
Vincent Chung
Modelling, Simulation & Computing Laboratory
Faculty of Engineering, Universiti Malaysia Sabah
Kota Kinabalu, Malaysia.
vincentchung@ieee.org, msclab@ums.edu.my

Norah Tuah
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Kota Kinabalu, Malaysia.
norahntuah@uitm.edu.my

Min Keng Tan
Modelling, Simulation & Computing Laboratory
Faculty of Engineering, Universiti Malaysia Sabah
Kota Kinabalu, Malaysia.
mktan@ieee.org, msclab@ums.edu.my

Ismail Saad
Faculty of Engineering
Universiti Malaysia Sabah
Kota Kinabalu, Malaysia.
ismail_s@ums.edu.my

Kit Guan Lim
Modelling, Simulation & Computing Laboratory
Faculty of Engineering, Universiti Malaysia Sabah
Kota Kinabalu, Malaysia.
kittguanlim.my@ieee.org, msclab@ums.edu.my

Abstract—Energy-efficient optimization algorithm in wireless sensor network (WSN) is often based on solely cluster routing or multi-hop routing. The cluster optimization algorithm will form a cluster network by dividing the sensor nodes into few clusters where each cluster has a cluster head (CH) for data collection. On the other hand, multi-hop optimization algorithm will form a multi-hop network by transmitting data to base station (BS) through data multi-hopping between sensor nodes. However, cluster optimization algorithm suffers from the overburdens of CH nodes, while multi-hop optimization algorithm suffers from the overburdens of nodes which are near to the BS. Therefore, Genetic Algorithm-Cuckoo Search (GACS) is proposed and developed based on the multi-hop clustering model in this paper. GACS optimizes both intra-cluster and inter-cluster communications to enhance energy efficiency in WSN, extending the network lifetime. Based on the performance evaluation, GACS outperforms both Genetic Algorithm (GA)-based cluster optimization algorithm and Cuckoo Search (CS)-based multi-hop optimization algorithm.

Keywords—genetic algorithm, cuckoo search, multi-hop clustering, wireless sensor network, metaheuristic mechanism

I. INTRODUCTION

Wireless sensor network (WSN) has attracted much attention in recent years for its application in various areas such as precision agriculture, military application and intelligent buildings [1]. Each sensor node is a portable data logger to store local data and transmit the collected data to BS. These sensor nodes will be deployed in the region of interest for long-term continuous monitoring. Thus, a sustainable power supply is important where the usage of energy should be optimized [2], leading to a longer network lifetime. Hence, it is obvious that an energy-efficient optimization algorithm is necessary.

As one of the effective metaheuristic optimization algorithms, evolutionary algorithm is commonly used in WSN to optimize the energy consumption. It is designed to provide a sufficiently good solution to an optimization problem, especially with incomplete information and limited computation capacity through bio-inspired computing [3]. Evolutionary algorithm samples a set of solutions which is too large to be completely sampled. Thus, it could save a lot of computation time. In WSN optimization problem, evolutionary algorithm is developed based on the routing protocol to achieve the optimization of energy consumption.

Routing protocol specifies the method used for WSN to communicate with base station (BS). There are many types of routing protocols that can be classified based on three different points of view which are path discovery, participation style and network structure [4]. However, the most applied routing protocols in the formulation of evolutionary algorithms are single-hop transmission, cluster routing and multi-hop routing. Evolutionary algorithms based on different routing protocol will impose different effects on the optimization of energy consumption in WSN.

Single-hop transmission is the most primitive routing for data transmission, but it is not energy-efficient due to the significantly long transmission distance. Later, cluster and multi-hop routing protocols are introduced to counter with the deficiency of single-hop transmission. Basically, sensor nodes are grouped into few clusters in cluster routing protocol. By only allowing cluster heads (CHs) to directly transmit aggregated data to BS, it reduces the total transmission distance. On the other hand, multi-hop routing protocol controls the data transmissions from sensor nodes to BS, where transmission is achieved through several hops, so the total transmission distance is reduced too.

Nevertheless, cluster and multi-hop routing protocols still have a major limitation which is the overburdens of specific nodes though they do outperform single-hop transmission due to shorter total transmission distance. CHs with cluster routing protocol and the nodes near to BS with multi-hop routing protocol are overburdened due to the routing of a large number of data messages to BS [5, 6]. Thus, evolutionary algorithms such as Genetic Algorithm (GA)-based cluster optimization algorithm and Cuckoo Search (CS)-based multi-hop optimization algorithm will inherit such limitation. Hence, Genetic Algorithm-Cuckoo Search (GACS) is proposed and developed based on the multi-hop clustering model in this paper to further optimize the energy consumption in WSN.

This paper is organized as follow: Section II introduces the elements in GACS which are genetic operators (selection, crossover, mutation, replacement) and Lévy flight. Section III explains the formulation of multi-objective fitness function. Section IV presents a network model as the testbed for robustness evaluation. Section V discusses the performance of GACS in terms of network lifetime, residual energy, energy consumption, CH number and transmission failure through the comparative analysis with GA-based cluster optimization algorithm and CS-based multi-hop optimization algorithm. Section VI summarizes the findings of this work and provides a recommendation for future study.

* corresponding author

This work is supported by UMS SDK, grant no. DN20086, UMSSweat, grant no. GUG0450-1/2020 and UMS Postgraduate Assistant Scheme (SBP).
II. GENETIC ALGORITHM-CUCKOO SEARCH

GACS is a multi-hop clustering optimization algorithm that is inspired by the drawbacks of both cluster and multi-hop optimization algorithms. In GACS, multi-hop routing is applied between CHs to reduce their total transmission distance. Multi-hop routing between CHs will not cause the overburdens of nodes which are near to the BS because CH selection is adaptive and selected CHs do not receive and aggregate a lot of data due to the low CH number. Therefore, workload is evenly distributed. In short, GACS optimizes both intra-cluster and inter-cluster communications via genetic operators and Lévy flight respectively.

A. Selection

Selection is the operator for parent selection to produce next generation. Roulette wheel selection (RWS) has been known for its simplest selection scheme in which the chromosomes are mapped to the contiguous segments that resemble a roulette [7]. The size of each segment is associated to their fitness. Thus, the probability of each chromosome to be selected from the roulette wheel can be expressed by (1).

\[
\text{Probability} = \frac{\text{Fitness Value of Chromosome}}{\text{Total Fitness Value of Population}} \tag{1}
\]

B. Crossover

Crossover is the operator that is analogous to biological crossover. One-point crossover is the most common crossover technique, in which a random crossover point will be selected to decide the gene position for swapping as illustrated in Fig. 1 [8]. Through the crossover at this gene position, two offspring will be generated. However, this crossover between two parents occurs only with a pre-defined crossover probability. If no crossover occurs, the offspring will inherit encoded information of their parents.

![Image](47x308 to 287x353)

Fig. 1. One-point crossover.

C. Mutation

Mutation is the operator used to maintain diversity in the population. If the mutation probability is too low, premature convergence may occur. Nevertheless, mutation probability that is too high may also leads to the difficulty of convergence. Hence, an optimum mutation probability is required. One of the popular mutation techniques is multi-fit flip mutation, in which random number of genes are selected for mutation [9]. The selected genes will be bit flipped with a pre-defined mutation probability.

D. Replacement

Replacement is the operator used to form a new population. Weak parent replacement as a steady-state replacement technique is applied to replace the parents with fitter offspring [10]. If the parents are fitter, the offspring will be eliminated. This is to ensure the next descendants to be better as the number of generations increases.

E. Lévy Flight

Lévy flight is a random walk performed to generate new possible multi-hop data transmission path from CHs to BS, in which each egg represents a multi-hop data transmission path. It is a modification of the standard random walk which uses a random step length as well as a random direction. The step length of this random walk is drawn from the Lévy distribution. Generally, it is often defined in terms of a simple power-law formula expressed by (2) that has an infinite variance and mean [11]. Mathematically speaking, a simple version of Lévy distribution can also be defined as shown by (3).

\[
L(s) \sim |s|^{-1-\beta} \tag{2}
\]

\[
L(s, \gamma, \mu) = \begin{cases} \frac{1}{2\pi (s-\mu)^{3/2}} \exp\left(-\frac{\gamma}{2(s-\mu)}\right), & 0 < \mu < s < \infty; \\ 0, & \text{otherwise}. \end{cases} \tag{3}
\]

where \(s\) is the step length; \(\beta\) is an index with its value between 0 and 2, \(\mu\) is the minimum step length and \(\gamma\) is the scale parameter. Clearly, (4) can be obtained as \(s \to \infty\).

\[
L(s, \gamma, \mu) \approx \frac{\gamma}{2\pi s^{3/2}} \tag{4}
\]

This is a special case of the generalized Lévy distribution.

The easiest method to calculate step length is by applying Mantegna algorithm for a symmetric Lévy distribution as expressed by (5) [12]. The term symmetric means the step length can be either positive or negative.

\[
s = \frac{u}{|v|^{1/\beta}} \tag{5}
\]

where \(u\) and \(v\) are drawn from the normal distributions shown by (6) with their standard deviations denoted by (7).

\[
u \sim N(0, \sigma_v^2), \quad v \sim N(0, \sigma_v^2) \tag{6}
\]

\[
\sigma_u = \left(\frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma(1+\beta/2)\beta 2^{\beta-1/2}}\right)^{1/\beta}, \quad \sigma_v = 1 \tag{7}
\]

Using the step length calculated based on Mantegna algorithm, new path can be generated as expressed by (7).

\[
x_i^{(t+1)} = x_i^{(t)} + (\alpha \times s) \tag{7}
\]

where \(x_i^{(t)}\) is the current path and \(\alpha\) is the step length adjustment parameter.

Lévy flight is a random walk that delivers superior performance than the Brownian random walk in exploring unknown and large-scale search space. One of the proofs is that its variance shown by (8) increases much faster than the linear relationship of Brownian random walk shown by (9).

\[
\hat{h}_U^2 \rightarrow \infty \quad \text{as} \quad h \rightarrow \infty \tag{8}
\]

\[
\hat{h}_B^2 \rightarrow \infty \quad \text{as} \quad h \rightarrow \infty \tag{9}
\]
\[ \sigma^2(t) \sim t^{3-\beta} \quad (8) \]

\[ \sigma^2(t) \sim t \quad (9) \]

where \( \beta \) is an index with its value ranges from 1 to 2.

F. Algorithm

Since GACS is a multi-hop clustering optimization algorithm, it optimizes both intra-cluster and inter-cluster communications. In intra-cluster communication, GACS will optimize the CH selection. A population of possible CH selections are initialized by integrating the prior knowledge based on the improved low-energy adaptive clustering hierarchy (LEACH) routing protocol and the condition of average residual energy. Then, the selection, crossover and mutation operators are used to generate new possible CH selections. The repetitive process throughout the generations will evolve the CH selection. Hence, the intra-cluster communication will be optimized before subjecting into the optimization of inter-cluster communication. In inter-cluster communication, GACS will optimize the multi-hop data transmission path discovery from CHs to BS. The obligation brood parasitism of some cuckoo species is applied for the path discovery. In this process, the Lévy flight is used to generate new possible paths. It is a random walk with great performance in exploring unknown and large-scale search space. This is because any large step is possible with the heavy-tailed Lévy distribution. At last, optimum multi-hop data transmission path will be discovered in the form of multi-chain. For clearer description, the pseudo code of GACS is given in Table 1.

<table>
<thead>
<tr>
<th>Genetic Algorithm–Cuckoo Search Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initialize the base station</td>
</tr>
<tr>
<td>2. for number of rounds until maximum number of rounds</td>
</tr>
<tr>
<td>3. Check and remove the dead sensor nodes</td>
</tr>
<tr>
<td>4. Initialize the population of possible cluster head selections</td>
</tr>
<tr>
<td>5. while (( t &lt; ) maximum number of generations) or (stop criterion)</td>
</tr>
<tr>
<td>6. Optimize the cluster head selection</td>
</tr>
<tr>
<td>7. end while</td>
</tr>
<tr>
<td>8. Evaluate the selections in population and find the best selection</td>
</tr>
<tr>
<td>9. Assign sensor nodes as cluster head</td>
</tr>
<tr>
<td>10. Form the clusters in network</td>
</tr>
<tr>
<td>11. Collect the data within every cluster</td>
</tr>
<tr>
<td>12. Initialize the population of possible paths</td>
</tr>
<tr>
<td>13. while (( t &lt; ) maximum number of generations) or (stop criterion)</td>
</tr>
<tr>
<td>14. Optimize the multi-hop data transmission path discovery</td>
</tr>
<tr>
<td>15. end while</td>
</tr>
<tr>
<td>16. Evaluate the multi-chain paths in population and find the best path</td>
</tr>
<tr>
<td>17. Form the multi-chain paths from cluster heads to base station</td>
</tr>
<tr>
<td>18. Collect the multi-chain paths to base station</td>
</tr>
<tr>
<td>19. Collect the data to base station</td>
</tr>
<tr>
<td>20. end for</td>
</tr>
</tbody>
</table>

III. MULTI-OBJECTIVE FITNESS FUNCTION

In evolutionary algorithm, fitness function is one of the essential elements as it is used to evaluate the fitness of the possible solution [13]. In GACS, the implemented objective functions are multi-hop clustering-based distance (MCD), transfer energy (E) and transmission status (TS). They will then be formulated into a single multi-objective fitness function through weighted-sum method.

A. Multi-Hop Clustering-Based Distance

Longer transmission distance commonly leads to higher energy consumption and shorter network lifetime[14]. Hence MCD is desired to be low and its calculation is given by (10).

\[ A = \sum_{i=1}^{n} \left( \left( \sum_{j=1}^{m} d_{ij} \right) + D_{is} \right) \quad (10) \]

where \( A \) is the objective value of MCD, \( n \) is the number of clusters, \( m \) is the number of associated nodes in the cluster, \( d_{ij} \) is the distance between a node and its CH and \( D_{is} \) is the distance between CH and its target.

B. Transfer Energy

This objective function stands for an amount of energy consumption required to transmit all collected data to the BS, so it is desired to be low [15]. It involves two elements that are differentiated into energy consumption of CH and energy consumption of member node as shown by (11).

\[ B = \sum_{i=1}^{n} \left( \left( \sum_{j=1}^{m} E_{jm} \right) + E_{i} \right) \quad (11) \]

where \( B \) is the objective value of E, \( n \) is the number of clusters, \( m \) is the number of associated nodes in the cluster, \( E_{jm} \) is the energy required to transfer data from a node to the corresponding CH and \( E_{i} \) is the required energy for transmission from CH to its target.

C. Transmission Status

In this work, transmission failure is due to the insufficient residual energy of sensor node in accomplishing the given tasks. Therefore, fitness evaluation based on (12) can reduce the number of transmission failures.

\[ C = \sum_{i=1}^{k} (E_{ir} - E_{ic}) > 0 \quad (12) \]

where \( C \) is the objective value of TS, \( k \) is the number of alive sensor nodes, \( E_{ir} \) is the residual energy and \( E_{ic} \) is the required energy for transmission.

D. Fitness Function

The multi-objective fitness function as shown by (13).

\[ f = w_1 \left( \frac{1}{A} \right) + w_2 \left( \frac{1}{B} \right) + w_3 (C) \quad (13) \]

where \( f \) is the fitness value whereas \( w_1,w_2 \) and \( w_3 \) are the weightages of MCD, E and TS respectively, in which the summation of these weightages is unity. The weightages are pre-defined depending upon the high-level information about the problem.
IV. NETWORK MODEL

Energy usage in wireless communication is far higher than sensing and computing activities. The tasks for radio module are transmitting data, listening at idle state and receiving data. Therefore, energy consumption in WSN can be calculated by focusing on the radio model of sensor node as illustrated in Fig. 2 [16].

Energy cost to transmit $ke$ bit data from the transmitter module of one node to the receiver module of another node over a $de$ distance is expended by (14) [17]. This energy cost includes the extra energy consumptions due to the path loss and data aggregation. In the aspect of path loss, free space propagation will be used for shorter transmission range while the multi-path propagation will be used for longer transmission range. The threshold distance used to swap the amplification models can be calculated using (15) [17]. In aspect of data aggregation, this process is performed by CHs when collecting data from the cluster members and multi-hopping data to the BS.

$$E_{Tx}(ke, de) = \begin{cases} ke(E_{elec} + \epsilon_{fs}de^2) + ke_{DA}E_{DA} & \text{if } de < d_0; \\ ke(E_{elec} + \epsilon_{mp}de^4) + ke_{DA}E_{DA} & \text{if } de \geq d_0 \end{cases}$$ (14)

$$d_0 = \sqrt{\epsilon_{fs}/\epsilon_{mp}}$$ (15)

where $E_{Tx}(ke, de)$ is the overall transmission energy, $E_{elec}$ is the electronics energy required by the circuitry to transmit and receive 1-bit data, $\epsilon_{fs}$ is the free space propagation constant, $\epsilon_{mp}$ is the multi-path propagation constant, $ke_{DA}$ is the number of bits subjected to data aggregation and $E_{DA}$ is the energy consumption required to aggregate, compress 1-bit data and $d_0$ is the threshold distance.

On the other hand, energy cost to receive $ke$ bits of data message by the receiver module is denoted by (16) [18].

$$E_{Rx}(ke) = ke \times E_{elec}$$ (16)

where $E_{Rx}(ke)$ is the overall reception energy.

Table IV summarizes the parameter settings for the simulation of the network model [19].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS Coordinate</td>
<td>$x = -80$ m, $y = -80$ m</td>
</tr>
<tr>
<td>Sensor Nodes Number</td>
<td>200</td>
</tr>
<tr>
<td>Network Dimension</td>
<td>$100$ m $\times$ 100 m</td>
</tr>
<tr>
<td>Bit Size, $ke$</td>
<td>1000 bits</td>
</tr>
<tr>
<td>Initial Energy</td>
<td>0.05 J</td>
</tr>
<tr>
<td>Electronics Energy, $E_{elec}$</td>
<td>50 nJ bit$^{-1}$</td>
</tr>
<tr>
<td>Transmitter Amplifier (Free Space), $\epsilon_{fs}$</td>
<td>10 pJ bit$^{-1}$ m$^{-2}$</td>
</tr>
<tr>
<td>Transmitter Amplifier (Multi-Path), $\epsilon_{mp}$</td>
<td>0.0013 pJ bit$^{-1}$ m$^{-4}$</td>
</tr>
<tr>
<td>Energy of Data Aggregation, $E_{DA}$</td>
<td>5 nJ bit$^{-1}$</td>
</tr>
</tbody>
</table>

V. RESULTS AND DISCUSSION

The performance of GACS is evaluated against GA-based cluster optimization algorithm and CS-based multi-hop optimization algorithm. Network lifetime is the main performance metric while the other four metrics are used for more deeper discussion.

A. Network Lifetime

Network lifetimes of WSN with GACS, GA and CS are illustrated in Fig. 3.

GACS prolongs the round number for first node dies (FND) of GA and CS by 42.95 % and 77.88 % respectively. FND is the metric to measure network lifetime as WSN requires all sensor nodes to be cooperative for objective achievement. GACS outperforms GA due to the multi-hop optimization between CHs. It leads to a shorter total transmission distance and avoids the overburdens of CHs located far from BS. On the other hand, overburdens of nodes located near to BS with CS lead to the shortest network lifetime. These nodes need to continuously aggregate a lot of data and transmit to BS. This is avoided in GACS because CH selection is adaptive and the CHs receive and aggregate lesser data due to the low CH number.
B. Residual Energy

The residual energy of WSN with GACS, GA and CS is shown in Fig. 4.

![Residual Energy](image)

In the correlated stability period (68 rounds), the residual energy of network with GACS and CS are the highest and the lowest respectively. This is because the energy efficiencies in the network are highest and lowest with GACS and CS respectively. GACS leads to the least energy consumption as it collects the data with identical size of RF data and transmits to the BS. Residual energy of network with GA exceeds GACS after 311th round implies that the residual energy of some sensor nodes differ greatly from the average residual energy. At 311th round, all sensor nodes with GACS have low residual energy with small standard deviation. This is proven by the short instability network, in which the number of alive sensor nodes with GACS drops drastically after the round number for FND, which is 312th round.

C. Energy Consumption

The energy consumptions of WSN with GACS, GA and CS are exhibited in Fig. 5.

![Energy Consumption](image)

Energy consumption with GACS has the smallest average and standard deviation in the correlated stability period. Besides, GA protocol shows drastic rise at the end of each cycle while CS shows the highest average in the correlated stability period. GA experiences the drastic rise of energy consumption due to the drastic rise of CH number. The CHs in GA also consume more energy in overall due to the heavy workload. By contrast, the overall energy consumption of CHs with GACS is lower than GA due to the implementation of multi-hop optimization between CHs. On the other hand, average energy consumption in the correlated stability period with CS is the highest as most of the sensor nodes need to continuously receive and aggregate a lot of data.

D. Cluster Head Number

CH numbers in GACS and GA are illustrated in Fig. 6. CS is not shown because it is a multi-hop optimization algorithm that does not form clusters.

![Cluster Head Number](image)

GACS shows more balanced CH with smaller standard deviation than GA. Drastic rise of CH number at the end of each cycle is larger with GA. As both populations of GACS and GA for CH selection are initialized based on CH selection mechanism of the improved low-energy adaptive clustering hierarchy (LEACH) routing protocol with a desired CH percentage over total number of sensor nodes of 5%, the expected CH number is 10 sensor nodes. However, CH numbers with both GACS and GA before drastic rises are lower than the expected CH number due to the evolutionary operations. This causes the number of sensor nodes without experience as CH increased at the end of each cycle. Thus, effect of increasing probability of each sensor node to be selected as CH is not well countered. As a consequence, CH number rises drastically. This also causes the energy consumption with GA to rise accordingly. Nevertheless, CH number with GACS is more balanced due to the condition of average residual energy.

E. Transmission Failure

In this work, transmission failure is due to the insufficient residual energy of sensor node in accomplishing the given tasks. The transmission failures in WSN with GACS, GA, CS are shown in Fig. 7.
GACS reduces the number of transmission failures of GA and CS by 78.70% and 73.51% respectively. 92.67% of the transmission failures with GA and CS are due to high energy consumption for a single transmission. By implementing GACS, CS located far from the BS can avoid significantly long transmission distance and they receive and aggregate lesser data as the CH number in each round is not high.

VI. CONCLUSION

This paper improves the optimization of energy consumption in WSN by hybridizing GA and CS based on multi-hop clustering model. Many researches are carried out to optimize the network lifework of WSN by implementing evolutionary algorithms based on solely cluster routing or multi-hop routing. However, such evolutionary algorithms suffer from the overburdens of specific sensor nodes. Therefore, this work is carried out to cope with this limitation by implementing GACS. Based on the performance evaluation, it is also proved that GACS outperforms both GA-based cluster optimization algorithm and CS-based multi-hop optimization algorithm. Nevertheless, the effect of genetic operator combination is yet to be investigated.

ACKNOWLEDGMENT

The authors would like to acknowledge Universiti Malaysia Sabah (UMS) for supporting this research under Niche Research Fund Scheme (SDN), grant no. DN20086, Postgraduate Research Grant Scheme (UMSGreat), grant no. GUG0450-1/2020 and Postgraduate Assistance Scheme (SBP).

REFERENCES


![Fig. 7. Transmission failure.](image-url)