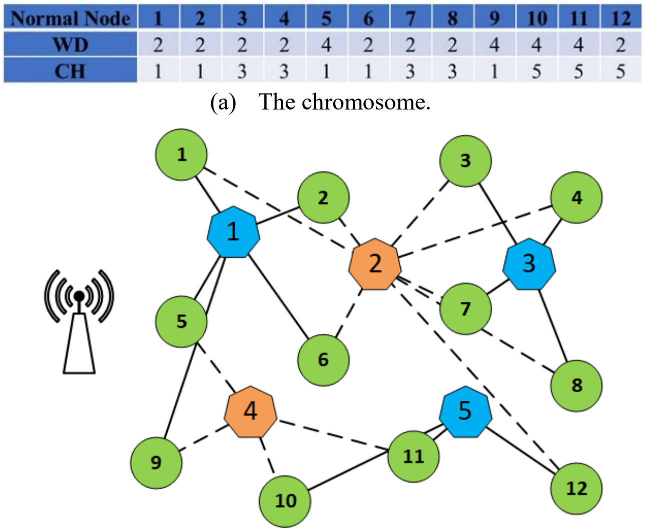
4.1. Clustering and WD Selection

The first phase of the algorithm focuses on organizing the network through clustering and selecting WD nodes. In this phase, each normal node is assigned to a super node that serves as its CH. Simultaneously, a subset of the super nodes is designated as WDs. This monitoring includes detecting potential anomalies and verifying the correct transmission of sensed data.

It is crucial to highlight that a super node cannot simultaneously serve as both a CH and a WD. These roles require distinct functionalities and impose different computational and communication overheads. A WD must remain independent of cluster management duties to dedicate its full capacity to surveillance tasks, whereas a CH is primarily responsible for aggregating data and maintaining intra-cluster coordination.

4.1.1. Chromosome Representation and Initialization

Since the optimization process simultaneously addresses CH selection and WD assignment, each chromosome is structured as a matrix with two rows. The first row encodes the selection of WD nodes for each normal node, while the second row encodes the corresponding CH selection. Given that both a WD and a CH must be designated for every normal node, the chromosome consists of $|NN|$ columns. Thus, the complete chromosome has a dimensionality of $2 \times |NN|$, with each column representing the assignment of one normal node to a specific WD and CH. This structure ensures that both roles are simultaneously optimized across the population. Figure 2 illustrates an example chromosome and its corresponding network topology. In this figure, the circles demonstrate the normal nodes, and the heptagons show the super nodes. Additionally, the dashed lines show the monitoring and the complete lines show data transmission.



Normal Node	1	2	3	4	5	6	7	8	9	10	11	12
WD	2	2	2	2	4	2	2	2	4	4	4	2
CH	1	1	3	3	1	1	3	3	1	5	5	5

(a) The chromosome.

(b) The corresponding network.

Figure 2. An example chromosome for the first phase.

To initiate the population, a first generation must be generated as follows: Candidate WDs are selected from the pool of super nodes based on their associated trust scores from previous rounds. Once the eligible super nodes for WD roles are determined (for the current chromosome), the first row of the chromosome is populated accordingly, taking into account their respective monitoring ranges. If a normal node has no suitable WD within range, its corresponding slot in the chromosome remains empty. If exactly one WD is available, it is assigned directly. In cases where multiple WDs are within range, one is selected at random. CHs are selected from the set of super nodes that not assigned as WDs in the current chromosome. For each normal node, a CH is chosen randomly from among the super nodes within its transmission range.

4.1.2. Fitness Function

Equation (5) shows the fitness function of a sample chromosome X^C:

$$fitness(X^C) = w_1\, fitness(CH) + w_2\, fitness(WD^C), \qquad w_1 + w_2 = 1 \qquad (5)$$

This equation consists of two parts: first, the fitness of CH selection as shown in (6), and second, the fitness of WD selection as shown in (7).

The fitness of CH selection is calculated by (6). In this equation, er_i stands for the remaining energy of super node sn_i based on the current chromosome structure. Additionally, e_{init} stands for the initial energy of super nodes. The division by e_{init} is for normalization into the range of [0, 1]. Consequently, the min function finds the minimum remaining energy using the current chromosome structure and we try to maximum it.

$$fitness(CH) = \min_{sn_i \in CH}\left(\frac{er_i}{e_{init}}\right) \qquad (6)$$

Equation (7) calculates the fitness of proposed WD selection structure. In this equation, $trust(sn_i)$ stands for the current trust of sn_i. This score is a dynamic value in the range of [0, 1] and always is updated based on how this node works in the network. The better sn_i forwards its data, the higher score it gets. At the beginning of the network work all the super nodes have the trust of one.

$$fitness(WD^{\mathrm{C}}) = \frac{\sum_{sn_i \in WD} \mathrm{trust}(sn_i)}{|WD|} \qquad (7)$$

In the GA algorithm, we try to maximize the associate fitness function mentioned in (5) through operators that are discussed in the following.

4.1.3. Operators

The proposed GA explores and exploits the solution space using customized operators. Below, we describe how each operator functions.

Crossover: The crossover operator first selects two parent chromosomes and determines a common crossover point. The offspring are generated by combining the left-hand side of one parent with the right-hand side of the other, and vice versa. This operation produces two offspring. The resulting chromosomes may be either valid or invalid, which are mitigated as follows:

- Valid chromosome: Meets the predefined constraints on the number of WDs.
- Invalid chromosome: Contains more or less WDs than allowed, due to the combination of genetic material from the two parents.

For an invalid chromosome, a validation process is applied:

1. WD validation: First, the algorithm counts the number of WD nodes in the chromosome. If this number is higher than the allowed limit, nodes with the lowest trust levels are removed and replaced with neighboring WDs until the total number of WDs meets the predefined threshold. Similarly, if this number is less than the allowed limit, CHs with the highest trust levels are converted to WDs, which monitor their nearby nodes.
2. CH validation: The selected CHs in the chromosome are checked. If a super node is assigned as both a WD and a CH at the same time, it is replaced with a nearby super node chosen from the valid CH list.

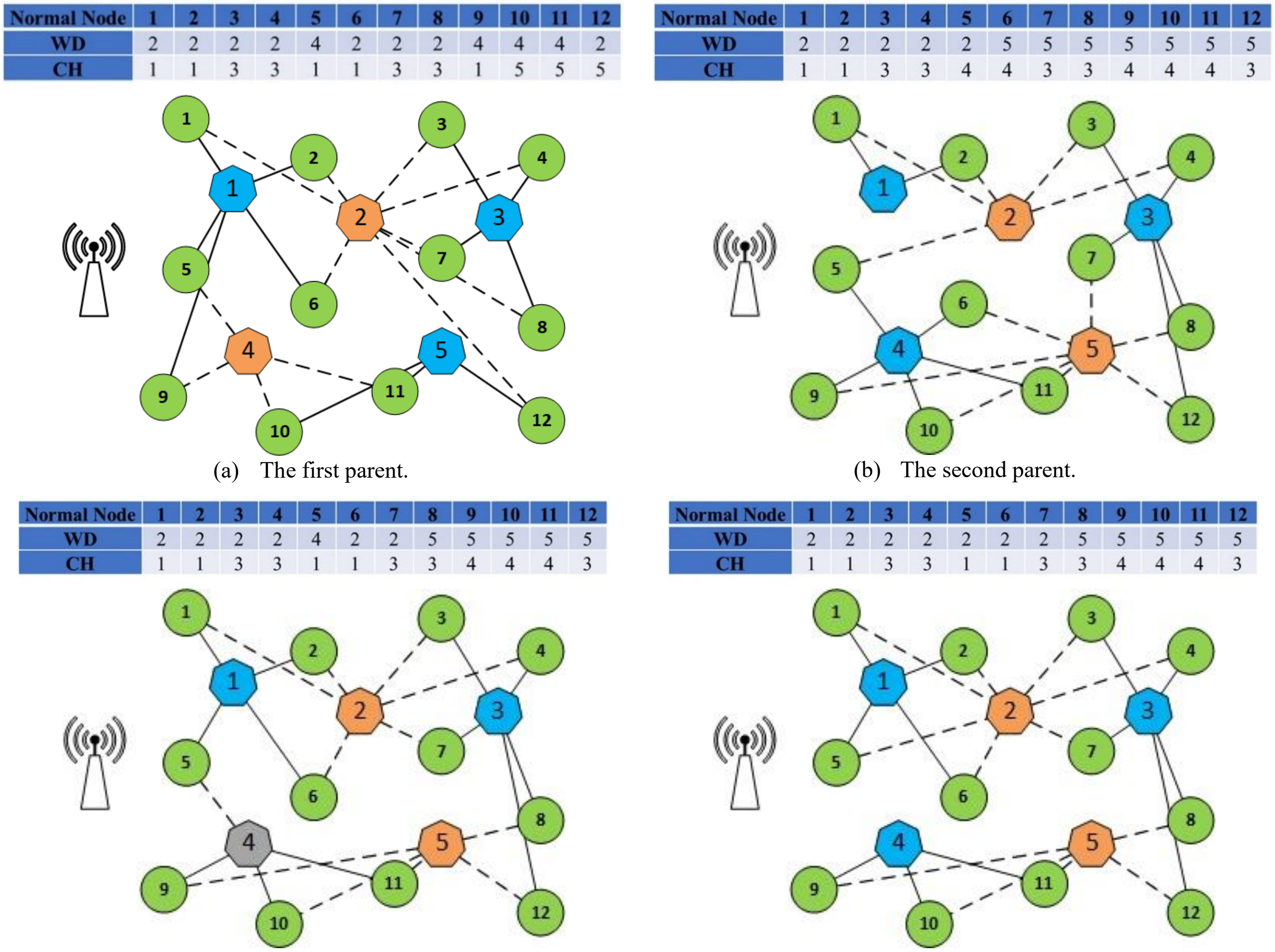
Figure 3 demonstrates a crossover example along with its validation process. The chromosomes shown in Figure 3(a) and Figure 3(b) are selected as the parents. After performing crossover at the point that separates sn_7 and sn_8, two offspring are generated. For demonstration purposes, only one of them is illustrated in Figure 3(c). As the figure shows, after crossover, we obtain three WDs, whereas in this example we should only have two. It is assumed that the trust level of super node sn_4 is less than that of sn_2 and sn_5. Consequently, it is revoked from its role as a WD, and sn_2 is assigned as the WD of nn_5. The final chromosome is demonstrated in Figure 3(d).

Mutation: The mutation operator exploits the solution space of the problem. It is applied to a chromosome as described in the following:

1. Gene selection: First, some genes of selected chromosome are selected to be mutated. In the earlier iterations, as we look to explore the solution space the number of selected genes for mutation is higher. As we proceed and we look for exploitation rather than exploration, it becomes lower. The number of selected genes iteration t, $m(t)$, is computed as (8). In this equation, m_{init} and itr denote the initial number of genes undergoing mutation and number of iterations, respectively.

$$m(t) = m_{init}\left(1 - \frac{t}{itr}\right) \tag{8}$$

2. Mutating selected genes: We change the WD and CH for selected genes (i.e., normal nodes) from possible valid sets of WDs and CHs.



Normal Node	1	2	3	4	5	6	7	8	9	10	11	12
WD	2	2	2	2	4	2	2	2	4	4	4	2
CH	1	1	3	3	1	1	3	3	1	5	5	5

(a) The first parent.

Normal Node	1	2	3	4	5	6	7	8	9	10	11	12
WD	2	2	2	2	2	5	5	5	5	5	5	5
CH	1	1	3	3	4	4	3	3	4	4	4	3

(b) The second parent.

Normal Node	1	2	3	4	5	6	7	8	9	10	11	12
WD	2	2	2	2	4	2	2	5	5	5	5	5
CH	1	1	3	3	1	1	3	3	4	4	4	3

(c) The first offspring before repairment.

Normal Node	1	2	3	4	5	6	7	8	9	10	11	12
WD	2	2	2	2	2	2	2	5	5	5	5	5
CH	1	1	3	3	1	1	3	3	4	4	4	3

(d) The final first offspring.

Figure 3. Crossover example for the first phase.

Selection: We use Roulette Wheel Selection (RWS) as described in (9) as our selection operator. As it is shown in this equation, RWS gives more chance to the chromosomes with better solution (higher fitness values) to participate in the next generations. However, it gives some chances to other chromosomes in the case that may a part of optimal solution be on their genes.

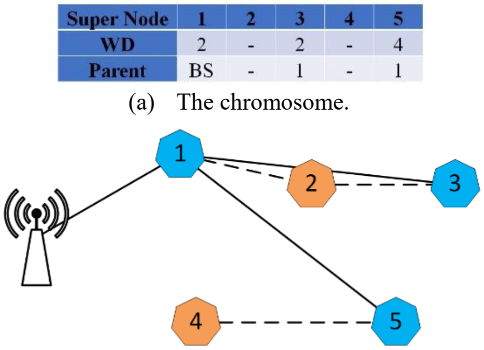
$$P\left(X_i^C\right) = \frac{fitness\left(X_i^C\right)}{\sum_{j \in pop} fitness\left(X_j^C\right)} \tag{9}$$

4.2. Tree Construction and WD Assignment

The gathered data must be transmitted to the BS to be processed. Consequently, this phase involves constructing a spanning tree on the non-WD super nodes. Also, to ensure proper delivery, WDs are designated to monitor the transmission process. As in the previous phase, we use GA to optimize the process. The subsequent sections detail the steps involved in this process.

4.2.1. Chromosome Representation and Initialization

As described, this phase entails constructing a spanning tree over the super nodes and assigning WDs. The chromosome is structured as a two-row matrix: the first row indicates the corresponding WD assignments, while the second row specifies the parent for each super node. Consequently, each chromosome forms a $2 \times |SN|$ matrix. Figure 4 illustrates an example chromosome alongside its associated network topology. In this figure, sn_2 and sn_4 are WDs that are selected in the previous phase and sn_1, sn_3, and sn_5 are the super nodes that acted as CHs in the previous phase and should deliver their data to the BS. As it is evident, sn_2 acts as WD for sn_1 and $sn_3$1, and sn_5 is monitored by sn_4.



Super Node	1	2	3	4	5
WD	2	-	2	-	4
Parent	BS	-	1	-	1

(a) The chromosome.

(b) The corresponding network.

Figure 4. An example chromosome for the first phase.

4.2.2. Fitness Function

We define the fitness function for this phase in (10). It uses two main criteria that we discuss in the following.

$$fitness(X^T) = w_3\, fitness(tree) + w_4\, fitness(WD^{\mathrm{T}}), \qquad w_3 + w_4 = 1 \qquad (10)$$

where X^T indicates a chromosome for tree construction and WD assignment.

The first criterion is calculated as (11). It evaluates the quality of the spanning tree represented by chromosome X^T by calculating the average remaining energy of all CHs at the end of the round. In this equation, er_i represents the remaining energy of node sn_i at the end of the round according to the structure defined by the current chromosome X^T. This value is normalized by dividing it by e_{init} in order to be comparable with the other criterion.

$$fitness(tree) = \frac{\sum_{sn_i \in CH} \frac{er_i}{e_{init}}}{|CH|} \tag{11}$$

Equation (12) is defined to improve the efficiency of WD selection. This function assigns the best WDs (the WDs with the most trust) to the CHs that have more data to relay. In this function, $monitor(wd_i)$ stands for the CHs that are monitored by wd_i. Additionally, $relay(ch_j)$ stands for the amount of data relayed by ch_j based on the structure of the network proposed by X^T.

$$fitness(WD^{\mathrm{T}}) = \frac{\sum_{wd_i \in WD}\left(\sum_{ch_j \in monitor(wd_i)}\left(trust(wd_i) \times relay(ch_j)\right)\right)}{|WD|} \tag{12}$$

4.2.3. Operators

In the following, we discuss the used GA operators in the second phase.

Crossover: The crossover operator employed in this phase is the single-point crossover. A crossover point is randomly selected on the parent chromosomes, and the genetic material to the left and right of this point is exchanged, producing two new offspring. Since the WDs are determined in the previous phase, and this phase only assigns the WDs to CHs WD, the generated offspring remain to be valid by the design.

Mutation: A chromosome can undergo mutation in two different ways:

1. The WD responsible for monitoring the behavior of a super node may be reassigned.
2. The parent of a CH may be altered.

In both cases, the mutated gene acquires its new value from the valid sets of WDs and CHs, respectively. This scheme ensures that the resulting configuration remains feasible and applicable within the network.

Selection: The RWS method is employed as the selection operator in this phase. Individuals are assigned selection probabilities proportional to their fitness values. This technique ensures that fitter chromosomes have a higher chance of being chosen for reproduction, while still maintaining diversity by allowing fewer fit candidates to be selected occasionally.

5. Experimental Results

This section reports the results obtained from the proposed algorithm. Its performance is evaluated against four existing algorithms: HEDHMG [19], EFEBPSO [21], EFCRPSO [22], and CRCGA [23]. All algorithms are implemented using WSNSimPy [37], a Python-based discrete event simulation framework for WSNs. The evaluation is carried out using three performance metrics: bits received by the BS, total energy consumption, and network lifetime. To ensure precision, each experiment is repeated five times, and the average values are considered as the final results.

The simulated WSNs cover an area of 200m×200m. Three different network settings are assumed, each varying in node density and the position of the BS. The first network consists of 30 super nodes and 200 normal nodes, with the BS located at the upper-left corner. The second network is denser, containing 50 super nodes and 300 normal nodes, with the BS also placed in the upper-left corner. The third network has the same node distribution as the second one but differs in that its BS is centrally located. In all cases, both super nodes and normal nodes are randomly distributed. The detailed simulation parameters and GA algorithm settings are provided in Table 1 and Table 2, respectively.

Table 1. Simulation parameters.

Parameter	Value
Network Dimension	200m × 200m
Initial Energy of Normal Nodes	1J
Initial Energy of Super Nodes	2J
Transmission Range of Normal Nodes	30m
Transmission Range of Super Nodes	90m
Time Slices per Round	50
θ	0.1
Percentage of Malicious Nodes	0.2

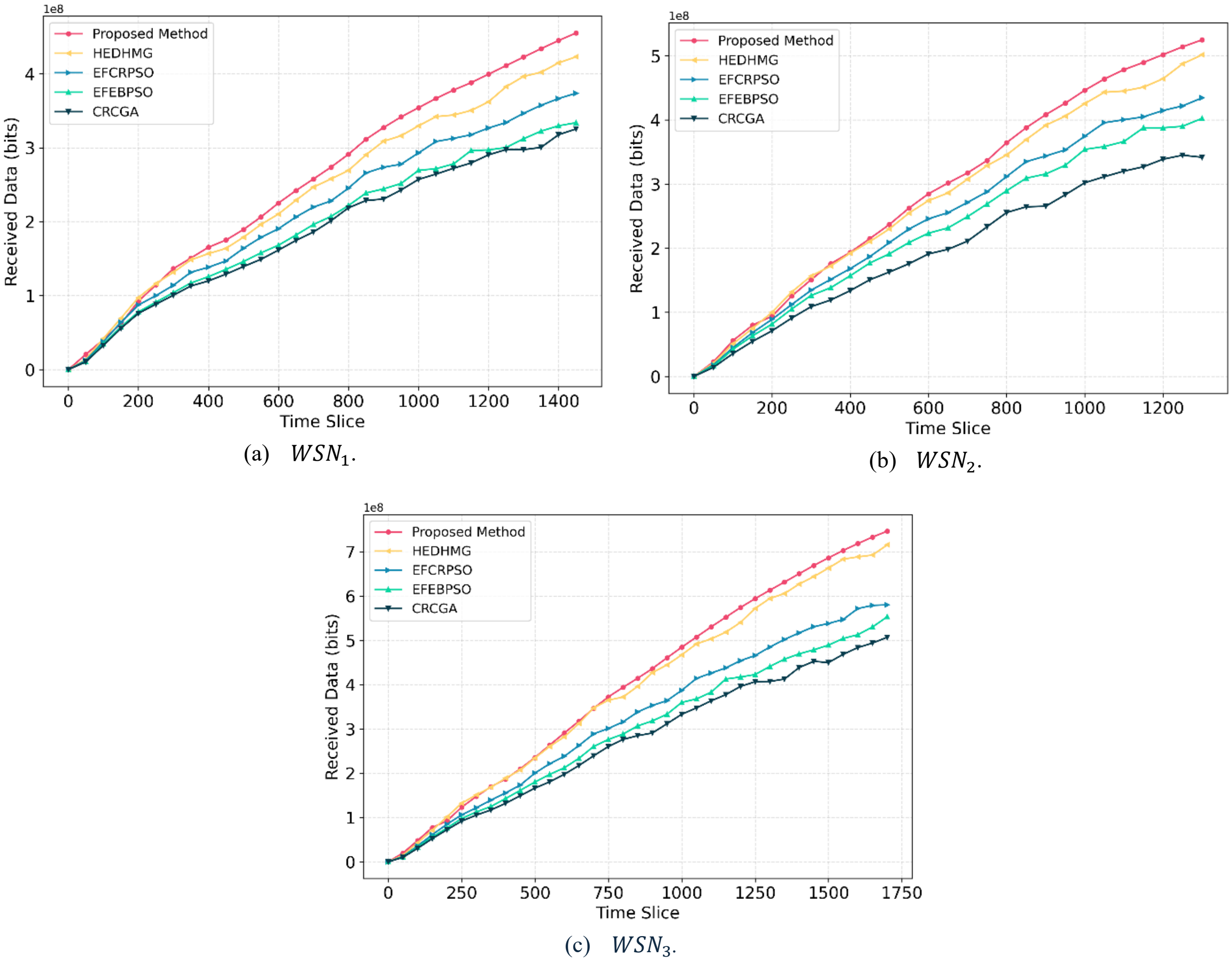
Table 2. GA parameters.

Parameter	Value
Population size	30
Number of iterations	50
w_1	0.5
w_2	0.5
w_3	0.3

w_4	0.7

5.1. Bits Received By the BS

This metric measures the delivered data to the BS. Figure 55 compares this metric for the competitive algorithms. As evident, the proposed method improves the amount of bit received by the BS in all networks. To be more precise, in WSN_1, the improvements over HEDHMG, EFCRPSO, EFEBPSO, and CRCGA are 6.5%, 19.4, 32.7%, and 38% respectively. These values are 3.8%, 17.3%, 26.8%, and 47.6% for WSN_2, and 3.7% 24.9%, 37.2%, and 47.2% for WSN_3.



(a) WSN_1.

(b) WSN_2.

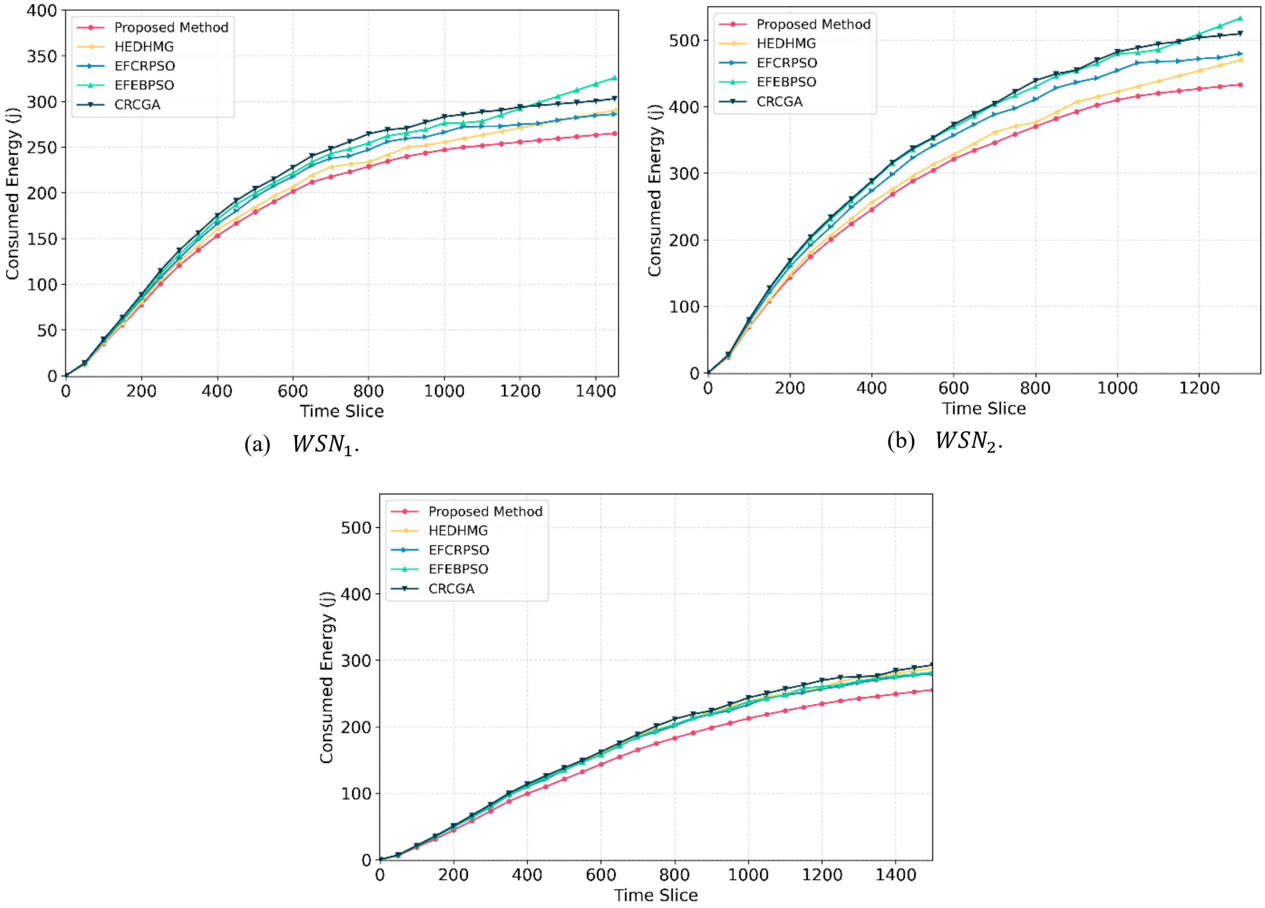
(c) WSN_3.

Figure 5. Number of bits received by the BS comparison.

As it is shown in the figure, in the earlier rounds, when the algorithm does not recognize the malicious super nodes, the amount of delivered data to the BS using the proposed method is close to the other algorithms. However, as time passes, the proposed method learns about the malicious nodes and does not use them. More precisely, in the first phase, it is preferred to use CHs with higher trusts. In the second phase, the algorithm avoids using the malicious nodes as relay nodes in the spanning tree. In both phases, low-trust super nodes are assigned fewer tasks. Accordingly, more data is delivered to the BS.

5.2. Total Consumed Energy

Limited battery capacity makes total energy consumption one of the most important factors in designing WSNs. **Error! Reference source not found.** investigates this metric. As it is evident by the figure, in all three networks, the proposed method has the least total consumed energy, despite that it delivers the most data to the BS. To be more precise, in WSN_1, the proposed method reduces the total consumed energy by 4.2%, 7.5%, 11.1%, and 12.4% over HEDHMG, EFCRPSO, EFEBPSO, and CRCGA respectively. The values are 3.5%, 9.9%, 14.2%, and 14.6% for WSN_2 and 10.8%, 9%, 9.8%, and 12.3% for WSN_3.



(a) WSN_1.

(b) WSN_2.

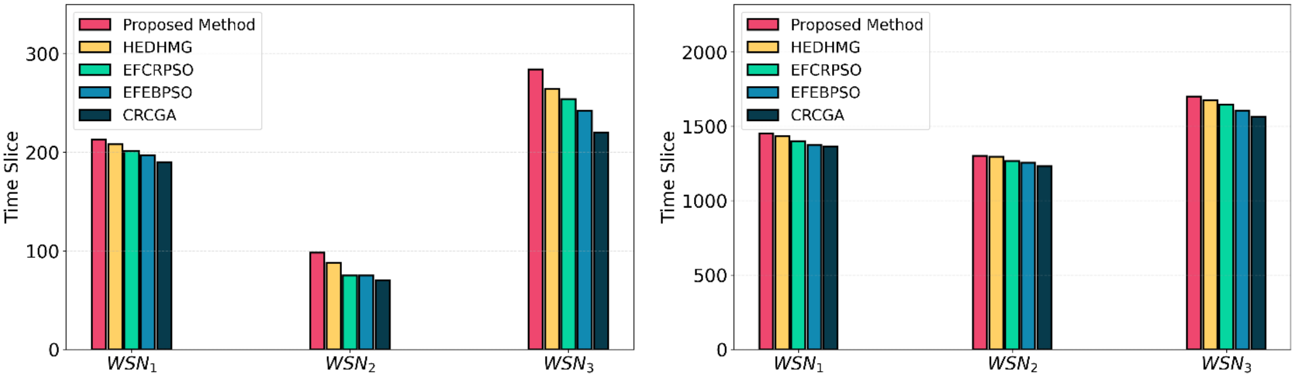
(c) WSN_3.

Figure 6. Total consumed energy comparison.

The better results of the proposed method are the direct effect of using energy-aware metrics in both phases. In the first phase, the super nodes with the most remaining energy are selected as CHs. Also, in the second phase, the algorithm constructed the tree in such a way that the load on the super nodes is distributed. These two factors, combined, leads to a network which excels in the usage of energy.

5.3. The Network Lifetime

Another metric that is important for WSNs is the network lifetime. We consider two metrics. The first measure, First Node Die (FND), shows the time slice in which the first super node of the network loses all its energy and becomes unusable. The second metric, Last Node Die (LND), indicates the time slice that all of the super nodes become unusable. **Error! Reference source not found.** compares FND and LND for the competitive algorithms.



(a) FND. (b) LND.

Figure 7. Network Lifetime comparison.

As it is comprehendible from the figure, the proposed method shows better network lifetime in all three networks. To be more precise, in WSN_1, the proposed method performs 5, 12, 16, and 23 more time slices compared to HEDHMG, EFCRPSO, EFEBPSO, and CRCGA, respectively until the FND happens. These values are 10, 23, 23, and 28 for WSN_2, and 20, 30, 42, 64 for WSN_3. In all three networks, the proposed method is the last method to fail. It outperforms its competitors in the following number of time slices: for $WSN_1$19, 51, 78, and 85; for WSN_2 6, 35, 46, and 70; and for WSN_3 25, 55, 95, and 135. It should be noted that the proposed algorithm achieves better FND and LND despite of delivering more data to the BS compared to other algorithms.

The better results of the proposed method are due to its more efficient clustering and tree construction methods. In the clustering phase, our energy-aware metric selects CHs by considering their remaining energy at the end of the round, which covers the current amount of energy and energy consumption rate of CHs. This metric helps to balance energy use among all nodes. In the tree construction phase, the well-designed routing structure shortens communication paths, thereby reducing overall energy dissipation. Together, these mechanisms extend the lifetime of super nodes while maintaining high data delivery rates to the BS, which explains the superior performance of the proposed method compared to the competing algorithms.

6. Conclusion and Future Works

This paper introduced a novel two-phase clustering, routing, and WD selection algorithm designed for HWSNs. The first phase included clustering and determining the WDs, while in the second phase a spanning tree was constructed over the CHs and proper WD was assigned per CH. By enforcing strict, mutually exclusive roles between CHs and WDs, the proposed GA-driven framework optimized both network longevity and security. Simulation results demonstrated that this joint optimization method significantly improved energy efficiency, and resilience against node compromise when compared to established baseline schemes. Future research directions could include extending the approach to dynamic network conditions, supporting mobile nodes, and developing adaptive mechanisms for real-time trust assessment and WD re-deployment. Additionally, exploring integrations with other advanced metaheuristics for intrusion detection could further enhance both scalability and security in large-scale deployments.

References

[1] S. Kumari and A. K. Tyagi, "Wireless sensor networks: An introduction," *Digital Twin and Blockchain for Smart Cities,* pp. 495-528, 2024.

[2] S. Verma, N. Sood, and A. K. Sharma, "Genetic algorithm-based optimized cluster head selection for single and multiple data sinks in heterogeneous wireless sensor network," *Applied Soft Computing,* vol. 85, p. 105788, 2019.

[3] V. Pal, G. Singh, and R. Yadav, "Cluster head selection optimization based on genetic algorithm to prolong lifetime of wireless sensor networks," *Procedia Computer Science,* vol. 57, pp. 1417-1423, 2015.

[4] Z. Al Aghbari, A. M. Khedr, W. Osamy, I. Arif, and D. P. Agrawal, "Routing in wireless sensor networks using optimization techniques: A survey," *Wireless Personal Communications,* vol. 111, no. 4, pp. 2407-2434, 2020.

[5] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Communications Surveys & Tutorials,* vol. 19, no. 1, pp. 550-586, 2016.

[6] Y. Wang, X. Wang, B. Xie, D. Wang, and D. P. Agrawal, "Intrusion detection in homogeneous and heterogeneous wireless sensor networks," *IEEE Transactions on Mobile Computing,* vol. 7, no. 6, pp. 698-711, 2008.

[7] M. Krzysztoń and M. Marks, "Simulation of watchdog placement for cooperative anomaly detection in bluetooth mesh intrusion detection system," *Simulation Modelling Practice and Theory,* vol. 101, p. 102041, 2020.

[8] N. A. Alrajeh, S. Khan, and B. Shams, "Intrusion detection systems in wireless sensor networks: A review," *International Journal of Distributed Sensor Networks,* vol. 9, no. 5, p. 167575, 2013.

[9] L. Du, Q. Wang, and Z. Zhang, "Reinforcement learning-driven cluster head selection for reliable data transmission in dense wireless sensor networks," *International Journal of Advanced Computer Science & Applications,* vol. 16, no. 4, 2025.

[10] A. K. Jemla Naik, M. Parameswarappa, and M. N. Ramachandra, "Multiobjective, trust-aware, artificial hummingbird algorithm-based secure clustering and routing with mobile sink for wireless sensor networks," *ETRI Journal,* vol. 46, no. 6, pp. 950-964, 2024.

[11] F. Bao, R. Chen, M. Chang, and J.-H. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection," *IEEE Transactions on Network and Service Management,* vol. 9, no. 2, pp. 169-183, 2012.

[12] Y. F. Ebobissé Djéné, M. S. El Idrissi, P.-M. Tardif, A. Jorio, B. El Bhiri, and Y. Fakhri, "A formal energy consumption analysis to secure cluster-based WSN: A case study of multi-hop clustering algorithm based on spectral classification using lightweight blockchain," *Sensors,* vol. 22, no. 20, p. 7730, 2022.

[13] P. Anusuya and C. Vanitha, "A comprehensive review of sensor node deployment strategies for maximized coverage and energy efficiency in wireless sensor networks," *PeerJ Computer Science,* vol. 10, p. e2407, 2024.

[14] R. Priyadarshi, "Efficient node deployment for enhancing coverage and connectivity in Wireless Sensor Networks," *Scientific Reports,* vol. 15, no. 1, p. 29052, 2025.

[15] M. Li, S. Zhang, Y. Cao, and S. Xu, "NMSFRA: Heterogeneous routing protocol for balanced energy consumption in mobile wireless sensor network," *Ad Hoc Networks,* vol. 145, p. 103176, 2023.

[16] P. Divya and B. Sudhakar, "Route optimization and optimal cluster head selection for cluster-oriented wireless sensor network utilizing circle-inspired optimization algorithm," *International Journal of Computational Intelligence Systems,* vol. 17, no. 1, p. 302, 2024.

[17] K. Khedhiri, I. Ben Omrane, D. Djabour, and A. Cherif, "Clustering for lifetime enhancement in wireless sensor networks," *Telecom*, vol. 6, no. 2, p. 30, 2025.

[18] A. Ojha and B. Gupta, "Evolving landscape of wireless sensor networks: a survey of trends, timelines, and future perspectives," *Discover Applied Sciences,* vol. 7, no. 8, p. 825, 2025.

[19] M.-S. Shahryari, L. Farzinvash, M.-R. Feizi-Derakhshi, and A. Taherkordi, "High-throughput and energy-efficient data gathering in heterogeneous multi-channel wireless sensor networks using genetic algorithm," *Ad Hoc Networks,* vol. 139, p. 103041, 2023.

[20] M.-S. Shahryari, L. Farzinvash, and M.-R. Feizi-Derakhshi, "Cost-efficient network design in multichannel WSNs with power control: A grey wolf optimization approach to routing and clustering," *International Journal of Distributed Sensor Networks,* vol. 2024, no. 1, p. 1357195, 2024.

[21] M. Azharuddin and P. K. Jana, "PSO-based approach for energy-efficient and energy-balanced routing and clustering in wireless sensor networks," *Soft Computing,* vol. 21, no. 22, pp. 6825-6839, 2017.

[22] P. Kuila and P. K. Jana, "Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach," *Engineering Applications of Artificial Intelligence,* vol. 33, pp. 127-140, 2014.

[23] C. Wang, X. Liu, H. Hu, Y. Han, and M. Yao, "Energy-efficient and load-balanced clustering routing protocol for wireless sensor networks using a chaotic genetic algorithm," *IEEE Access,* vol. 8, pp. 158082-158096, 2020.

[24] E. H. Houssein, M. R. Saad, Y. Djenouri, G. Hu, A. A. Ali, and H. Shaban, "Metaheuristic algorithms and their applications in wireless sensor networks: review, open issues, and challenges," *Cluster Computing,* vol. 27, no. 10, pp. 13643-13673, 2024.

[25] M. Kaedi, A. Bohlooli, and R. Pakrooh, "Simultaneous optimization of cluster head selection and inter-cluster routing in wireless sensor networks using a 2-level genetic algorithm," *Applied Soft Computing,* vol. 128, p. 109444, 2022.

[26] V. Prakash and S. Pandey, "Metaheuristic algorithm for energy efficient clustering scheme in wireless sensor networks," *Microprocessors and Microsystems,* vol. 101, p. 104898, 2023.

[27] B. M. Sahoo, H. M. Pandey, and T. Amgoth, "A genetic algorithm inspired optimized cluster head selection method in wireless sensor networks," *Swarm and Evolutionary Computation,* vol. 75, p. 101151, 2022.

[28] H. Khujamatov, M. Pitchai, A. Shamsiev, A. Mukhamadiyev, and J. Cho, "Clustered routing using chaotic genetic algorithm with grey wolf optimization to enhance energy efficiency in sensor networks," *Sensors*, vol. 24, no. 13, p. 4406, 2024.

[29] K. Sharada, T. Mahesh, S. Chandrasekaran, R. Shashikumar, V. V. Kumar, and J. R. Annand, "Improved energy efficiency using adaptive ant colony distributed intelligent based clustering in wireless sensor networks," *Scientific Reports,* vol. 14, no. 1, p. 4391, 2024.

[30] M. K. Roberts, J. Thangavel, and H. Aldawsari, "An improved dual-phased meta-heuristic optimization-based framework for energy efficient cluster-based routing in wireless sensor networks," *Alexandria Engineering Journal,* vol. 101, pp. 306-317, 2024.

[31] C. Kurangi, K. K. Paidipati, A. S. K. Reddy, J. Uthayakumar, G. Kadiravan, and S. Parveen, "Metaheuristic optimization-based clustering with routing protocol in wireless sensor networks," *International Journal of Communication Systems,* vol. 37, no. 16, p. e5914, 2024.

[32] M.-S. Shahryari, L. Mohammad-Khanli, M. Ramezani, L. Farzinvash, and M.-R. Feizi-Derakhshi, "Nesting circles: An interactive visualization paradigm for network intrusion detection system alerts," *Security and Communication Networks,* vol. 2023, no. 1, p. 8043619, 2023.

[33] V. Sivagaminathan, M. Sharma, and S. K. Henge, "Intrusion detection systems for wireless sensor networks using computational intelligence techniques," *Cybersecurity,* vol. 6, no. 1, p. 27, 2023.

[34] M. Hosseini Shirvani and A. Akbarifar, "A survey study on intrusion detection system in wireless sensor network: Challenges and considerations," *Journal of Electrical and Computer Engineering Innovations*, vol. 12, no. 2, pp. 449-474, 2024.

[35] C. K. K. Reddy, V. S. Kaza, P. R. Anisha, M. M. Khubrani, M. Shuaib, S. Alam, and S. Ahmad, "Optimising barrier placement for intrusion detection and prevention in WSNs," *Plos one,* vol. 19, no. 2, p. e0299334, 2024.

[36] N. Liu, S. Liu, and W.-M. Zheng, "PPSO and Bayesian game for intrusion detection in WSN from a macro perspective," *Complex & Intelligent Systems,* vol. 10, no. 6, pp. 7645-7659, 2024.

[37] A. Marchiori, L. Guo, J. Thomas, and Q. Han, "Realistic performance analysis of WSN protocols through trace based simulation," in *Proceedings of the 7th ACM workshop on performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, pp. 87-94, 2010.