Abstract—This study proposes a genetic algorithm-based (GA-based) adaptive clustering protocol with an optimal probability prediction to achieve good performance in terms of lifetime of network in wireless sensor networks. The proposed GA-based protocol is based on LEACH, called LEACH-GA herein, which basically has set-up and steady-state phases for each round in the protocol and an additional preparation phase before the beginning of the first round. In the period of preparation phase, all nodes initially perform cluster head selection process and then send their messages with statuses of being a candidate cluster head or not, node IDs, and geographical positions to the base station. As the base station received the messages from all nodes, it then searches for an optimal probability of nodes being cluster heads via a genetic algorithm by minimizing the total energy consumption required for completing one round in the sensor field. Thereafter, the base station broadcasts an advertisement message with the optimal value of probability to all nodes in order to form clusters in the following set-up phase. The preparation phase is performed only once before the set-up phase of the first round. The processes of following set-up and steady-state phases in every round are the same as LEACH. Simulation results show that the proposed genetic-algorithm-based adaptive clustering protocol effectively produces optimal energy consumption for the wireless sensor networks, and resulting in an extension of lifetime for the network.

Index Terms—Adaptive clustering protocol, clustering head, genetic algorithm, optimal probability, lifetime.

I. INTRODUCTION

Wireless sensor networks (WSNs), which consist of a number of small battery-powered devices, are frequently to obtain various sorts of useful data from surroundings. These devices sense physical properties, such as sound, humidity, pressure, luminosity, temperature, or chemical concentration, and transmit the gathered data to a base station (BS) for further analysis and processing. WSNs have been effectively deployed in tactical combat situations, habitat monitoring, home security, and so on [1-5]. Since WSNs consist of many sensors with limited energy, an energy-efficient network protocol is an important consideration in WSN applications. Many routing protocols for WSNs have appeared in the literature. In applications using direct transmission (DT) protocols [6], sensor nodes transmit their sensed data directly to a BS. Thus, the nodes located far from the BS will die quickly since they dissipate much energy in transmitting data packets. DT protocols are inefficient since energy levels of nodes are drained rapidly when the BS is located far. On the other hand, minimum transmission energy (MTE) protocols [7, 8] transmit data packets to the BS by way of multi-hop relay. As a result, nodes located near the BS die quickly since they end up relaying lots of data on behalf of remote nodes. The results of simulations using the DT and MTE communication protocols are shown in Figs. 1(a)-(d), with the BS located at the point with coordinate (50, 200). Clearly, DT and MTE result in a poor distribution for energy consumption by nodes. Sensor nodes in some subregions have all died out, but nodes in other regions are still active. As a result, data for a part of the sensor field may not be detected.

(a) 20% nodes dead (DT)    (b) 50% nodes dead (DT)

(c) 20% nodes dead (MTE)    (d) 50% nodes dead (MTE)

Figure 1. Survival statuses of sensor nodes using DT ((a) and (b)) and MTE ((c) and (d)) protocols.

Clustering communication protocols represent a superior approach, and result in more balanced patterns of energy use
in WSNs [9]. The first low-energy adaptive clustering hierarchy was LEACH, proposed by Heinzelman et al. [10, 11]. It showed how energy loads could be well amortized by dynamically creating a small number of clusters based on a threshold function $T(s)$ with a priori probability $p$ (say, 5%), in a set-up phase. The technique uses cluster heads (CHs) to mediate data transmission. Simulation results in [10, 11] show that all node tend to dissipate the same level of energy over time since the CH roles are rotated among nodes. Although LEACH clearly outperforms the DT and MTE protocols, it retains several shortcomings. Thus several enhanced versions of LEACH have appeared in the literature [12]. LEACH uses a threshold function parameterized by a probability $p$ input by user. However, the performance of sensor network is very sensitive to the value of $p$. When $p$ is large, many clusters are formed and could result in high energy consumption since many CHs dissipate energy in transmitting to the BS. On the other hand, when $p$ is small, only a few clusters are formed, which may increase energy dissipation when member nodes transmit to CHs. The literature suggests that the optimal $p$ value $p_{opt}$, or the optimal cluster number $k_{opt}$ depends on parameters such as the total number of nodes distributed in the sensor field, the size of sensor field, the location of BS, and so on [13, 14]. Our work proposes a genetic algorithm-based (GA-based) adaptive clustering protocol, termed LEACH-GA, to predict the optimal values of probability effectively.

II. ENERGY-EFFICIENT COMMUNICATION PROTOCOLS

We now briefly describe the LEACH protocol, and then present our genetic algorithm-based adaptive clustering protocol.

A. Clustering Hierarchy in LEACH

LEACH operates in several rounds, each consisting of a set-up and a steady-state phase. Each node transmits sensed data to its closest CH. The CH for each cluster receives and aggregates the data from cluster members and then transmits the aggregated data to the BS through a single-hop relay (shown in Fig. 2). LEACH creates a set-up phase for CHs’ selection, and a steady-state phase for time slot scheduling and transmission. Each sensor node $s$ decides independently of other sensor nodes whether it will claim to be a CH or not, by picking a random $r$ between 0 and 1 and comparing $r$ with a threshold $T(s)$ based on a user-specified probability $p$. The threshold is defined as follows [10]:

$$T(s) = \begin{cases} 
    p & \text{if } s \in G \\
    1 - p \left( r \mod \left( \frac{1}{p} \right) \right) & \text{otherwise} 
\end{cases} \quad (1)$$

Figure 2. The LEACH clustering communication hierarchy for WSNs.

After the cluster-setup sub-phase, the CH recognizes the number of member nodes and IDs of the nodes. Based on all join-request messages received within the cluster, the CH creates a TDMA schedule in addition to a unique spreading code, and transmits them to cluster members at the beginning of steady-state phase. Thereafter, all nodes in the cluster transmit their data packets to their CHs in the pre-specified TDMA time slot, using this code. As we known that TDMA-based protocols are naturally energy preserving, because they have time slots built-in, and do not suffer from collisions. Also, each member node can situate in a sleep mode at all times except during its corresponding time slots in order to decrease node’s energy dissipation. When the data packets sent by a node have been received by a CH, the CH aggregates and forwards them to the BS. These actions are repeated in each round. The plots of simulation results by LEACH are shown in Figs. 3(a) and (b), for the BS located at coordinate (50, 200). It is clearly shown that the nodes dead obtained using LEACH are more uniform than that of DT and MTE protocols.

(a) 20% nodes dead  (b) 50% nodes dead

Figure 3. Survival status of sensor nodes under LEACH.

B. Our Proposed Genetic Algorithm-based Adaptive Clustering Protocol

Our work introduces a genetic algorithm-based variant of LEACH to determine the optimal value of $p$ for various base station placements. The GA-based optimization procedure is performed only once, before the set-up phase of the first
round. The pseudo-code of the proposed protocol is described as follows.

**Pseudo-code of the Proposed LEACH-GA Protocol:**

BEGIN
1: Specify the probability \( p_{cand} \), number of nodes \( n \);
2: \( E_{init}(s)=E_0, s=1,2, \ldots, n; \)

(I) PREPARATION PHASE
1: if \( (E_{init}(s)>0 \) \& \( r\text{mod}(1/p_{cand})=0) \) then \( \rightarrow \) /p_{cand} can set\( \geq 0.5 \)
2: \( r\leftarrow\text{random}(0,1) \) and compute \( T(s); \) /given by (1)
3: if \( (r < T(s)) \) then
4: \( \text{CCH}\{s\}=\text{FALSE}; \) //node s not be a candidate CH
5: \( \) else
6: \( \text{CCH}\{s\}=\text{TRUE}; \) //node s be a candidate CH
7: \( \) end if
8: \( \) end if
9: SendToBS(IDu, (x u,y u), CCH(u))
10: \( \) else
11: \( \text{GAinBS}(); \) //node s not be a candidate CH
12: \( \) end if

(II) SET-UP PHASE
1: do \{ \( \) //repeat for r rounds
2: \( r\leftarrow\text{random}(0,1); \)
3: if \( (E_{init}(s)>0 \) \& \( r\text{mod}(1/p_{cand})=0) \) then
4: \( \) compute \( T(s); \) /given by (1)
5: \( \) end if
6: if \( (r < T(s)) \) then
7: \( \text{CH}\{s\}=\text{TRUE}; \) //node s be a CH
8: \( \) else
9: \( \text{CH}\{s\}=\text{FALSE}; \) //node s not be a CH
10: \( \) end if
11: \( \) if \( \text{CH}\{s\}=\text{TRUE} \) then
12: \( \text{BC (ADV)} \) ← broadcast an advertisement message;
13: \( \text{Join}(ID_i); \) //non-cluster head node i join
14: \( \) into the closest CH
15: \( \) Cluster(c); //form a cluster c;
16: \( \) end if

(III) STEADY-STATE PHASE
1: if \( \text{CH}\{s\}=\text{TRUE} \) then
2: \( \text{Receive}(ID_i, \text{DataPCK}); \) //receive data from members;
3: \( \text{Aggregate}(ID_i, \text{DataPCK}); \) //aggregate received data;
4: \( \text{TansToBS}(ID_i, \text{DataPCK}); \) /transmit received data;
5: \( \) else
6: \( \) if \( \text{MyTimeSlot}=\text{TRUE} \) then
7: \( \text{TansToCH}(ID_i, \text{DataPCK}); \) /transmit sensed data;
8: \( \) else
9: \( \text{SleepMode}(i)=\text{TRUE}; \) //node i at a sleep state
10: \( \) end if
11: \( \) end if
12: \} // one round is completed

END

### III. OPTIMAL CLUSTERING ANALYSIS

We evaluate our protocol using the first-order radio model of [10]. The parameter settings used in the simulation for the model are listed in TABLE I. According to the radio energy dissipation model of Fig. 4, the energy required by the transmit amplifier \( E_{ETx}(l,d) \) to transmit an l-bit message over a distance \( d \) between a transmitter and receiver is

\[
E_{ETx}(l,d) = \begin{cases} 
1\times E_{elec} + l\times E_{fs} \times d^2 & \text{if } d \leq d_0 \\
1\times E_{elec} + l\times E_{mp} \times d^4 & \text{if } d \geq d_0 
\end{cases}
\]  

where \( d_0 = \sqrt{E_{fs}/E_{mp}} \) denotes the threshold distance, \( E_{elec} \) represents the energy consumption in the electronics for sending or receiving one bit, and \( E_{fs} \) and \( E_{mp} \) represent amplifier energy consumptions for a short- and long-distance transmissions, respectively. To receive an l-bit message, the energy \( E_{ERx}(l) \) required by the receiver is given by

\[
E_{ERx}(l) = l\times E_{elec}
\]

**TABLE I.** \( \) PARAMETER SETTINGS OF THE FIRST-ORDER RADIO MODEL

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial energy ( (E_0) )</td>
<td>0.5 J/node</td>
</tr>
<tr>
<td>Transmitter Electronics ( (E_{elec}) )</td>
<td>50 nJ/bit</td>
</tr>
<tr>
<td>Receiver Electronics ( (E_{elec}) )</td>
<td>50 nJ/bit</td>
</tr>
<tr>
<td>Data Packet Size ( (l) )</td>
<td>2000 bits</td>
</tr>
<tr>
<td>Transmitter Amplifier ( (\varepsilon_p) ) if ( d \leq d_0 )</td>
<td>10 or 100 pJ/bit/m²</td>
</tr>
<tr>
<td>Transmitter Amplifier ( (\varepsilon_p) ) if ( d \geq d_0 )</td>
<td>0.0013 pJ/bit/m²</td>
</tr>
</tbody>
</table>

Let a total of \( n \) sensor nodes be distributed uniformly in the sensor field of size \( M \times M \) meters, and be grouped into \( k \) clusters. The energy required per round for a CH to receive data packets from member nodes, and aggregate and forward them a distance \( d_{toBS} \) to the BS is

\[
E_{CH}(l,d) = \begin{cases} 
1\times \left[ E_{elec} \times \frac{n}{k} + E_{DA} \times \frac{n}{k} + E_{elec} \times d_{toBS}^2 \right] & \text{if } d_{toBS} < d_0 \\
1\times \left[ E_{elec} \times \frac{n}{k} + E_{DA} \times \frac{n}{k} + E_{elec} \times d_{toBS}^2 \right] & \text{if } d_{toBS} \geq d_0 
\end{cases}
\]  

where \( E_{DA} \) represents the energy dissipation for aggregating data. The energy dissipation for a non-cluster head node is

\[
E_{non-CH}(l,d) = l\times E_{elec} + l\times E_{fs} \times d_{toCH}^2
\]

where \( d_{toCH} \) represents the distance between a cluster member and its CH. Since the nodes are assumed to be uniformly distributed in the sensor field, the expected value of squared distance from a member nodes to its CH, which located at the
point \((a, b)\), is given by

\[
E[d^2_{\text{toCH}}] = \frac{1}{A} \iint (x-a)^2 + (y-b)^2 \, dx \, dy
\] (6)

Assuming the shape of clusters is a circle, thus (6) becomes

\[
E[d^2_{\text{toCH}}] = \frac{2}{A} \int (x^2 + y^2) \, dx \, dy
\]

\[
= \frac{M^2}{\pi} \quad \text{(7)}
\]

The value of \(d^2_{\text{toCH}}\) in (7) is twice that of Heinzelman et al., who assumed that the CH is placed at the center of cluster. Moreover, the energy dissipated in a cluster is obtained as

\[
E_{\text{total}} = k \times \left( E_{\text{CH}} + \frac{n}{k} E_{\text{non-CH}} \right) = k \times \left( E_{\text{CH}} + \frac{n}{k} E_{\text{non-CH}} \right) \quad \text{(8)}
\]

Thus, the total energy dissipation for a round is given by

\[
E_{\text{Total}} = \begin{cases} 
1 & \text{if } d_{\text{ubs}} < d_0 \\
2nE_{\text{elec}} + nE_{\text{DA}} + kE_{\text{rc}} E[d^2_{\text{to}}] + \epsilon_n \frac{nM^2}{\pi} & \text{if } d_{\text{ubs}} \geq d_0 
\end{cases}
\]

(9)

From (9), the analytical optimal solutions for \(k_{\text{opt}}\) and \(p_{\text{opt}}\) are obtained.

\[
k_{\text{opt}} = \begin{cases} 
\frac{M}{\sqrt{\pi E[d^2_{\text{ubs}}]}} & \text{if } d_{\text{ubs}} < d_0 \\
\frac{E_{\text{fs}}}{\sqrt{\pi E_{\text{mp}} E[d^2_{\text{ubs}}]}} & \text{if } d_{\text{ubs}} \geq d_0 
\end{cases}
\]

(10)

and

\[
p_{\text{opt}} = \frac{k_{\text{opt}}}{n} = \begin{cases} 
\frac{1}{\sqrt{n\pi E[d^2_{\text{ubs}}]}} & \text{if } d_{\text{ubs}} < d_0 \\
\frac{E_{\text{fs}}}{\sqrt{n\pi E_{\text{mp}} E[d^2_{\text{ubs}}]}} & \text{if } d_{\text{ubs}} \geq d_0 
\end{cases}
\]

(11)

We assume the coordinates of the BS to be \((0.5M, 0.5M+B)\), and calculate the values of \(E[d^2_{\text{to}}]\) and \(E[d^4_{\text{ubs}}]\) to be

\[
E[d^2_{\text{to}}] = \frac{M^2}{6} + B^2; \quad E[d^4_{\text{ubs}}] = \frac{7M^4}{180} + \frac{2}{3} B^2 M^2 + B^4
\]

(12)

Therefore the values of \(k_{\text{opt}}\) and \(p_{\text{opt}}\) are related to the total number of sensor nodes, domain size of sensor field, and the location of BS. In addition, Heinzelman et al. assumed the BS is far from the nodes, so the energy dissipation follows the multipath model. Thus, their formula for \(k_{\text{opt}}\) is only the lower part of (10). The original formula, denoted as original form in this work, for \(k_{\text{opt}}\) from Heinzelman et al. is [11]

\[
k_{\text{opt}} = \sqrt{\frac{n}{2\pi} E_{\text{fs}} M \epsilon_n d_{\text{ubs}}^2}
\]

(13)

Therefore, the \(p_{\text{opt}}\) can be formulated as

\[
p_{\text{opt}} = \frac{k_{\text{opt}}}{n} = \frac{1}{2m} E_{\text{fs}} M \epsilon_{\text{mp}} d_{\text{ubs}}^2
\]

(14)

In this work, (10) and (11) are used as the corrected forms of analytical solution for \(k_{\text{opt}}\) and \(p_{\text{opt}}\) without assuming the positions of BS located near or far from the sensor field.

### IV. GENETIC ALGORITHM-BASED CLUSTERING

At the beginning of preparation phase, each node initially determines whether or not it should be a candidate cluster head (CCH), using the following cluster head selection procedure. First, every sensor node selects a random number \(r\) from the interval [0, 1]. If \(r\) is smaller than \(T(s)\), based on a prescribed probability \(p_{\text{set}}\), then the node is a CCH. The value of \(p_{\text{set}}\) can be a large value in our protocol, \(p_{\text{set}}=0.5\), say. Thereafter, each node sends its ID, location information, and whether or not it is a CCH to the BS. As the BS receives messages sent by all nodes, it performs GA operations to determine the optimal probability, \(p_{\text{opt}}=k_{\text{opt}}/n\), by minimizing the total amount of energy consumption in each round. Therefore, the objective function used in the GA can be formulated as

\[
f(\bar{x}) = \sum_{c=1}^{k} \sum_{i=1}^{q} (E_{\text{elec}} + E_{\text{DA}} + \epsilon_{\text{mp}} d^2(\text{CCH}(c), \text{BS})) \times x_c + \sum_{c=1}^{k} (E_{\text{elec}} + E_{\text{DA}} + E_{\text{elec}} + \epsilon_{\text{mp}} d^2(\text{CCH}(c), \text{BS})) \times x_c
\]

(15)

where \(\bar{x} = [x_1, x_2, ..., x_c, ..., x_k]\). The values of \(x_c\) are one for our binary-GA when it is a CCH, otherwise, it is zero. The parameters \(e=E_{\text{fs}}\) and \(\alpha=2\) were used for \(d\text{ubSD}\); while, \(e=E_{\text{mp}}\) and \(\alpha=4\) were set for \(d\text{ubSp}\). The symbol \(q\) represents the number of member nodes in a CCH. The optimal probability \(p_{\text{opt}}\) is determined by the GA by searching the solution space through an evolutionary optimization process incorporating probabilistic transitions and non-deterministic rules, and applying selection, crossover and mutation operators. Once the optimal probability \(p_{\text{opt}}\) is found, the BS broadcasts the value of \(p_{\text{opt}}\) to all nodes. The set-up and steady-state phases begin. The procedures of set-up and steady-state phase are the same as in LEACH.

### V. SIMULATION RESULTS

Our work assumes that all sensor nodes are homogeneous and distributed uniformly over the sensor field with limited energy that the links between nodes are symmetric, and that messages from all nodes can reach the BS. The nodes are distributed randomly in a square of size \(MxM\). Each
simulation is repeated for 30 independent runs. In addition, control packet sizes for broadcasting packet and packet header were 50 bits long, and the energy dissipation for aggregating data was 5 and 10 nJ/bit/signal.

A. Comparison of Optimal Probability of Cluster Heads

In this section, the energy dissipation for aggregating data and the parameter $\varepsilon_f$ were specified as 5 nJ/bit/signal and 10 pJ/bit/m², respectively. The total number of sensor nodes was 100. Figures 5(a) and (b) show the comparison of optimal probability obtained from model analysis and GA-based computation for a variety of locations of BS for the sensor fields of 50m×50m and 100m×100m. The comparison of solutions depicts that the distribution of present modified analytical formulas. Moreover, the results show that the optimal probability, $p_{opt}$, is clearly affected by the BS location, the center of sensor field. When the BS is located near the sensor field, the values of $p_{opt}$ are large. On the contrary, the values of optimal probability decrease as the BS moves farther from the sensor field. When the BS is located at the center of sensor field, the values of $\sqrt{E[d_{obs}^2]}$ is given by [15, 16]

$$\sqrt{E[d_{obs}^2]} = \frac{1}{A} \int \sqrt{x^2 + y^2} \, dA = 0.765 \frac{M}{2}$$

(16)

and the form of $p_{opt}$ can be simplified as

$$p_{opt} = \frac{1}{n} \frac{2}{0.765}$$

(17)

Equation (17) states that the parameter $p_{opt}$ is just function of the total number of sensor nodes only when the BS is located at the center of sensor field. Namely, the value of probability at the center of sensor field is independent of the domain size.

Moreover, the values of $p_{opt}$ are clearly dependent of the total number of sensor nodes $n$ from the expression of (11). When the number of sensor nodes increases, the values of optimal probability will decrease based on (11). Two cases are conducted using $n=200$ and $n=400$ to study the effect of $n$ to the value of $p_{opt}$. The simulation results for $n=200$ and $n=400$ are displayed in Figs. 6 and 7, respectively. Figures 6(a) and (b) show that the predicted distributions of optimal probability by present GA were as well as the present corrected form governed by (11) for BS located at different positions. In these two figures, the values of optimal probability at the center of sensor field were independent of the domain size of sensor field based on (17). This work also performed the simulation of the case with $n=400$, and the results are shown in Figs. 7(a) and (b). From the results shown in Figs. 5 and 7, the values of optimal probability at the center of sensor field for the case of $n=400$ were equal to the half of that of the case for $n=100$.

B. Comparison of the Presented LEACH-GA and LEACH

In this section, the nodes with 100 are distributed randomly in the $M\times M$ sensor field with $50m\times50m$. Each simulation is also repeated for 30 independent runs, and solutions are obtained from the average of the runs. In addition, control packet sizes for broadcasting packet and packet header were 50 bits length for the present computations, and the energy dissipation for aggregating data and the parameter $\varepsilon_f$ were specified as 10 nJ/bit/signal and 100 pJ/bit/m² [11], respectively. Figure 8 is the solution distribution of $p_{opt}$ predicted by using the presented LACH-GA for BS located at different positions. The comparison of the values of $p_{opt}$ computed by LEACH-GA
and corrected formula is agreeable, whereas, the data using Heinzelman et al.’s formula shown in (14) were clearly discrepant compared to LEACH-GA and our corrected form. Especially, there are large errors performed using the original form when the BS located near to the center of sensor field.

TABLE II lists the simulation results obtained using LEACH and presented LEACH-GA protocols for BS located at different positions. The initial energy for all nodes was 0.5(J) and the probability $p$ used in LEACH is 5%, same as the settings in [10, 11]. The number of rounds required when the number dead of nodes is 1%, 20%, 50%, and 100% are recorded during simulations. From our results, the values of $p_{opt}$ clearly depend on the positions of BS. The value of optimal probability is the largest when the BS is at the center of sensor field, and it decreases when the BS moves outward. Moreover, the proposed LEACH-GA outperforms LEACH in terms of lifetime of network.

VI. CONCLUSIONS

This work proposed a GA-based adaptive clustering protocol to determine the optimal thresholding probability for cluster formation in WSNs. The LEACH protocol requires the user to specify this probability for use with the threshold function in determining whether a node becomes a CH or not. However, the network performance is extremely sensitive to this probability, and it is very hard to obtain an optimum setting from available prior knowledge. Hence, our approach uses a preparation phase prior to the set-up phase of the first round to gather information about node status, IDs, and location and sends it to the BS, which determines the optimal probability to use in the CH selection mechanism. Our simulation results show that the optimum distribution of probability matched the analytical results proposed by our corrected formulas well. Moreover, our proposed LEACH-GA method outperforms MTE, DT, and LEACH in terms of network lifetime, since the use of the optimal probability yields optimal energy-efficient clustering.
ACKNOWLEDGMENT

The work was supported in part by National Science Council of Republic of China under Grant Number NSC 98-2221-E-214-038-MY2.

REFERENCES