An Intelligent Energy Efficient Clustering in Wireless Sensor Networks

Jamshid Shanbehzadeh, Saeed Mehrjoo, Abdolhossein Sarrafzadeh

Abstract— **One of the main challenges of wireless sensor network is how to improve its life time. The limited energy of nodes is the main obstacle. We may overcome this problem by optimizing the nodes' power consumption. A solution is clustering, but optimum clustering of wireless sensor network is an NP-Hard problem. This paper proposes a hybrid algorithm based on Genetic Algorithm and Particle Swarm Optimization to overcome this clustering problem by finding the number of clusters, the cluster heads and the clusters members. Simulation results reveal that this algorithm outperforms LEACH and Genetic Algorithm based clustering scheme.**

Index Terms—Clustering, Genetic algorithm, Lifetime, Particle swarm optimization, Transmission, Wireless sensor network

I. Introduction

Wireless Sensor Network (WSN) consists of a large number of sensor nodes where these nodes are low-power, low-cost and energy limited with constrained communication and computational capabilities [1]. WSN is employed in monitoring a specific region specially places that are hardly accessible such as battlefields [2] and volcanoes [3], detection of fire events in forests and jungles [4], measuring temperature and humidity in specific places [5] and many other applications. While replacing or recharging the batteries of nodes is infeasible, energy consumption of nodes is the major factor of WSN during communication. A negative effect is WSN lifetime reduction. A method to improve the lifetime is to reduce the number of transmissions by clustering the nodes. But, choosing cluster-heads, their number and cluster members are NP-Hard problem.

The LEACH protocol, a self-organized, hierarchical, cluster-based approach was proposed by Heinzelman et al. [12]. LEACH divides the date collection area into several pre-determined clusters, randomly. The sensor nodes transmit data to the cluster heads based on time division multiple access (TDMA), and cluster heads aggregate and transmit the data to the base station. After specific time intervals, LEACH chooses a new set of cluster heads. Only when all the remaining candidates have been elected, then a node can be re-elected.

Genetic algorithm (GA) is an adaptive method which is generally employed to solve search and optimization problems [6]. It is based on the genetic processes of biological organisms. We employ it to find the number of clusters and their heads. Particle Swarm optimization (PSO), motivated by the social behaviors of animals such as bird flocking and fish schooling, it is widely applied in optimization [8]. We employ PSO to overcome the problem of assigning nodes to cluster heads and constructing the clusters.

GA is extensively employed in clustering. Sajid Hussain and et al [7] successfully utilized a genetic algorithm approach to cluster the nodes but they used GA just in choosing the cluster heads. There are several reasons why PSO based methods are more preferable in comparison to Genetic algorithm approaches. For example GA is inherently a discrete method and each entity has no communication with the global optimization process. On the other hand PSO employs the ***gbest*** and ***lbest*** models to determine the next velocity vector, and is inherently continuous in nature. PSO converges to the solution faster and with less computation than a standard genetic algorithm [8]. Therefore, we employed PSO in clustering.

Ying Liang and et al [9] utilized PSO for clustering. They proposed a hybrid algorithm using PSO and LEACH but, their simulations show little improvement compared to LEACH. Yang and et al proposed an improved PSO to construct energy efficient clusters without simulating their works on WSN [10]. This paper employs GA and PSO algorithms in WSN. The simulations reveal significant performance of the novel algorithm over LEACH or GA.

The rest of our paper is organized as follows. Section II describes the proposed artificial intelligence based clustering. We will talk about general steps of data gathering in section III. Section IV discusses the simulations results and finally Section V concludes the paper and suggests further research.

II. Artificial Intelligence based clustering

This section introduces GA, PSO and their application in clustering WSN in brief. We Consider a network with N nodes named S1,S2,…,SN and a base station (BS). Each node has a unique ID from 1 to N and the ID of BS is 0.

A. Genetic algorithm

Manuscript received November, 2010 and revised in January 9, 2011. This work is supported by educational center of the Science and Research Branch -Islamic Azad University Tehran, I. R. Iran and Dezfoul Branch I. R. Iran. .

Dr. J. Shanbehzadeh is with the Department of Computer Engineering at Science and Research Branch-Islamic Azad University - Tehran, I. R. Iran (e-mail: jamshid@tmu.ac.ir).

S. Mehrjoo is a postgraduate student in the field of computer in Dezfoul branch-Islamic Azad University , Young Researchers Club. He is interested in wireless sensor networks (email:sa.mehrjoo@yahoo.com).

Dr. Abdolhossein Sarrafzadeh is an Associate Professor and Head of Department of Computing at Unitec, New Zealand. (e-mail: hsarrafzadeh @unitec.ac.nz).

Genetic algorithm is an adaptive method that is generally employed in optimization problems. It is based on the genetic processes of biological organisms. GA upholds a population of chromosomes that develop according to selection, mutation, crossover, replacement rules, etc. Each chromosome has a measure of goodness called fitness. Selection function focuses on high fitness chromosomes. Mutation and crossover offer common heuristics that simulate the reproduction procedure.

We utilize GA to determine the number of cluster heads and choosing the best ones. Each chromosome consists of a sequence of bits in which every bit is represented as a sensor node that can be 0 or 1. Indexes of the chromosome are the nodes ID. The existence of 1 in the *ith* index of the chromosome means that node *i* is a cluster head and 0 means that the corresponding node is only a member of a cluster. For example in Fig. 1 node 1 is a member and node 3 is a cluster head.

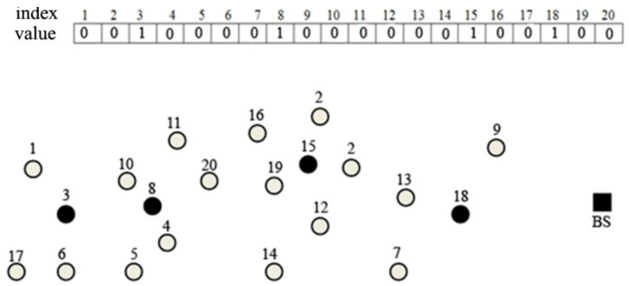

Fig. 1 Examples of chromosome and respective cluster heads; black dots are the cluster heads

Fitness function of GA algorithm is one of the most important elements of algorithm and specifies the goodness of the solutions. We explain the parameters of our GA fitness function in brief.

a) Sum of the cluster Heads Distance (SHD) to base station: SHD is the sum of all cluster heads from the base station and is defined as Equation (1):

$$\text{SHD} = \sum_{i=1}^{k} D_{ib} \qquad (1)$$

where k is the number of cluster heads and D_{ib} is the distance between cluster head i and the base station.

b) Sum of the cluster Heads Density (SHDS): The number of nodes around one node such that their distances are less than a threshold *Dst1,* is determined as density parameter of every node. SHDS is the sum of density of cluster heads and the greater, the SHDS, the better the chromosome.

$$\text{SHDS} = \sum_{i=1}^{k} Den_i \qquad (2)$$

Where k is the number of cluster heads and Den_i is the density of the cluster head i.

c) Sum of the cluster Heads Centrality (SHC): When several neighbor nodes have the same density parameter, the node at the center is the best choice to be the cluster head. SHC is the sum of centrality of cluster heads.

$$\text{SHC} = \sum_{i=1}^{k} Cent_i \qquad (3)$$

In Equation (3) k is the number of cluster heads and $Cent_i$ is the centrality of cluster head i.

d) Sum of the cluster Heads Residual Energy (SHRE): As cluster heads receive more packets and transmit them to long distance than member nodes, those nodes with a great residual energy are better choices to become cluster heads. We define SHRE as sum of the residual energy of the cluster heads.

$$\text{SHRE} = \sum_{i=1}^{k} RE_i \qquad (4)$$

In Equation (4) k is the number of cluster heads and RE_i is the residual energy of cluster head i.

Fitness function uses all the explained parameters to evaluate the goodness of each chromosome, Equation (5).

$$F = w1 * \left(\frac{1}{\text{SHD}}\right) + w2 * \text{SHC} + w3 * SHDS + w3 * SHRE \qquad (5)$$

where coefficients are initialized first and then updated as Equation (6),

$$\text{Wc} = \frac{Wp + \mid fc - fp \mid}{1 + e^{-fp}} \qquad (6)$$

Wp and Wc are the previous and current coefficients, fp and fc are the fitness values of the previous and current best chromosomes respectively.

B. *Particle Swarm Optimization*

Particle swarm optimization is one of the latest population based evolutionary optimization techniques which is based on the behaviors of bird flocking and fish schooling [11]. PSO is based on this scenario: there is a group of birds (fish) who search for food without knowledge about the exact place of it but, they know how far it is. Each bird (particle) can be informed about its best previous position and the best previous position of all other birds and how to follow these two positions. In PSO, each solution (particle) behaves like a bird in the search space. Each particle has a velocity too, which shows the direction of its flying and also has a fitness that shows how good this particle is. This fitness is calculated by a function. PSO initializes the population by randomly generated solutions and saves the best found position by all the particles and the best found position by particles in iterations. A potential solution can be achieved by the particle who updates its position and velocity based on Equation (7) and (8).

$$\text{V}_i^{(t+1)} = w * \text{V}_i^{(t)} + c_1 * \text{rand1()} * (\text{P}_i - \text{X}_i^{(t)}) + c_2 * \text{rand2()} * (\text{P}_g - \text{X}_i^{(t)}) \qquad (7)$$

$$\text{X}_i^{(t+1)} = \text{X}_i^{(t)} + \text{V}_i^{(t+1)} \qquad (8)$$

where $\text{X}_i^{(t)}$ and $\text{V}_i^{(t)}$ are the position and the velocity of particle i in the t^{th} iteration, respectively, and P_i is the best previous position of particle i and P_g is the best previous position of all the particles that have been found so far. W is the inertia factor that controls the trade-off between the local and global position direction. *rand1()* and *rand2()* are two

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

random numbers in interval [0,1]. The last two factors, c_1 and c_2, are scaling constants and are usually taken as $c_1 = c_2 = 2.0$.

In the proposed AI-based clustering, an array of length number of the live nodes of IDs is utilized to represent the particles. Indexes of the array represent node IDs and the values of respective cells are their cluster head ID. In this structure, those nodes which are chosen as cluster head, in the previous step, have 0 in their particular cells. By the use of such structure, we can assign nodes to the suboptimal cluster heads. Fig. 2 shows an example of clustering according to an instant particle.

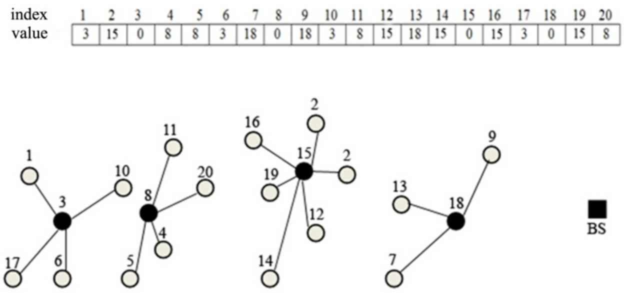

Fig.2 particle example and respective clustering

The energy consumption will be lower for transmitting and receiving the packets between members and cluster heads but it may make a great load on some cluster heads. Therefore, we employ some features to choose the semi-optimal cluster heads for nodes; here we describe some of our PSO fitness function parameters.

a) Sum of the Members' Distance (SMD) to the cluster heads: This parameter is the sum of the distance of members of all clusters from their cluster head and is defined as Equation.(9) :

$$\mathrm{SMD} = \sum_{i=1}^{n} D_{ih} \qquad (9)$$

where n is the number of nodes and D_{ih} is the distance between node i and its cluster head.

b) Energy Difference (ED): ED is the difference between energy of the particle and the best particle of the previous generation. Smaller energy difference illustrates better particle. This Energy difference is defined by Equation.(10) :

$$\mathrm{ED} = \mid E_{Net}^{k} - E_{Net}^{k+1} \mid \qquad (10)$$

c) Sum of the ratio of residual energy to the square of load (SREL): SREL is one of the most important parameters to balance the loads of nodes in tree and makes the node with the less residual energy to have fewer loads, so such a node will be considered as a relay node. This parameter is defined in Equation (11):

$$\mathrm{SREL} = \sum_{i=1}^{n} \frac{E_i}{(Load_i)^2} \qquad (11)$$

where, E_i is the residual energy of node i and $Load_i$ is the number of nodes which transmit their messages to node i.

All the mentioned parameters are used in fitness function; like as Equation.(12) to estimate the goodness of each particle which actually is a clustering.

$$F = w1 * \left(\frac{1}{\mathrm{SMD}}\right) + w2 * \left(\frac{1}{\mathrm{ED}}\right) + w3 * \mathrm{SREL} \qquad (12)$$

where coefficients are calculated same as coefficients of GA fitness function.

III. General Steps of Data Gathering

The whole operation on the network is divided into two phases:

1) Configuration phase
2) Data aggregation phase

1) Configuration phase: In this phase, the base station finds the best Hybrid-Way clustering configuration and then the base station broadcasts messages which contain complete network details to the sensor nodes. Some details are; the query execution plan, the number of cluster heads, the members associated with each cluster head and the number of transmissions for this configuration. Base station calculates the number of times that each clustering should be used to configure the WSN as follows;

$$T_i = \frac{fitness}{average\ fitness} * T_0 \qquad (13)$$

where T_0 is the default number of transmissions and is set to 10, $average\ fitness$ is the average of fitness value of previous clusters and T_i is the number of times which i'th clustering would be used. Equation.(13) shows that the fitness improvement of a clustering , the more that clustering would be used but if the fitness value is small, the respective clustering will be used less frequently. Also, as we approach the end of the lifetime of the network, the number of times that a specific clustering is used, is reduced in order to avoid the death of nodes due to over using of the configuration (for example the root node).

2) Data aggregation phase: In this phase, member nodes transmit the sensed data to their cluster heads and, each cluster head chooses the nearest neighbor cluster head or base station to transmit its packet and construct a multi hop configuration as shown in Fig. 3.

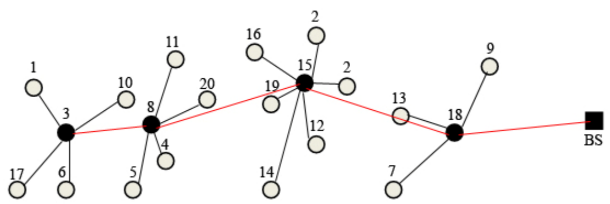

Fig.3 communication between cluster heads

IV. SIMULATION RESULTS

For each experiment, sensors are placed randomly in the field and the average of ten different experiments is used for performance evaluation.

The number of times that nodes transmit their messages to the cluster heads and cluster heads transmit the message to the base station from the beginning of the network life is defined as number of transmissions and is employed as a scale to compare the network lifetime and different

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

clustering methods. By the way the network will be alive until ten percents of the nodes are alive.

We employed two types of layouts in our simulation, random and grid. Wireless channel is assumed to be ideal and so there is no retransmission of control packets because of collision. We used a radio setting model the same as the one in [12]. In this model we used Equation (14) to transmit a message of length ℓ between a cluster member and its cluster head and used Equation (15) to transmit a message of length ℓ between a cluster head and the base station.

$$E_{Tih} = \ell E_e + \ell \varepsilon_s d_{ih}^2 \quad (14)$$
$$E_{Thb} = \ell E_e + \ell \varepsilon_\ell d_{hb}^4 \quad (15)$$
$$E_R = \ell E_e + \ell E_{BF} \quad (16)$$

where E_{Tih} is the energy of transmitting a message of length ℓ bits between node i and its cluster head h and E_{Thb} is the energy of transmitting a message of length ℓ bits between a cluster head *h* and the base station, d_{ih} is the distance between node *i* and its cluster head, d_{hb} is the distance between cluster head h and the base station and the energy of receiving a message of length ℓ bits is E_R and it is calculated using Equation (16). E_{BF} represents the cost of beam forming approach to reduce the energy consumption, ε_s is the energy consumed by the amplifier to transmit at a short distance, ε_l is the energy consumed by the amplifier to transmit at a long distance, E_e is the energy consumed in the electronics circuit to transmit or receive the signal. Values of these parameters are presented in Table 1.

Our hybrid GA-PSO-based clustering method is compared with LEACH, a famous clustering protocol, and with a genetic algorithm method employed in [7]. The parameters of the simulated network are briefly shown in Table 2 and the GA and PSO parameters are shown in Table 3 and 4 respectively.

TABLE.1 RADIO MODEL PARAMETERS

Parameter	Value
ε_s	10 pJ/bit/m2
ε_ℓ	0.0013 pJ/bit/m4
E_e	50 nJ/bit
E_{BF}	5 nJ/bit
ℓ	4000 bit
rate of data transfer	250e3

TABLE.2 NETWORK PARAMETERS

Parameter	value
Number of nodes	200
Network size	100*100 m^2
Distance to base station	200 m (away from the network)
Initial node energy	2 j
Minimum energy	0.0001 j
Network threshold	0.1 of nodes be alive
data transfer rate	bandwidth = 1Mbps

TABLE.3 GA PARAMETERS

Parameter	value
Population size	50
Number of generations	30
mutation rate	0.006
crossover rate	0.8
tournament selection	probability of 0.9

TABLE.4 PSO PARAMETERS

Parameter	value
Population size	50
Number of generations	30
C1	2.0
C2	2.0
V_{max}	4.0
V_{min}	4.0
ε	20

Fig.4 shows the percentage of alive nodes with respect to the number of transmissions (lifetime) for two types of layout. This graph shows that hybrid GA-PSO clustering method extends the lifetime of WSN much more than Simple GA and LEACH. Fig.5 shows the number of nodes with respect to the number of transmissions. Fig.6 shows the number of cluster heads with respect to the number of transmissions. The graph of Fig.6 is the simulation result of a random layout network with 100 nodes.

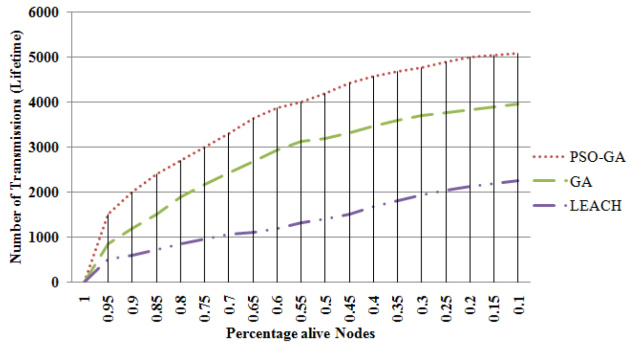

Fig.4 a) random layout

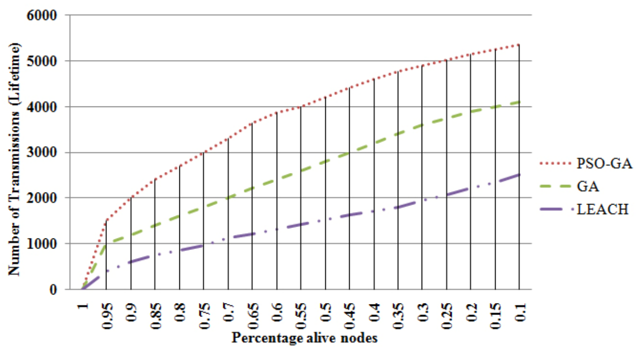

Fig.4 b) grid layout

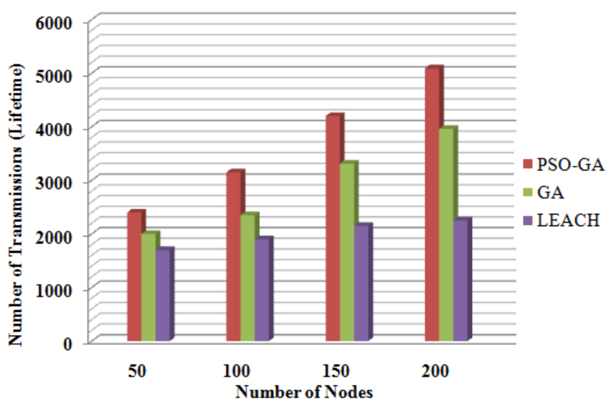

Fig.5 a) random layout

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

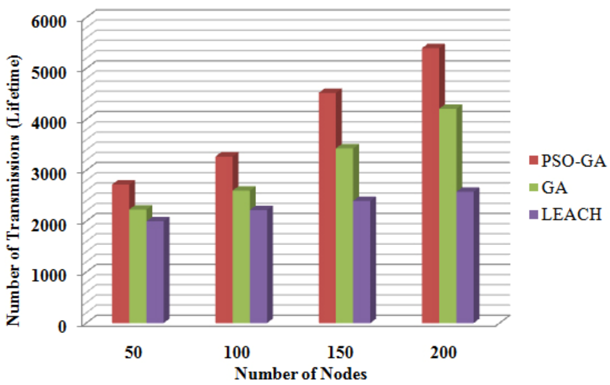

Fig.5 b) grid layout

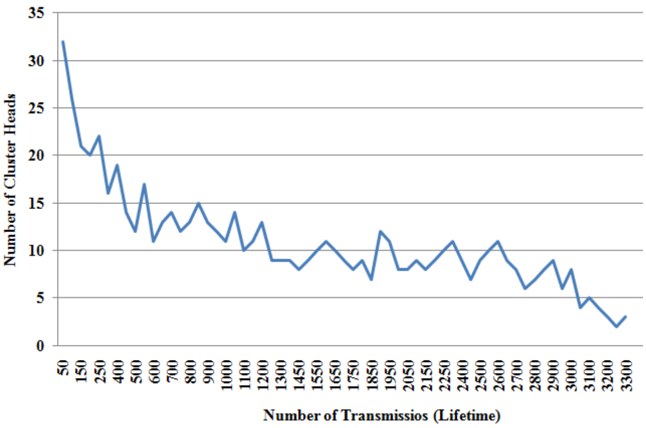

Fig.6 Number of cluster heads

V. CONCLUSION AND FUTURE WORKS

This paper proposed a hybrid GA-PSO based clustering algorithm that improved the lifetime of WSN effectively. We utilized GA to choose the cluster heads and their number and PSO to select the clusters' members. Simulation results showed that this method was much better than classic LEACH and Genetic algorithm clustering method.

Further investigations may include the use of other intelligent algorithms instead of PSO. Also, data aggregation tree may be developed in WSN which presents better results in energy consumption compared with clustering.

REFERENCES

[1] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam, E.Cayirci,"Wireless sensor networks: a survey", Computer Networks 38, pp. 393 422, 2002.

[2] Tatiana Bokareva, Wen Hu, SalilKanhere, "Wireless Sensor Networks for Battlefield Surveillance", Land Warfare Conference Brisbane, October 2006.

[3] Werner Allen, G. Johnson, J. Ruiz, "Monitoring volcanic eruptions with a wireless sensor network ", Proceedings of the Second European workshop on Wireless Sensor Network,2005.

[4] Junguo Zhang, Wenbin Li, Zhongxing Yin, Shengbo Liu, XiaolinGuo, "Forest fire detection system based on wireless sensor network ", 4th IEEE Conference on Industrial Electronics and Applications, 2009.

[5] Lee Angeles, Talampas Sison, "MotesArt: Wireless Sensor Network for Monitoring Relative Humidity and Temperature in an Art Gallery ", IEEE International Conference on networking, sensing and control,.ICNSC 2008.

[6] D. Goldberg, B. Karp, Y. Ke, S. Nath, and S. Seshan. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, 1989.

[7] SajidHussain, Obidul Islam, AbdulWaseyMatin, "Genetic algorithm for energy efficient clusters in wireless sensor networks.", In Proceedings of the 4th International Conference on Information Technology). IEEE Computer Society, April 2007.

[8] Jason Sidabras," A Survey on Particle Swarm Optimization", MSCS 252 - Deterministic Models of Operations Research, December 14, 2007.

[9] Ying Liang, Haibin Yu, "PSO-Based Energy Efficient Gathering in Sensor Networks", Springer-Verlag Berlin Heidelberg 2005.

[10] Erfu Yang, Ahmet T. Erdogan, TughrulArslan, Nick Barton," An Improved Particle Swarm Optimization Algorithm for Power-Efficient Wireless Sensor Networks", IEEE 0-7695-2919-4/07 2007.

[11] J. Kennedy, R. Eberhart, Particle Swarm Optimization, IEEE International Conference on Neural Networks, Perth, Australia. 1995.

[12] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, "Energy efficient communication protocol for wireless micro sensor networks," in Proceedings of the Hawaii International Conference on System Sciences, January 2000.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)