

Genetic Algorithm Energy Optimization in 3D WSNs with Different Node Distributions

Yousef Jaradat*, Mohammad Masoud, Ismael Jannoud and Dema Zeidan

Department of Electrical Engineering, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan

*Corresponding Author: Yousef Jaradat. Email: y.jaradat@zu.jo

Received: 09 October 2021; Accepted: 24 November 2021

Abstract: Optimal node clustering in wireless sensor networks (WSNs) is a major issue in reducing energy consumption and extending network node life time and reliability measures. Many techniques for optimizing the node clustering process in WSN have been proposed in the literature. The metaheuristic algorithms are a subset of these techniques. Genetic algorithm (GA) is an evolutionary metaheuristic technique utilized to improve the network reliability and extending the network life time by optimizing the clustering process in the network. The GA dynamic clustering (GA-DC) algorithm is proposed in this paper to extend the network reliability and node life time of three dimensional (3D) WSN. The GA-DC algorithm made use of an improved fitness function that takes into account a variety of metrics such as energy expenditure per protocol round, clustering distance, and the number of long-distance wireless connections. There have been two types of simulation scenarios run. First, simulation results show that the GA-DC algorithm increases network life time by 80% and network throughput by 55% when compared to the well-known LEACH protocol. Second, simulation results show that the uniform node distribution outperforms the normal and exponential distributions in terms of network life time by 5.7% and 7%, network reliability by 4.2% and 76%, and data throughput by 10.85% and 19.54%, respectively

Keywords: Dynamic GA; selection; mutation; crossover; clustering; energy efficient

1 Introduction

Rapid technological advances in the field of micro electro mechanical systems (MEMS) have led to the development of cheap and autonomous miniature sensor nodes [1]. These sensor nodes have the capabilities of sensing and monitoring the environment, processing and aggregating data, and communicating data reports to each other or to a central location, usually referred to as the sink or the base station (BS). Wireless sensor network (WSN) is the interconnection of large number of these sensor nodes to serve an application-specific purpose [2]. Sensor nodes have limited battery energy sources and these batteries are mostly not rechargeable. The communication function of a sensor node usually consume energy the most [3]. If the energy source of a particular node gets drained, the node is considered dead and it becomes



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

useless. WSNs have a wide variety of real-world applications. They are used in both civilian and military applications, including agriculture, industry, healthcare, surveillance, target tracking, and security management [4]. Fig. 1 shows a typical sensor node block diagram [5].

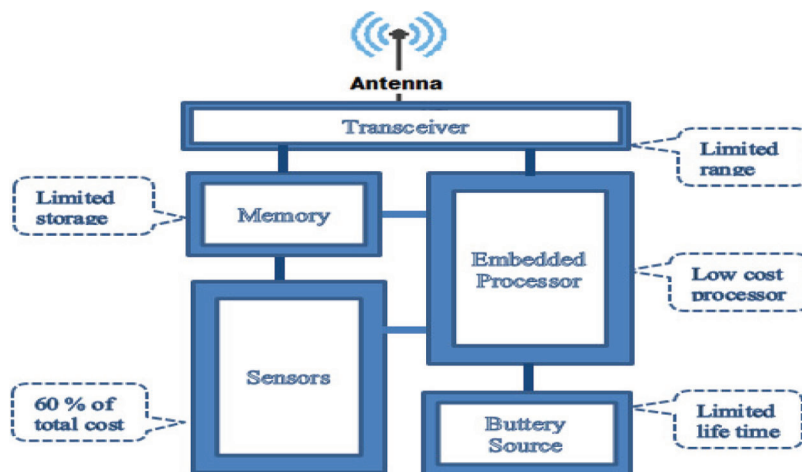


Figure 1: Typical sensor node block diagram

Sensor nodes can send their collected data directly to the sink node [6], but this consumes more energy and leads to the node's premature death, introducing other problems into the network such as the coverage/hole problem [7]. Clustering is utilized to balance energy consumption between sensor nodes. Clustering is based on grouping nearby nodes or nodes with similar characteristics or purpose. Nodes in a cluster can be classified into two categories: cluster head (CH) and member nodes (MN). MNs collect data reports from the environment and send them to the CHs. Moreover, CHs aggregate, compress and send the information to the BS node. The CH to BS communication can be done directly or through multiple hop techniques [8]. CHs usually consume more energy compared to the MNs. Rotation of the CH role between nodes in the network is required to balance the overall network energy and extend the network life time [9]. Many protocols have been developed for clusters creation, operation, and maintenance. These include low energy adaptive clustering hierarchy (LEACH), stable energy protocol (SEP), hybrid energy efficient distributed (HEED) and many more [10]. The inability of most clustering protocols in WSNs to reduce overall network energy and balance energy consumption among sensor nodes is a known issue [8]. In some cases, a significant reduction in overall network energy expenditure is achieved at the expense of unequal remaining energies among sensor nodes. In other cases, sensor node remaining energies are balanced at the expense of increasing the overall network consumed energy.

Nature-inspired or meta-heuristic algorithms have been applied to WSNs to achieve a balance between balancing energy consumption among sensor nodes and lowering overall energy consumption. Genetic algorithm (GA) is one of the most popular and few metaheuristics techniques used for solving the four fundamental problems of WSNs. These include: energy efficient clustering, localization of sensor nodes, data aggregation, and optimal coverage [11]. In this paper, GA dynamic clustering is used to improve the lifetime and reliability of three-dimensional (3D) WSNs by optimizing the number of clusters in the network and balancing consumed energies among sensor nodes for three different randomly distributed sensor nodes. In typical WSN settings, two dimensional network (2D) and uniform node distribution are studied. In real-world applications, the environment is 3D, and sensor nodes are distributed based on terrain and deployment location. Sensor deployment methods can be deterministic or random. Any

distribution can be used for random deployment. Three randomly distributed techniques are studied and compared in this work. Uniform, normal, and exponential distributions are our case studies.

This paper has the following contributions: First, GA dynamic clustering (GA-DC) algorithm is proposed to extend the life time of 3D WSN and network reliability. Second, three random deployment strategies of sensor nodes are investigated and evaluated for better network life time and reliability. Three, a new GA fitness function is used to optimize the number of clusters in the network, which essentially includes information about the type of connections MNs establish with CHs in a cluster and CHs establish with BSs in the network.

The rest of the paper is organized as follows. Section 2 overviews the related work. Section 3 describes GA operators' details. In Section 4 GA fitness function is introduced along with cluster formation and maintenance phases. Section 5 provides the simulation setups and configuration. Simulation results are explained in Section 6. Finally, we conclude our research in Section 7.

2 Related Work

A lot of researches have been conducted utilizing nature inspired algorithms for optimizing WSNs life time. GA has showed a great applicability and solutions to different challenges in WSNs. In [12], the authors have proposed GA based energy efficient clustering hierarchy (GAEECH) protocol, the purpose of this protocol is to reduce the overall network energy and extends the life span of the network. Their results have shown great increase in network life time compared to other renowned clustering protocols especially when the BS is located outside the network area. The authors in [13] have proposed GAEEC, GA based energy efficient clustering, which divided the network into optimal number of clusters utilizing GA, then GA is utilized again to select the suitable node within a cluster as a CH based on their fitness function per protocol round. Their results have shown improvement over the state-of-art LEACH protocol in terms of stable period and network throughput. In [14], the authors have proposed genetic algorithm-based optimized clustering (GAOC) and multiple data sinks based GAOC (MS-GAOC) for CH optimization in heterogeneous WSN. GAOC is utilized to select the optimal number of CHs in the network by evaluating a cost function that depends on node density, remaining energy and distance to the BS. To overcome the hot-spot problem and saving communication energies to the BS, MS-GAOC is utilized. Their results have outperformed state-of-the-art clustering protocols for different performance metrics viz. network life span, throughput, stability period and remaining energy. The authors in [15] have proposed (FCM-GA), Fuzzy-C means genetic algorithm, protocol. The protocol has two stages. In the first stage, FCM is utilized to perform clustering based on three objectives including appropriate online energy distribution of each cluster, intra-cluster distances, and distances between CHs and the BS. FCM-GA results have shown great improvement over other protocols viz. direct transmission, MH-FEER, and SHMEER in terms of network life span and network throughput. In [5], the authors have proposed variants of PEGASIS protocol utilizing genetic algorithm for energy efficient clustering and routing in 3D WSN. GA has been used to build near-minimal length chain for data transmission. New approach for devising CHs based on the distance and remaining energy for better load balancing have been introduced. Their results have shown superiority over normal PEGASIS protocol in terms of first node-die (FND) metric. The authors in [16] have proposed optimized version of the LEACH protocol (OLEACH). OLEACH uses GA for finding the optimal route. Their results have shown a reduction of 17.39% in energy consumption. In [17], the authors have studied the implication of different node deployments on the energy consumption in WSN utilizing k-means and k-medoids clustering algorithms. It has been shown that the uniform and beta distributions of node deployment with k-medoids have the least energy consumption compared to other distributions. The authors in [18] have proposed genetic algorithm based energy efficient clusters (GABEEC). GABEEC protocol is based on two phases: the

setup phase and the steady state phase. In their proposed method, the clusters created during the setup phase are not changed throughout the network. As a result, there are static clusters with dynamically changed CHs in each round. The new CH is chosen utilizing the GA and based on the residual energy of the current CH and its member nodes. GABEEC protocol outperforms the LEACH protocol in terms of extending network life time and throughput. The downside of GABEEC is that it is used in small size network in which all of the node connections utilize the free space model for communication. No long range connections are established between member nodes and CHs and between CHs and the BS. In [19] the authors have proposed 3D-cluster, a new clustering routing protocol, which innovatively combines node scheduling with routing optimization, in order to optimize network energy consumption and network lifetime. The improved genetic algorithm determines the optimal node set by minimizing the number of working nodes while satisfying the full coverage of discrete point of interest (DPOI). 3D-cluster outperforms other protocols in terms of network lifetime, throughput, and energy consumption, according to simulation results. However, the proposed algorithm is centralized, whereas the network is distributed. Furthermore, the co-frequency interference of sensor nodes is not taken into account in this protocol.

3 Genetic Algorithm Overview

GA is an evolutionary, bio-inspired, meta-heuristic and stochastic optimization algorithm [11]. GA is composed of three operators; selection (reproduction), crossover and mutation. The solution of GA is encoded in a string (binary or decimal) referred to as chromosome or individual. Elements of a chromosome are called genes. All chromosomes should have the same number of genes. A fitness (cost) function is used to evaluate the fitness value of a chromosome. A collection of chromosomes are called population. Fig. 2, Shows a general GA flow diagram [13].

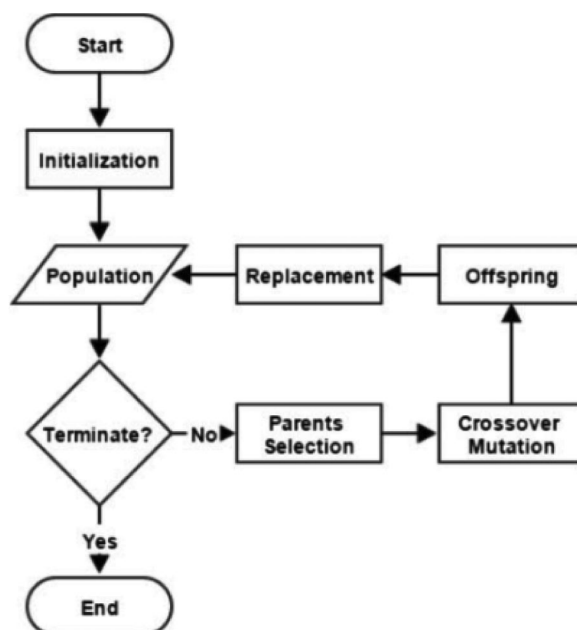


Figure 2: GA flow diagram

GA is an iterative algorithm. It applies the three operators at each run to find new chromosome which may yield better solution. The GA operators are [20,21]:

- **Selection:** At this stage, chromosomes with higher fitness values (parents) are selected to produce next generation's offspring. Roulette wheel, tournament, rank, and elitism are some examples of the selection operator.
- **Crossover:** At this stage, the genetic material of the two selected parents are merged to produce offspring. Single point, double-point and uniform-point techniques are used individually or collectively in this phase.
- **Mutation:** At this stage, single or multiple gene modifications are applied to the newly produced chromosomes in order to broaden the search space.

GA is terminated by checking two conditions: achieving a threshold optimal solution or reaching the maximum number of algorithm repetitions. The GA then suggests the best chromosome as a solution.

4 Genetic Algorithm Dynamic Clustering Algorithm

We propose genetic algorithm dynamic clustering (GA-DC) algorithm to improve network reliability, extend the network life time and reduce the overall energy expenditure per protocol round. GA-DC algorithm uses an enhanced fitness function (F) compared to other clustering algorithms. GA-DC fitness function is comprised of the following components.

4.1 Total Energy Expenditure per Protocol Round

In order to extend network life time, energy consumption per protocol round must be minimized. The chromosome with the minimal energy expenditure is chosen as a solution. Total energy expenditure is the sum of the energy consumed within a cluster (E_C) and the energy consumed by the CH to transfer the aggregated data to the BS (E_{BS}). For a cluster of n nodes, the energy consumption is given by:

$$E_C = \sum_{i=1}^{n-1} E_{TX}(i, CH) + (n-1)E_{RX} + E_{DA} \quad (1)$$

For m clusters in the network, the energy consumption by all CHs is given by:

$$E_{BC} = \sum_{j=1}^m E_{TX}(j, BS) \quad (2)$$

The total energy (E_T) expenditure per protocol round is:

$$E_T = \sum_{j=1}^m E_C(j) + E_{BS} \quad (3)$$

$E_{TX}(i, CH)$ represents the consumed transmitted energy of the i^{th} node to its corresponding CH. $(n-1)E_{RX}$ represents the energy consumed by the CH while receiving $(n-1)$ data reports from its member nodes. E_{DA} represents the energy consumed by the CH for aggregating the $(n-1)$ reports into single data packet to be sent to the BS. $E_{TX}(j, BS)$ shows the transmitted energy consumed by the j^{th} CH to the BS. This paper uses the simple node's radio energy model [22,23]. In this model E_{TX} depends on the distance between the transmitter and the receiver to switch between the free-space and multiple-path amplification models and the number of bits (k) in a data report. E_{TX} is given by:

$$E_{TX}(k, d) = \begin{cases} k\varepsilon_{fs}d^2, & \text{if } d < d_0 \\ k\varepsilon_{mp}d^4, & \text{if } d \geq d_0 \end{cases} \quad (4)$$

d_0 is the distance threshold to swap between the free-space and multiple-path models. d_0 is given by $d_0 = \sqrt{\epsilon_{fs}/\epsilon_{mp}}$.

4.2 Clustering Distance

The clustering distance (d_C), is the sum of the distances from member nodes to their corresponding CHs and the distances from CHs to the BS per protocol round as shown in Fig. 3. d_C is given by:

$$d_C = \sum_{j=1}^m \left[\sum_{i=1}^{n-1} d(i, CH(j)) + d(CH(j), BS) \right] \quad (5)$$

$d(i, CH(j))$ and $d(CH(j), BS)$ represent the Euclidean distance from i^{th} member node to its corresponding j^{th} CH, and the distance from the j^{th} CH to the BS respectively.

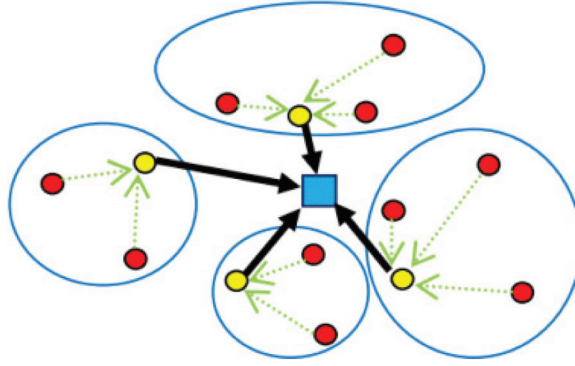


Figure 3: Clustering distance (d_C)

The total distance in the network (d_T) is the sum of distances of all nodes (N) to the BS and is given by:

$$d_T = \sum_{i=1}^N d(i, BS) \quad (6)$$

4.3 Number of CHs

The number of CHs (m) plays an important role in extending the network's life time. As previously stated, CHs consume more energy than other member nodes. Then, it is critical to reduce the number of CHs to a bare minimum.

4.4 Number of Multiple Path Propagation Connections

A node can connect to either a CH or to the BS. If the distance between a CH or the BS is greater than a threshold value (d_0), a multiple-path amplification model is used. Moreover, multiple-path amplification model uses more energy for sending data reports compared to free-space model, see Eq. (4). Multiple-path connections (C_{mp}) is the sum of the multiple-path connections within a cluster (C_{mpBS}) and the multiple-path connections to the BS (C_{mpBS}), as shown below.

$$C_{mp} = C_{mpC} + C_{mpBS} \quad (7)$$

As the value of the C_{mp} increases more energy is consumed and the life time of network decreases. Then, it is critical to keep the number of multiple-path connections to a minimum.

The fitness function (F) of GA-DC protocol is a function of four independent variables and is given by:

$$F = \omega_1 \times E_T + \omega_2 \times \frac{d_C}{d_T} + \omega_3 \times \frac{m}{N} + \omega_4 \times \frac{C_{mp}}{N} \quad (8)$$

ω_1 , ω_2 , ω_3 and ω_4 represent constant coefficients weights of the four fitness parameters. Their values are application-dependent.

GA-DC protocol has two working phases, cluster formation and data collection. The following explains the two phases.

- **Cluster formation:** At this phase GA-DC algorithm runs in the BS since it knows the geographical location of all nodes. Binary encoding technique is utilized in the GA-DC algorithm. In this approach, CHs are encoded as binary '1' and other member nodes as binary '0'. An example of binary encoding is shown in Fig. 4, in which sensor nodes S4 and S9 are considered CHs while others are considered member nodes connecting to one of these two CHs in a particular protocol run. GA-DC runs all the GA operators to a randomly generated initial population till the termination condition is satisfied. After that, the best fit chromosome solution is produced and that represents the new cluster configuration. Then, the BS will distribute the cluster settings to the network's nodes.

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
0	0	0	1	0	0	0	0	1	0

Figure 4: GA-DC binary encoded chromosome example

- **Data processing and transfer:** At this phase data collected by MNs are transferred to the CHs. Then, CHs process, aggregate and compress the collected data into packets before being transferred to the BS. CHs organize data collection from their MNs through TDMA scheduling while sending their data reports utilizing CDMA technique.

Algorithm 1 shows the pseudo code for the GA-DC algorithm.

Algorithm 1: GA-DC Algorithm

```

Random population is generated
Evaluate the Initial chromosomes' solutions using F
for (condition is true) do
    Apply selection operator (Roulette Wheel)
    Apply crossover operator (Single-point)
    Apply mutation operator (Gene bit flipping)
    Population update with new offspring
end for
Cluster formation by selecting best fit chromosome solution
Data processing and communication

```

5 Performance Metrics and Experimental Setup

In this section a number of network metrics are utilized to evaluate the performance of the GA-DC algorithm for various network configurations. These metrics include:

- Network life time (NLT): Represents the time duration from the beginning of the network to the death of the last node. NLT is usually measured by the number of protocol rounds till the network is down.
- Energy metric: Represents a number of energy parameters to measure. It includes, total remaining energy per protocol round, average energy consumption per protocol round.
- Stability period: Represents the time interval (number of rounds) till the death of the first node in the network, usually referred to as the network reliability measure. The time interval from the death of the first node (FND) to the death of the last node (LND) is referred to as the instability period.
- Network throughput: Represents the accumulated data reports reaching the BS during the network life time.

Simulation and algorithm implementation are done using MATLAB. A number of simulation scenarios are implemented. The first scenario compares the performance of the GA-DC algorithm to the state-of-the-art LEACH protocol in a 3D network setting with the BS in the middle of the network and a uniform node distribution. The second scenario involves evaluating the performance of the GA-DC algorithm when different node distributions are used. In this paper, three common node distributions are used viz. uniform, normal and exponential. Most of the literature assumes that nodes are uniformly distributed in a 2D environment. The actual location of a node is specified by generating a uniform random number in the interval $[0,1]$ and then multiplying it by the network dimensions. Uniform distribution does not require any location information to determine a node position. All network locations are equally likely to be chosen as a potential node position. The need to define distribution parameters is a major issue when distributing network nodes according to other non-uniform distributions. Normal node distribution requires the sample mean and standard deviation, $(\hat{\mu}, \hat{\sigma})$, as the defining parameters to distribute the nodes accordingly. Exponential node distribution requires the sample rate parameter, $(\hat{\lambda})$, as the defining parameter to distribute the nodes accordingly. The maximum likelihood estimation (MLE) technique is used to estimate the defining parameters of the normal and exponential distributions using the uniform distribution's node locations. Appendix (A) describes the mathematical foundation of the MLE technique for computing the defining parameters of the normal and exponential node distributions.

Three dimensional (3D) networks are more realistic than 2D networks. The third dimension is important in determining the CHs to connect with. It is also important in calculating the energy expenditure budget during protocol round. It is shown that 3D networks life span is less than the life span of the 2D networks [4]. Fig. 5 depicts an example of the three node distributions on a 3D network. The BS is shown in the middle of the network.

Tab. 1 shows the parameter settings used in the simulation.

6 Simulation Results

Two sets of simulation scenarios are investigated in this section. The first scenario compares and evaluates the performance of the GA-DC algorithm with that of the well-known LEACH algorithm for clustering and routing in a 3D network with uniformly distributed nodes. Figs. 6 and 7 show the distribution of the number of CHs during network operation. A number of observations are noticed regarding these two figures. First, the range of the number of CHs for the LEACH algorithm is $[1\ 19]$, while the range in the GA-DC algorithm is $[3\ 14]$. The huge gap in the LEACH algorithm require the network to consume more energy for intra-cluster data communications especially when the number of CHs is one or two, and in inter-cluster data communications when the number of CHs is more than

fourteen. The percentage of protocol rounds that uses one, two and fourteen and more CHs in LEACH based network is around 19%. This huge percentage contributes to minimizing the life span of the network compared to the GA-DC based network. The optimal number of CHs used in LEACH and GA-DC algorithms during the network operation are 8 and 4 respectively. They contribute to 10.5% and 28.5% of the number of CHs used during the network life time.

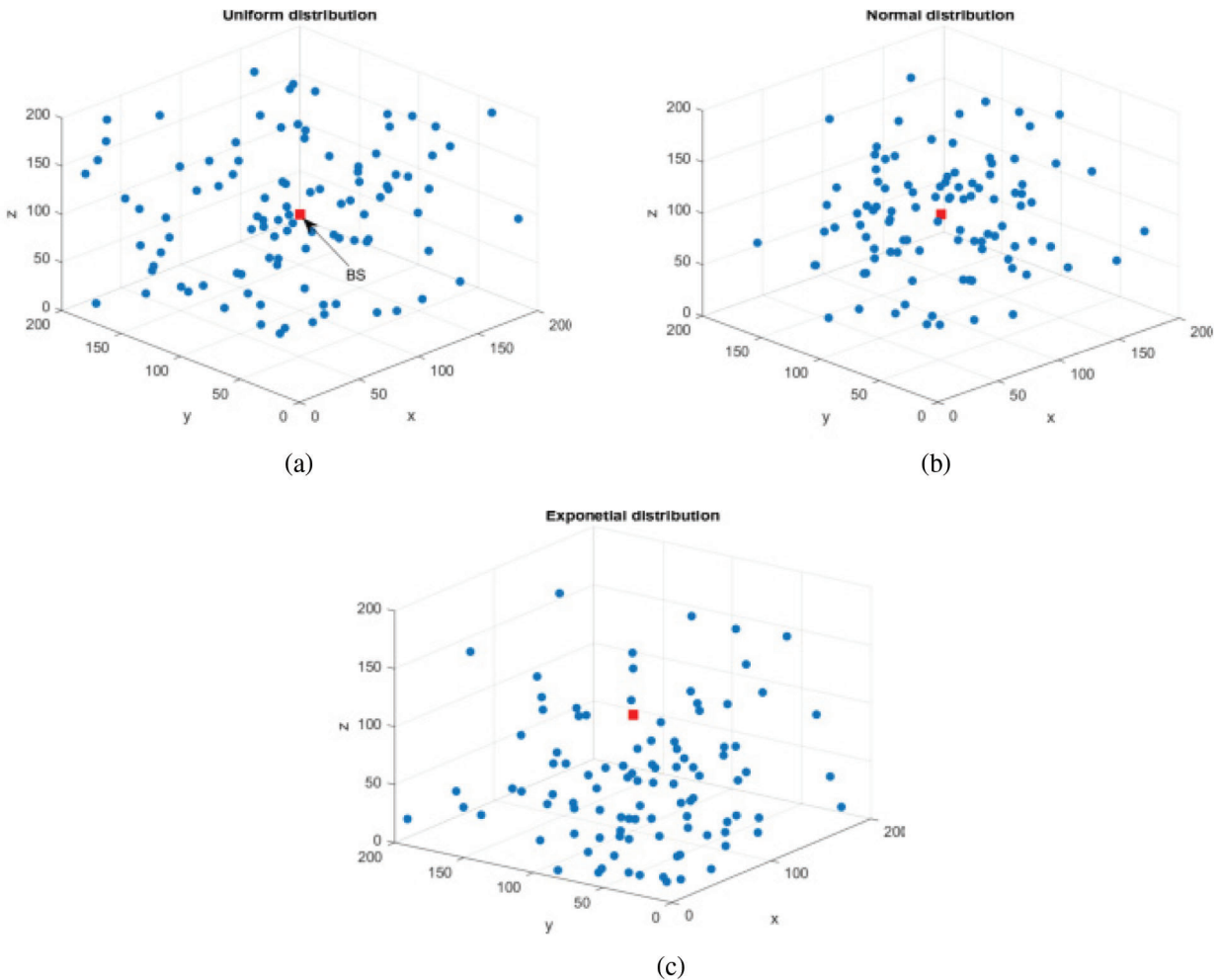


Figure 5: 3D WSN random distributions: (a) Uniform (b) Normal (c) Exponential

Fig. 8 depicts the total remaining energy for the two protocols. It clearly shows that the LEACH protocol performs better than the GA-DC protocol up to an equilibrium point (P_0) in which the GA-DC algorithm outperforms the LEACH algorithm. P_0 represents the point at which the remaining energy in the network for the two algorithms is the same. P_0 occurs at round number 738 in which 11.5% of the energy is available. GA-DC algorithm utilizes the remaining energy efficiently to extend the network life time.

Fig. 9 shows the packet throughput for the two clustering algorithms. The GA-DC algorithm has much higher throughput than the LEACH algorithm because it extends the network's life time compared to the LEACH protocol. When compared to the LEACH algorithm, the GA-DC algorithm increases throughput by 55%.

Table 1: Network parameters settings

Parameter	Value
N	100
Node energy	0.5 J
E_{elec}	50 nJ/bit
E_{DA}	5 pJ/bit
ϵ_{fs}	10 pJ/bit/m ²
ϵ_{mp}	0.0013 pJ/bit/m ⁴
d_0	87 m
k	4000 bits
Network volume	$200 \times 200 \times 200$
Crossover rate	1.0
Mutation rate	0.02
Rounds	2500

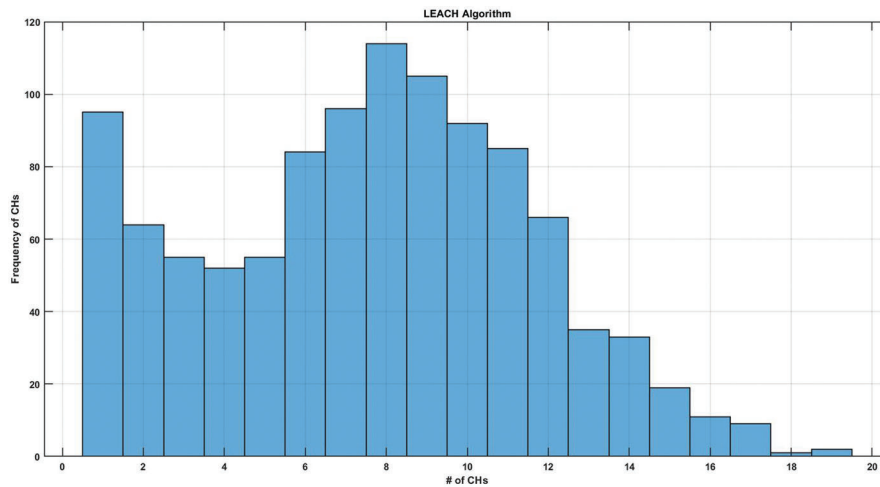
**Figure 6:** Distribution of CHs during network operation using LEACH algorithm

Fig. 10 shows the network life time (NLT) of the two algorithms. NLT is defined as the sum of the stability and instability periods of the clustering algorithm. It is shown that the stability period of the LEACH protocol is larger than the stability period of the GA-DC algorithm. This indicates that LEACH has more clustering reliability than the GA-DC algorithm. On the other and GA-DC algorithm has a better NLT because of the larger instability period compared to the LEACH algorithm. This shows that the network can still send data reports about the environment regardless the death of most of the network nodes. It is a trade-off between reliability and NLT. The stability periods of the LEACH and GA-DC algorithms comprise 38.3% and 12.7% of the NLT, respectively.

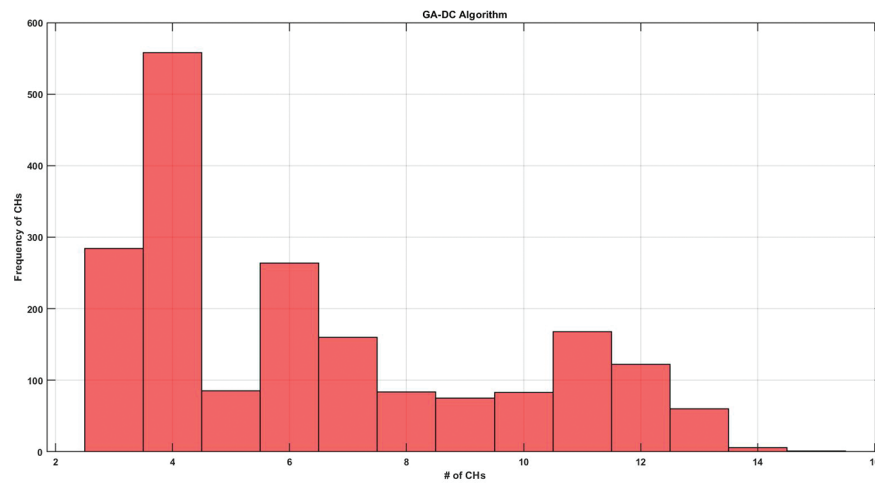


Figure 7: Distribution of CHs during network operation using GA-DC algorithm

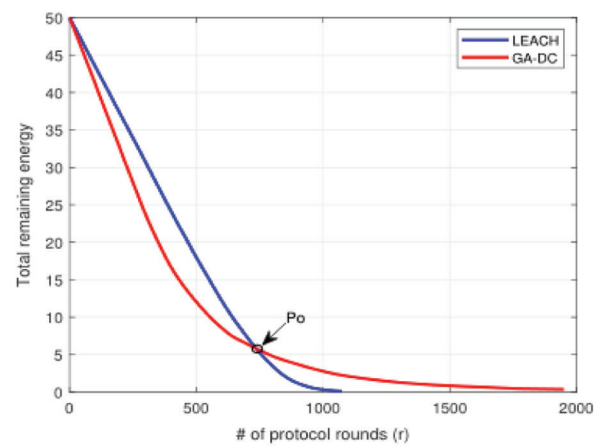


Figure 8: Total remaining energy per protocol round

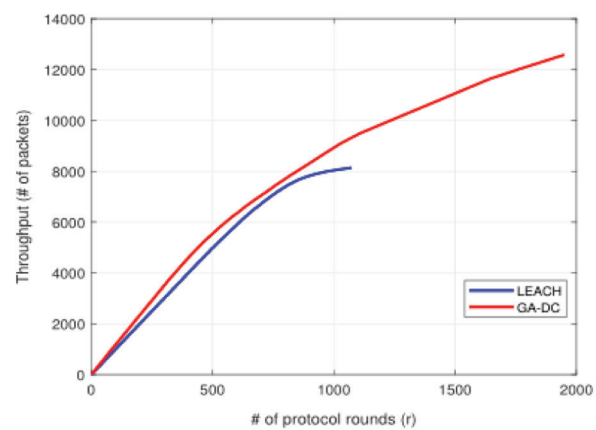


Figure 9: Total throughput packets per protocol round

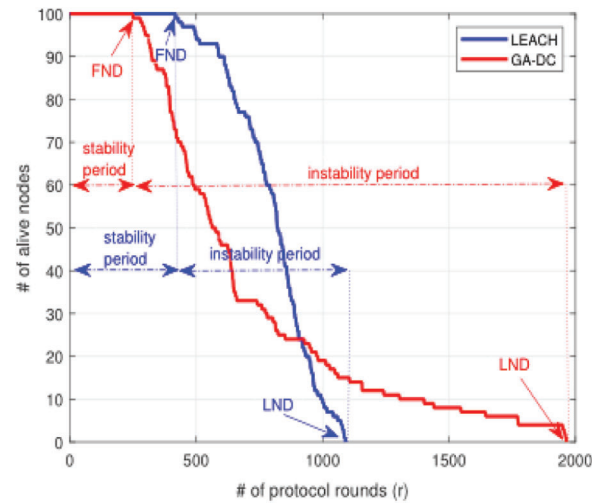


Figure 10: Network life time (NLT) of LEACH and GA-DC algorithms

The second scenario evaluates and compares the performance of GA-DC algorithm for three different node distributions. The goal of this scenario is to determine which node distribution is best in terms of energy consumption and extending the NLT. Fig. 11 shows the frequencies of the CHs per node distribution. It shows that the most used number of CHs in uniform and normal distributions is 4 which comprise 28.5% and 28% of the number of clusters formed respectively. The most used number of CHs in exponential distributions is 5 and it comprises 27.8% of the number of clusters formed. It is clearly shown that 4 is the best number of clusters used for extending the life time for both the uniform and normal distributions while 5 is best number of clusters used in exponential distribution.

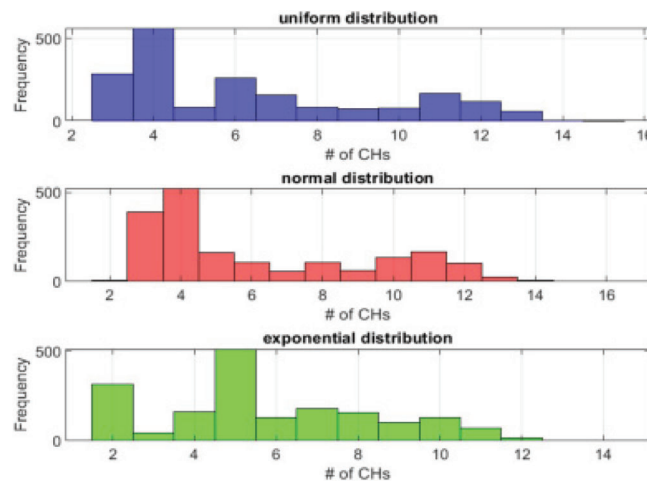


Figure 11: CHs frequencies for different node distributions

Fig. 12 shows that the uniform and normal node distributions consumed nearly the same amount of energy during the network's lifetime, with the uniform distribution consuming slightly less energy than the normal distribution. Exponential distribution consumes energy the most since it has less total remaining energy during the network life time.

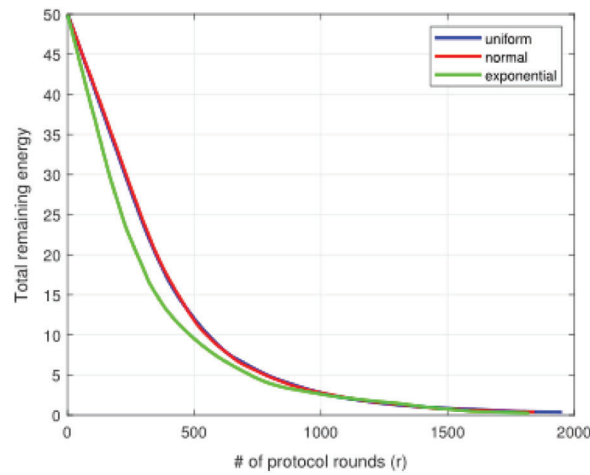


Figure 12: Total remaining energy per protocol round

Fig. 13 depicts that the exponential distribution has the least stability region which indicates a low reliability of the clustering process compared to the stability regions of the uniform and normal distributions. All three distributions favor increasing the NLT over clustering process reliability by extending the instability region. The stability regions end at round 142, 240 and 250 for the exponential, normal and uniform distributions respectively. This shows that GA-DC algorithm over uniform distribution is more reliable than the other distributions and this is important because when the node population begins to decline, the number of CHs per round becomes unstable (lower than intended), and there is no guarantee that an optimal number of CHs will be generated per round. Furthermore, because there are fewer alive nodes, the field is sampled (sensed) over fewer nodes than intended.

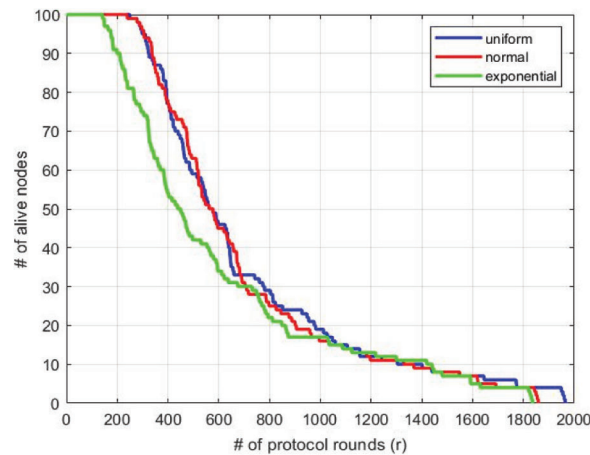


Figure 13: Network life time (NLT) for different node distribution

Fig. 14 shows the cumulative packet throughput reached the CHs by the MNs. During this phase, the CHs aggregate and compress the packets received from their corresponding MNs before transmitting them to the BS. When compared to the exponential distribution, the uniform and normal node distributions have the most packets transmitted from the MNs to the CHs, with the uniform distribution having the most packets because it manages its energy more efficiently.

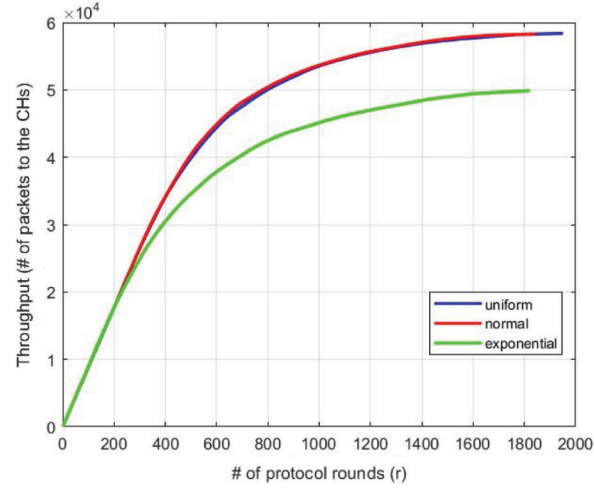


Figure 14: Cumulative throughput packets to the CHs per protocol round

Fig. 15 shows that the uniform distribution outperforms the other two distributions in terms of the total number of packet reports reached the BS. This is the useful throughput that gives information about the environment. The uniform distribution's high throughput performance is due to its ability to manage energy efficiently through the optimal selection of CHs over the network's lifetime.

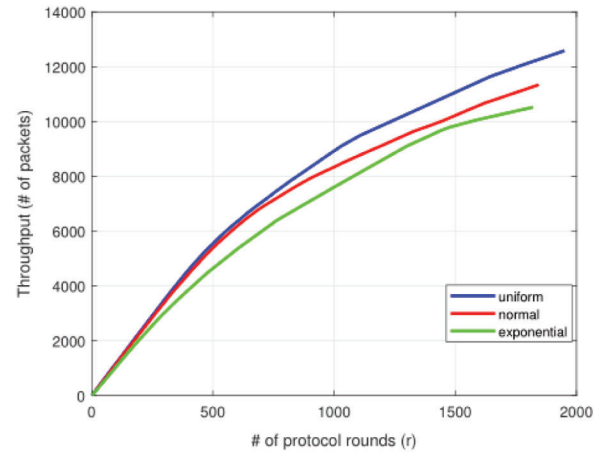


Figure 15: Cumulative throughput packets to the BS per protocol round

Tab. 2 summarizes the different network metrics for the three node distributions used in this work.

Tab. 3 shows the improvement ratio (γ) of utilizing the uniform distribution in GA-DC algorithm over the other two distributions. The improvement ratio (γ) is given by

$$\gamma = \frac{PI_u - PI_o}{PI_o} \quad (9)$$

where PI_u refers to any performance indicator (NLT, FND and throughput) under the uniform distribution, PI_o refers to any performance indicator under other distributions.

Table 2: Node distribution summary

Metric	Uniform	Normal	Exponential
NLT (rounds)	1965	1859	1836
FND	250	240	142
LND	1965	1859	1836
Throughput (packets)	12572	11341	10517
Frequent # of CHs	4	4	5

Table 3: Improvement ratio summary

Metric	Normal	Exponential
NLT (rounds)	5.7%	7%
FND	4.2%	76%
Throughput (packets)	10.85%	19.54%

7 Conclusion

In this work, the GA-DC algorithm is proposed for 3D network setting. The GA-DC algorithm is based on binary GA for distinguishing CHs from MNs. When compared to the state-of-the-art LEACH protocol, the GA-DC algorithm increased network life time by 80% and data throughput by 55%. The GA-DC algorithm demonstrated that the uniform distribution of network nodes in a 3D setting outperformed the normal and exponential distributions in terms of network life time, reliability, and network throughput. When compared to the normal distribution, the uniform distribution improved the NLT by 5.7%, reliability measure by 4.2% and throughput by 10.85%. Furthermore, when compared to the exponential distribution, the uniform distribution improved NLT by 7%, reliability measure by 76% and throughput by 19.54%. The GA-DC algorithm employs a new fitness function that implements various parameters to compute the cost value of the proposed solution. The parameters include information about the proposed solution's energy consumption, inter and intra clustering distances, and short and long range wireless link connections.

Data Availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding Statement: This research was supported by Al-Zaytoonah University of Jordan fund. Project Number 15/12/2019–2020.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. Masoud, Y. Jaradat, A. Manasrah and I. Jannoud, "Sensors of smart devices in the internet of everything (IOE) Era: Big opportunities and massive doubts," *Journal of Sensors*, vol. 2019, pp 1–26, 2019.
- [2] Y. Jaradat, H. Huang, M. Masoud and I. Janoud, "Capacity of wireless networks with directed energy links in the presence of obstacles," *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5273–5283, 2017.

- [3] Y. Jaradat, M. Masoud and I. Jannoud, "A mathematical framework of optimal number of clusters in 3D noise-prone WSN environment," *Institute of Electrical and Electronics Engineers (IEEE) Sensors Journal*, vol. 19, no. 6, pp. 2378–2388, 2018.
- [4] Y. Jaradat, M. Masoud and S. Al-Jazzar, "A comparative study of the effect of node distributions on 2D and 3D heterogeneous WSN," *International Journal of Sensor Networks*, vol. 33, no. 4, pp. 202–210, 2020.
- [5] A. Somauroo and V. Bassoo, "Energy-efficient genetic algorithm variants of PEGASIS for 3D wireless sensor networks," *Applied Computing and Informatics*, vol. ahead-of-print, no. ahead-of-print, vol. 16, no. 2, pp. 1–23, 2020.
- [6] M. Z. Masoud, Y. Jaradat, D. Zaidan and I. Jannoud, "To cluster or not to cluster: A hybrid clustering protocol for WSN," in *2019 IEEE Jordan Int. Joint Conf. on Electrical Engineering and Information Technology (JEEIT)*, IEEE, Amman, Jordan, pp. 678–682, 2019.
- [7] M. Z. Masoud, Y. Jaradat, I. Jannoud and M. A. Al Sibahee, "A hybrid clustering routing protocol based on machine learning and graph theory for energy conservation and hole detection in wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 15, no. 6, pp. 300–319, 2019.
- [8] R. Zagrouba and A. Kardi, "Comparative study of energy efficient routing techniques in wireless sensor networks," *Information*, vol. 12, no. 1, pp. 42, 2021.
- [9] H. Yetgin, K. T. K. Cheung, M. El-Hajjar and L. Hanzo, "Network-lifetime maximization of wireless sensor networks," *Institute of Electrical and Electronics Engineers (IEEE) Access*, vol. 3, pp. 2191–2226, 2015.
- [10] S. K. Singh, P. Kumar and J. P. Singh, "A survey on successors of LEACH protocol," *Institute of Electrical and Electronics Engineers (IEEE) Access*, vol. 5, pp. 4298–4328, 2017.
- [11] A. Singh, S. Sharma and J. Singh, "Nature-inspired algorithms for wireless sensor networks: A comprehensive survey," *Computer Science Review*, vol. 39, pp. 100342, 2021.
- [12] B. Baranidharan and B. Santhi, "GAECH: Genetic algorithm based energy efficient clustering hierarchy in wireless sensor networks," *Journal of Sensors*, vol. 2015, 2015.
- [13] S. P. Singh and S. Sharma, "Genetic-algorithm-based energy-efficient clustering (GAEEC) for homogenous wireless sensor networks," *Institute of Electronics and Telecommunication Engineers (IETE) Journal of Research*, vol. 64, no. 5, pp. 648–659, 2018.
- [14] 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, pp. 105788, 2019.
- [15] A. Rezaeipannah, H. Nazari and G. Ahmadi, "A hybrid approach for prolonging lifetime of wireless sensor networks using genetic algorithm and online clustering," *Journal of Computing Science and Engineering*, vol. 13, no. 4, pp. 163–174, 2019.
- [16] J. Bhola, S. Soni and G. K. Cheema, "Genetic algorithm based optimized LEACH protocol for energy efficient wireless sensor networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 1281–1288, 2020.
- [17] P. Joshi, S. Gavel and A. S. Raghuvanshi, "Estimating energy consumption for various sensor node distributions in wireless sensor networks," in *Nanoelectronics, Circuits and Communication Systems*, vol. 692, Springer, Switzerland, pp. 289–299, 2021.
- [18] S. Bayraklı, and S. Z. Erdogan. "Genetic algorithm based energy efficient clusters (GABEEC) in wireless sensor networks." *Procedia Computer Science*, vol. 10, pp. 247–254, 2012.
- [19] Y. Xu, W. Jiao and M. Tian. "An energy-efficient routing protocol for 3D wireless sensor networks." *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19550–19559, 2021.
- [20] S. Katoch, S. S. Chauhan and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2020.
- [21] M. T. Ahvanooey, Q. Li, M. Wu and S. Wang, "A survey of genetic programming and its applications," *Transactions on Internet and Information Systems*, vol. 13, no. 4, pp. 1765–1794, 2019.
- [22] Y. Jaradat, M. Masoud, S. Al-Jazzar and M. Alia, "Optimal network dimensions for energy conservation in clustered 3D WSN," *Wireless Networks*, vol. 27, no. 3, pp. 1821–1833, 2021.
- [23] A. Alkhatib and Q. Abed-Al. "Multivariate outlier detection for forest fire data aggregation accuracy." *Intelligent Automation and Soft Computing*, vol. 31, no. 2, pp. 1071–1087, 2022.

Appendix A.**Maximum Likelihood Estimation (MLE)**

MLE technique computes the defining parameter(s) of a random distribution by maximizing the likelihood function, L , of the samples of interest. $L(\theta; x_1, \dots, x_n)$ of n samples is defined by:

$$L(\theta; x_1, \dots, x_n) = f_X(x_1, \dots, x_n; \theta) \quad (10)$$

where θ is the parameter(s) of the distribution of interest and $f_X(x_1, \dots, x_n; \theta)$ is the joint probability function from which the samples are drawn. Assuming that the random samples are independent, identically distributed (IID) random variables, then L is simplified to:

$$L(\theta) = L(\theta; x_1, \dots, x_n) = f_X(x_1) \times \dots \times f_X(x_n) \quad (11)$$

To find the value of θ that maximize the L function, the derivative of the L function is computed and set to zero.

$$\frac{d}{d\theta} L(\theta) = 0 \quad (12)$$

It is usually easier to work with $\ln[L(\theta)]$ since the maxima of the $L(\theta)$ and $\ln[L(\theta)]$ occur at the same θ value especially when working with distributions that utilize an exponential function in their distribution.

A. Estimating Normal Distribution Parameters

For n independent random samples x_1, \dots, x_n , the L function is given by:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \quad (13)$$

Taking the natural logarithm of both sides of [Eq. \(13\)](#).

$$\begin{aligned} \ln[L(\theta)] &= \ln\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}\right] \\ &= -\frac{n}{2} \ln[2\pi] - \frac{n}{2} \ln[\sigma^2] - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned} \quad (14)$$

To find the estimators for μ and σ that maximize the L function.

$$\frac{\partial}{\partial \mu} \ln[\sigma^2] = 0 \quad (15)$$

$$\frac{\partial}{\partial \sigma} \ln[\sigma^2] = 0 \quad (16)$$

Solving [Eqs. \(15\)](#) and [\(16\)](#) for μ and σ we get:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (17)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (18)$$

B. Estimating Exponential Distribution Parameter

For n independent random samples x_1, \dots, x_n , the L function is given by:

$$L(\theta) = \prod_{i=1}^n \lambda \exp\{-\lambda x_i\} = \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\} \quad (19)$$

Taking the natural logarithm of both sides of [Eq. \(19\)](#).

$$\ln[L(\theta)] = n \ln[\lambda] - \lambda \sum_{i=1}^n x_i \quad (20)$$

To find the estimator for λ that maximize the L function we solve the following.

$$\frac{d}{d\lambda} \ln[L(\theta)] = 0 \quad (21)$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} \quad (22)$$