An Intelligent Evaluation Model based on the LEACH Protocol in Wireless Sensor Networks

Ning Cao  Russell Higgs  Gregory M.P. O'Hare

Abstract—This paper aims to introduce some key parameters for the tracking application in wireless sensor networks. In this work the LEACH protocol with J-sim simulation tool has been implemented, and consequently some useful trade-off analysis results among the EDCR (Energy, Density, Coverage and Reliability) parameters has been obtained. Based on these results, an intelligent evaluation model is proposed in this paper.

Keywords-J-sim; LEACH; EDCR; Evaluation Model

I. INTRODUCTION

Wireless sensor networks (WSNs) are employed in a number of different application areas [1], such as military, health, home and other commercial applications. A WSN typically consists of a given number of sensor nodes, which implement computation and communication functions.

Wireless sensor networks may consist of many different types of sensors. A typical sensor may be capable of monitoring a number of sensing modalities that may include temperature, humidity, vehicular movement, pressure, noise levels, together with the speed, size and direction of an object.

Several researchers are currently engaged in research based on different parameters in wireless sensor networks.

In [2] the goal for the authors to address the target coverage problem is to prolong the network lifetime of the energy constrained wireless sensor network. The sensor nodes are deployed randomly around the target node. If the target enters in the sensor's sensing range, the sensor will get the target location information and send this information to a sink node. In this paper the researchers propose an efficient approach to prolong the WSN's lifetime by organizing the nodes into a maximal number of set covers which can be activated continuously. The sensor nodes from the current active set have the responsibility to monitor the target and transmit the collected data, while all other nodes are in a low-energy sleep state. In the simulation work, the researchers have shown many results among radius, number of nodes and lifetime, however, the researchers haven’t proposed any mathematical model for these results.

In a wireless sensor network, all the nodes are deployed with a predefined communication radius. The network is obviously very stable with a fixed number of nodes and the number of nodes strictly follows a pure mathematical formula. The deployment of structured network architecture is not feasible in several areas, for example, temperature monitoring in a building. In this case, a random distributed wireless sensor network is a better choice. For a random structured wireless sensor network, the researchers need to take into account the minimum critical communication distance for large areas to ensure network connectivity and stability. Although certain algorithms based on mathematical formulae exist, Maity and Gupta [3] have proposed an analysis between the critical communication radius and number of nodes for a random distributed wireless sensor network.

Compared to the existing work, this paper focuses on the parameters of Energy, Density, Coverage and Reliability. Based on the results of these parameters, an intelligent evaluation model will be proposed.

There is no need to assume that all sensors measure the same phenomena; however, this paper will assume that all sensors can transmit data about the phenomena they detect and can receive and retransmit data from other sensors about any of the phenomena detected in the system.

The normal measure of accuracy in a wireless sensor network for one quantity is:

\[
\text{Accuracy} = 1 - \left(\frac{\text{real value of quantity} - \text{measured value of quantity}}{\text{real value of quantity}}\right) \tag{1}
\]

Accuracy is an average figure computed over the lifetime of the network.

There are a number of drawbacks when using this definition. Firstly, with a single quantity in order to be measured you need to know the real value of the quantity to work out the accuracy, whereas the whole point of a wireless system network is to remotely estimate the real value. In reality it may not be possible to measure the real value of the quantity in the field and so accuracy might have to be calculated by simulation before real deployment. In particular accuracy in the field may differ
from that measured under laboratory or controlled conditions, but the end-user will be unaware of this. Secondly, suppose the network is going to measure $n$ different quantities, where $n$ is greater than one. Then using the definition above this work will obtain $n$ accuracy values, making analysis of the system far harder and results from such analysis difficult or impossible to graph. Instead in this paper experiments are conducted using the concept of reliability, defined by:

$$\text{Reliability} = \frac{\text{the number of packets received by the sink node}}{\text{the number of packets sent to the sink node}}. \quad (2)$$

For the LEACH protocol considered in this paper the denominator will be taken to be the total number of packets sent to the sink node by the cluster heads during the lifetime of the network, whereas the numerator will taken to be the total number of such packets received by the sink node during the lifetime of the network.

Firstly, this definition can be used in the field and laboratory. In the field provided the sink node (end-user) knows how many cluster heads there are and the average number of packets of sensed data per sensor per time unit, then reliability can be estimated at any time during the lifetime of the network using the number of time units elapsed. Any difference between reliability measured in the field and laboratory could be used to detect how many packets of data are lost between sensors and cluster heads for LEACH (this would include counting sensors that have failed or have otherwise been lost to the system).

Secondly, this simple definition of reliability can be used in a network where there are sensors of various types monitoring different things to produce a single simple measure of accuracy. In this situation our definition gives a systemic measure of accuracy for the network as a whole independent of the applications. This definition also makes it easy to analyze communication among the sensor nodes; in particular it makes the estimation of data collision in the wireless sensor network possible.

A frequently used definition for tracking accuracy is to set the tracking error equal to the Euclidean distance between the estimated and actual locations of the target. The chief advantage of our definition for the tracking application, that this paper is going to simulate, is that it is independent of the target position.

This work has completed power control over the radio components on the J-sim simulation platform, which makes simulation of power consumption possible. Besides this, LEACH provides sensor networks with many good features, such as clustering architecture, localized coordination etc. So this paper is going to analyze some critical parameters based on LEACH protocol with J-sim.

In this reliability and lifetime will be evaluated using density as a parameter. As the number of nodes increases in a fixed space, the efficiency of the sensor network may become better or worse. So in some simulations the researchers may care about the number of sensors they are using in the network. It is obvious that some of the nodes will be out of power in a real wireless sensor network. Thus attention will move to the number of sensors that are alive.

In a wireless sensor network, all the nodes are deployed with a predefined communication radius. This limits the area that the sensors can detect phenomena in, so that this paper will also evaluate the effect that this limitation on coverage has on lifetime and reliability by varying the radius.

Network lifetime has become the key characteristic for evaluating sensor networks in an application-specific way. In particular the availability of nodes, sensor coverage, and connectivity have been included in discussions on network lifetime. In fact, even quality of service measures can be reduced to lifetime considerations. Network lifetime is the time span from the deployment of the sensors to the instant when the network is considered non-functional. When a network should be considered non-functional is application specific. It could be, for example, the instant when the first sensor dies, a percentage of sensor die, or a loss of coverage occurs. The definition for network lifetime in our experiments is the time at which the last sensor dies in the sensor network, which using the Leach protocol is essentially equal to the time at which the sensor network provides no quality of service at all.

Conserving sensor energy and increasing tracking accuracy are the two main goals for the research of target tracking applications in wireless sensor networks. Simulation is a common way to compare these two parameters. Before the real deployment of sensor nodes, users always perform some simulation tests for the tracking application. In the simulations, finding the trade-off point between energy consumption and tracking accuracy is one of the key questions to be addressed. In addition, evaluation analysis will be performed here among all the related parameters.

The rest of this paper is structured as follows: Section 2 contains a brief overview for the system architecture. Section 3 focuses on the experiment setup information. Section 4 describes the results which have been obtained so far. Section 5 proposed a model of Lifetime. Section 6 proposed a model of Reliability. Section 7 concludes this paper.

II. SYSTEM ARCHITECTURE

This work applies a standard protocol stack, whereby equivalent layers are able to communicate with each other via neighbor sensors. A MAC layer is required to assign the communication channel to the sensor nodes as the sensor nodes cannot share the channel. One approach to MAC design for sensor networks is to use time-division multiple access (TDMA) based protocols that conserve
more energy than contention-based protocols like carrier sense multiple access. Time-division multiple access allows several users to share the same frequency channel by dividing the signal into different time-slots. It has a natural advantage of collision free medium access. The approach of TDMA in a MAC layer is also the basic technique to realize the LEACH routing protocol. MAC layer provides the RTS/CTS mechanism to control communication among different sensor nodes. Unfortunately these management signals can be dropped due to collisions. These dropped packets may lead to an error for the sensed data.

J-Sim (formerly known as JavaSim) is a component-based, compositional simulation environment. It has been built upon the notion of the autonomous component programming model and has been developed entirely in JavaTM. The system is based on the IEEE 802.11 implementation provided with J-sim [4]. IEEE 802.11 is the first wireless LAN (WLAN) standard proposed in 1997. J-sim is implemented on top of the autonomous component architecture, so components are basic elements in this architecture. With these components, J-sim implements the data transmission process.

In this paper the LEACH protocol will be used as the routing protocol. LEACH [5] is a TDMA-based MAC protocol which is integrated with clustering and a simple routing protocol in wireless sensor networks.

The LEACH protocol differentiates itself in that:
- It uses randomized rotation of the cluster-heads.
- It reduces the amount of data that needs to be transmitted to the sink node.

Consequently the use of clusters decreases the number of intermediate nodes. Furthermore, using rotating cluster-heads and adaptive clusters, the energy requirements of the system are in general distributed among all the sensors.

LEACH is one of the hierarchical protocols, in which most sensor nodes transmit sensed data to cluster heads, and the cluster heads aggregate and compress the data and forward the data to the sink node. Each node uses a stochastic algorithm at each round to determine whether it will become a cluster head in that round. So, if the remaining energy for each node can be measured, it will make great contributions to this research area.

III. EXPERIMENT SETUP

There are three main ways in which packets of data can be dropped, these are:
- A sensor may receive data from the target, but be unable to transmit it to any cluster head through lack of power or insufficient radius.
- Packets from sensors can be dropped through data collisions with packets from other sensors or cluster heads.
- Aggregated data sent from cluster heads can be dropped through data collisions with packets from sensors or other cluster heads.

The definition of reliability excludes data dropped in the first and second ways and so the primary concern in this paper will be data collisions that affect at least one cluster head. A data collision is the simultaneous presence of signals from two nodes on the network. A collision can occur when two nodes each think the network is idle and both start transmitting at roughly the same time. Considering a single cluster in isolation our use of TDMA ensures that no data collisions occur within the cluster. The LEACH protocol also employs CDMA, code-division multiple access, this uses spread-spectrum technology and a special coding scheme to allow sensors from different clusters to transmit their data at the same time without data collisions occurring. However, data from different cluster heads can collide en route to or at the sink node leading to a severe loss of data. Finally it is possible that a sensor’s transmission can interfere with that of a cluster head or vice versa. Thus data collisions can be classified into two types:
- Cluster head with cluster head.
- Sensor with cluster head.

When a data collision occurs it is assumed that all the data from the two (or more) sources are lost. Now, because of the definition of reliability this work is only concerned about measuring packets of data sent to the sink node, which are those packets sent from the cluster heads. So our reliability measure will measure the first type of data collisions, but will only measure the data lost by the cluster head in the second type.

The simulated area for the experiments in the following sections is a 10 metre \( \times \) 10 metre square with randomly deployed nodes. The sink node for this application is located in the middle of this square. One of the primary reasons for selecting this setup is to allow the results to be generalized to large areas by concatenation of networks similar to this. For example, a 100 metre \( \times \) 100 metre square region could be configured using 100 instances of the setup used here in a 10 \( \times \) 10 grid formation.

IV. RESULTS

A. Density-Lifetime-Reliability

In order to analyze the effects of density, network lifetime and reliability on WSN performance, a series of experiments were undertaken varying the number of sensor nodes starting at 2, increasing to 4, 6, 9, 12, 15, 18, 21, 24, 30, 45 and finally 60. The transmission radius for each sensor remained at 10 metres and the number of clusters for this application was two. The transmission radius for the target node was 10 metres.

As the number of nodes increases in the fixed space, more sensors can be elected as cluster heads and so the energy for each sensor can be conserved. Thus the
lifetime of the network should increase as the number of sensors increases. Fig. 1 confirms as expected that network lifetime increases as density increases. In other words, the relationship between density and network lifetime is that of positive correlation. If the number of sensors, reliability and network lifetime are multiplied together the following figures are obtained in the 12 cases: 705, 1371, 1868, 2329, 2755, 3200, 3164, 3408, 3785, 3592, 4215, 4585. This gives a measure of how much data is received by the sink node, lifetime increases with density whilst reliability decreases, and our figures indicate that the total number of packets of data received by the sink node will increase with the number of sensors, although data lost from sensors before they reach a cluster head is being ignored.

From the data one can observe that reliability, lifetime and the number of sensors are related by the following formula:

\[
\frac{\text{Reliability}}{\text{Lifetime}} \times \text{Number of sensors} = \text{Constant} \quad (3)
\]

This equation says that the rate of successful packet reception per unit time is independent of the number of sensors. The constant in equation (3) will depend on the parameters in the simulation and it is a reasonable model if the number of cluster heads is small. The number of sensors in the cluster is one way of measuring packets received, but again ignores data lost from sensors before it reaches the cluster head. Such lost data will increase with the number of sensors, so that the constant on the right hand side of the equation may have to be found for a small, medium or large number of sensors.

This experiment gives the following ‘constants’ in the 12 cases (to three decimal places): 0.000297, 0.000312, 0.000298, 0.000302, 0.000291, 0.000293, 0.000269, 0.000282, 0.000292, 0.000261, 0.000277, 0.000280, which because these figures are all very close to each other indicates that the model is reasonable.

Now as there is a constant rate of successful packet reception per unit time it follows from calculus that plotting reliability multiplied by number of sensors against lifetime a line \( ax + b \) should be obtained, where \( x \) represents the lifetime and \( a \) is ‘the’ constant obtained above.

The least-squares approximation method from statistics was used to fit the line through the data points in Fig. 2, this yields the line

\[
0.00025x + 0.10993
\]

Invert this linear relationship to obtain

\[
\text{Lifetime} = 4000y - 439.72
\]

where \( y \) is as in Fig. 2. The 4000 is the rate at which lifetime (energy) is used per unit of \( y \). The constant 439.72 appears to be a measure of the lifetime used in cluster head activities that are independent of the number of sensors.
The initial experiment was then repeated with all the parameters as before except for the radius which was increased from 1 to 10 metres in 1 metre steps. The curves obtained were all similar to those in Fig. 1.

Finally in Fig. 3, the transmission radius for sensors and the target node have been increased from 10 to 15 metres. The lifetime for this experiment in Fig. 3(a) increases from 1421 to 2077, 2407, 2767, 3053, 3183, 3401, 3577, 3597, 3875, 3957 and finally to 4055. The reliability in Fig. 3(b) decreased from 24.38% to 15.66%, 12.38%, 9.35%, 7.48%, 6.07%, 5.31%, 4.98%, 4.08%, 3.65%, 2.35% and finally to 1.87%. Comparing Fig. 1 and Fig. 3, this paper concludes that lifetime and reliability are not affected much by the increase in radius with the shape of the curves much the same as before, this will be further analyzed in the next section. However, there is a degree of experimental variation in our results as the cluster heads will be randomly deployed in different positions. So when the number of sensors is just two the radius should have no effect on the reliability or lifetime as the only transmissions these sensors have to make is as cluster heads.

B. Coverage-Lifetime-Reliability

In order to analyze the effect of communication coverage, network lifetime and reliability on WSN performance, a series of experiments were carried out with a transmission radius of each node starting at 1 metre, and then increasing in 1 metre steps up to 8 metres. The ability to communicate with the sink node for the sensors is limited by the transmission radius. In this experiment, this work set a fixed value of 20 sensors for the density. The number of clusters for this application was two and the transmission radius for the target node was 15 metres.

In Fig. 4(a) the value of lifetime starts at 3487 and then becomes 3427, 3431, 3475, 3495, 3567, 3423, 3455.

In Fig. 4(b) the value of reliability starts at 4.79% and then becomes 4.75%, 4.67%, 4.96%, but at the fifth scenario it reaches its highest point of 5.09% and then drops to 5.08%, 5.03% and finally to 4.73%.
First two statistical results: Randomly pick two independent sets, $x$ and $y$, of points from the closed interval from 0 to 1 and place them at the points $(x, y)$ to obtain points uniformly distributed over the unit square. It is known from this distribution that the expected distance of a randomly chosen point in the unit square from the center of the square is

\[
\frac{\sqrt{2} + \sinh^{-1} 1}{6}
\]

and the expected distance between two randomly chosen points in the unit square is

\[
\frac{2 + \sqrt{2} + 5 \sinh^{-1} 1}{15}
\]

This means for a square of side 10 metres, the mean distance of a point from the center is approximately $3.825978582$ metres and the mean distance between two points is approximately $5.21405433$ metres.

For a given radius $r$ sensors may be divided into two types: isolated sensors and cluster sensors. An isolated sensor is a sensor which is not in any cluster. Isolated sensors will only be able to feature in data collisions with cluster heads if they are in the region(s) indicated in Fig. 5, and isolated sensors in this region are called central isolated sensors. Notice that central isolated sensors have the potential to collide with data transmitted from one or more cluster heads. A central isolated sensor within radius $r$ of the sink node has the potential to collide with all data transmitted from all the cluster heads. In reality the broadcast time of a sensor is very short, so that it is highly likely that if a central isolated sensor within radius $r$ of the sink node collides with the data from more than one cluster heads then these would have collided in any case, but such a sensor could certainly cause a ‘new’ data collision with just one cluster head.

Concentrating on central isolated sensors outside of radius $r$ of the sink node and look at the case with two cluster heads, such a sensor can only cause a collision with data from its nearest cluster head.

The area in Fig. 5 is obtained by imagining there are wired connections from cluster heads $A$ and $B$ to the sink node in the center of Fig. 4. A central isolated sensor for cluster $A$ must then be within radius $r$ of the wire but outside of the circle of radius $r$ and center $A$. The total area that central isolated clusters can occur in can be taken to be approximately

\[
f(r) = 2r(d_1 + d_2) - \pi r^2
\]

where $d_1$ and $d_2$ are the distances of the cluster heads from the sink node. Now

\[
f'(r) = 2(d_1 + d_2) - 2\pi r
\]

and

\[
f''(r) = -2\pi
\]

Thus a local maximum occurs when

\[
r = \frac{(d_1 + d_2)}{\pi}
\]

Setting $d_1 = d_2 = 3.825978582$ metres, the maximum occurs for $r = 2.435693614$ metres. There are no central isolated sensors when $f(r) = 0$, which occurs when

\[
r = \frac{2(d_1 + d_2)}{\pi}
\]

Setting $d_1 = d_2 = 3.825978582$ metres, this occurs for $r = 4.871387228$ metres. These values correspond roughly with the experimental values that a minimum value for reliability occurs with a radius of between 2 and 3 metres and a maximum with a radius of about 5 metres.

The second type of sensor that can occur is a cluster sensor, that is, a sensor that occurs within a cluster. Such a sensor does not give rise to a data collision with its own cluster head but could cause a collision with the data from another cluster head. Indeed once the radius passes the distance of the sensor to the sink node it has the capacity to collide with the data from all other cluster heads.
Continuing to increase the radius from there will have no further effect on data collisions from this one sensor; however, the number of sensors in each cluster will increase, which in turn will lead to more data collisions with cluster heads. Thus only considering cluster head sensors reliability should decrease with increasing radius and this does occur from a radius of 5 metres onwards. Considering just two clusters A and B, a cluster sensor in cluster B to cause a data collision with cluster head A has to be in the same area as the central isolated cluster region of cluster A in Fig. 5 and of course it also has to be in within a circle of radius r about cluster head B (so in Fig. 5, no such troublesome sensor would exist).

Our conclusions are that the dominant type of sensors for reliability for a small radius are central isolated ones, as the radius increases the number of central isolated sensors decrease as they transform into cluster sensors, these then become the dominant type of sensors and lead to a monotonic decrease in reliability. The results indicate that setting the radius to about half the length of the side of the sensor area will produce a good result for both reliability and lifetime independent of the other parameters.

Generally the effect of coverage seems very weak and our experiments indicate that the coverage parameter is not critical for LEACH.

V. LIFETIME MODEL

This section will propose a simple relationship between lifetime and the number of sensors assuming all other parameters remain constant. Lifetime can be being modeled by a two parameter equation involving E, the energy used in cluster head activities per time unit and e, the energy used in normal sensor activities per time unit.

The energy used in normal sensor activities should be independent of the number of sensors and therefore e is constant. On the other hand E will increase with the number of sensors as each cluster will contain more sensors and so its cluster head will have to aggregate (and transmit) more data. Nevertheless E is assumed as a constant by finding an average value for it.

Let K be the total battery energy of a sensor, let n be the total number of sensors and let r be the number of clusters. Then lifetime should be governed by the equation

\[ nK = rtE + (n - r)e, \]

where t denotes the lifetime.

Thus

\[ t = \frac{nK}{rE + (n - r)e} \]

Now as \( n \to \infty \), \( t \to K/e \), this of course is independent of \( r \) and \( E \). So the horizontal line \( t = K/e \) should form a horizontal asymptote in the density-lifetime graph. On the other hand when \( n = r \), \( t = K/E \) as one would expect.

In Fig. 1 there are the following twelve data points: (2, 1541), (4, 2097), (6, 2503), (9, 2779), (12, 3077), (15, 3307), (18, 3427), (21, 3475), (24, 3601), (30, 3707), (45, 3903), (60, 4043). The data for 15 and 30 sensors could be used to obtain an average value for \( E \).

Now

\[ 15K = 6614E + 42991e \]

and

\[ 30K = 7414E + 103796e \]

Eliminating \( K \) from these equations gives

\[ 13228E + 85982e = 7414E + 103796e \]

and so

\[ 5814E = 17814e \]

Thus

\[ e = 0.326372516E \]

and so from the first equation

\[ K = 1376.338722E \]

Using these approximations

\[ t = \frac{1376.338722n}{2 + 0.326372516(n - 2)} \]

So as \( n \to \infty \), \( t \to 4217.079119 \), this should be an overestimation of the lifetime, because of the average value taken for \( E \).

The graph of the function \( (n,t) \) with \( t \) represented by the formula above (the red curve in Fig. 6) has been overlaid with that of the lifetime in Fig. 1 (in blue in Fig. 6), showing that this is a good model.
If \( n \) is large compared to \( r \) a simpler relationship can be found by replacing \( n - r \) by \( n \) in the above equation, so that
\[
nK = rtE + nte,
\]
where \( t \) denotes the lifetime.

Then
\[
t = \frac{nK}{rE + ne}
\]
Repeating the above analysis (without assuming that \( r = 2 \)) yields
\[
e = 0.24225rE
\]
and
\[
K = 1135.237983rE
\]
Using these approximations
\[
t = \frac{1135.237983n}{1 + 0.24225n}
\]
indicates that
\[
K = 1135.237983rE
\]

Of course in reality the (average) total battery energy of a sensor will be known, since this can be measured before deployment.

VI. RELIABILITY MODEL

The formula connecting reliability and lifetime in section 4 will now be married with the equation for lifetime found in section 5 to obtain the following formula for reliability:

\[
\text{Reliability} = \frac{K}{rE + (n - r)e} \times \text{Constant}
\]

where the constant has to be determined by experimentation as shown in section 4. With the data in Fig. 1 (so \( r = 2 \)) and this time the arithmetic mean 0.000288 of the constants found in section 4, the following equation can be obtained

\[
\text{Reliability} = \frac{0.396385551}{2 + 0.326372516(n - 2)}
\]

using the formula found for lifetime in section 5.

Plotting this graph (in red in Fig. 7) over that of Fig. 1 for reliability (in blue in Fig. 7) shows that this is a good model for reliability.

VII. CONCLUSIONS

In this paper, the evaluation parameters of Lifetime and Reliability have been defined. Based on these definitions, some Energy, Density, Coverage and Reliability analysis has been undertaken for the LEACH protocol based upon a tracking application. From a series of experiments with J-sim simulation tool several conclusions were reached: (1) The network lifetime increases as the number of sensors increases. (2) There is a linear relationship between the number of packets received by the sink node and the lifetime. (3) The effect of coverage is weak indicating that this parameter is not critical for LEACH.

In this paper an intelligent evaluation model between the parameters of lifetime, reliability and density has also been proposed. Based on this intelligent evaluation model, wireless sensor network users can predict the lifetime and reliability directly. This means that sensor nodes can be deployed in such a network without further simulations.

ACKNOWLEDGMENT

This research is supported by Science Foundation Ireland under grant 07/CE/1147.

REFERENCES